arXiv:2301.08102v1 [physics.data-an] 19 Jan 2023

用于推断稀疏观测随机非线性系统的几何路径增强

Dimitra Maoutsa*柏林工业大学，德国柏林
dimitra.maoutsa@tu-berlin.de

摘要

描述系统在确定性力和随机力共同作用下动力学的随机演化方程，在各个科学领域都十分常见。然而，依据稀疏时间观测来识别这些系统仍然是一项艰巨的任务。现有方法要么依赖条件期望，着重考察观测值的时间结构，从而舍弃了系统不变密度几何结构中蕴含的信息；要么采用不变密度的几何近似，但这类方法仅适用于具有保守力的系统。本文提出一种方法，将这两种范式结合起来。我们提出一种考虑局部观测几何的数据驱动路径增强方案。通过对增广路径进行非参数推断，我们能够在低采样率观测条件下，高效识别底层系统的确定性驱动力。

1 引言

随机微分方程是一类表现力尤为丰富的动力学模型，天然适合描述具有多重时间尺度的演化系统 [1–4]。从这类系统中提取随机演化方程一直是许多科学领域关注的重要课题 [5–15]。过去，连续时间确定性模型的识别问题已基本得到解决 [16–19], ，但随机对应模型的识别并非如此。在观测以低采样率（观测间隔较长）采集的情况下，随机系统的推断尤其具有挑战性。

现有识别随机系统中确定性驱动力的方法，大多依赖于对**不变密度**的近似（例如密度估计 [20] 或扩散映射等谱方法 [21–25]）（几何方法），或者通过计算状态增量的条件期望来考虑观测值的**时间结构** [26–37]（时间方法）。然而，几何方法仅适用于具有保守力的系统，因为它们假设漂移是某个势能的梯度 [20, 24, 38], ，或者假设状态变量完全解耦 [21]。另一方面，当观测间隔较大（观测稀疏）时，时间方法表现不佳 [39], ，因为在这些情况下根据观测值计算出的状态增量（见附录公式（11））无法反映真实的底层动力学（图1）。

为减轻观测稀疏的影响，一部分时间方法采用**路径增广**[2]，通过采样**扩散桥**来近似连续观测值之间的转移密度；扩散桥是受初始状态和终止状态约束的扩散过程 [31–35]。然而，大多数非参数方法采用简化的桥动力学（例如布朗运动 [31,

*https://dimitra-maoutsa.gitlab.io/

[2]本文使用**路径增广**一词，指通常所称的**数据增强**。我们选择这一术语，是因为它更优雅，也更能准确描述所提出的增广过程；而“数据增强”一词则相当含糊。

Machine Learning and the Physical Sciences 研讨会，NeurIPS 2022。

35] 或奥恩斯坦-乌伦贝克桥 [39]），当观测到的系统为非线性时，这些方法无法准确反映底层的转移密度（图2 c 和 d）。

另一种路径增广策略是考虑粗略的漂移估计（例如，假设观测值之间服从高斯似然，见式（2）），随后采用随机桥采样器 [40–42]，利用估计的非线性漂移构造随机桥。然而，对于较大的观测间隔，在估计扩散过程的概率律下，观测值的概率为零（图2 b.）。因此，根据所采用的框架，任何试图遵循估计动力学在连续观测值之间构造扩散桥的方法，要么收敛速度较慢，要么完全失败。

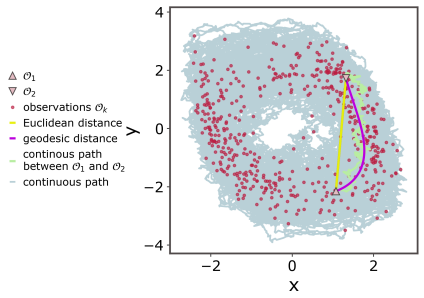


图 1：**高斯似然假设下低频观测所考虑的状态增量**。欧几里得距离（黄色线条）——用于计算连续观测值之间的状态增量——未考虑不变密度的曲率。测地曲线（紫色线条）能更好地近似系统在连续观测值之间未观测到的状态（浅绿色线条）。

在此，我们提出一种替代方法。我们假设，增广路径应位于**附近测地曲线**（图1b，洋红色）上。这些曲线连接由系统不变密度所诱导的流形上的连续观测值。为此，我们提出一种路径增广框架，构造**几何约束桥** [3]，通过将增广路径推向连接连续观测值的相应测地线（图2 e）。为此，我们采用 [40, 41]中提出的随机控制框架，并施加路径约束，引导增广路径趋向连接连续观测值的测地曲线。

2 设定

我们考虑由如下形式的随机微分方程（SDE）描述的随机系统：

$$\mathrm{d}X_t = f(X_t)\mathrm{d}t + \sigma\,\mathrm{d}\boldsymbol{\beta}_t, \qquad X_0 = x_0\,, \tag{1}$$

其中 $f(\cdot): R^d \to R^d$ 表示确定性驱动力（漂移函数），而 $\sigma\mathrm{d}\boldsymbol{\beta}_t$ 表示作用于系统的随机力（扩散）。这里，$\sigma \in \mathcal{R}^{d\times d}$ 表示噪声幅度，$\boldsymbol{\beta}$ 表示由 d 个相互独立的维纳过程组成的向量。下文中，我们采用随机积分的伊藤解释。我们通过观测过程 $\mathcal{O}_k = \psi(X_{k\tau})$ 观测系统，其中 $X_{k\tau} \doteq X_t|_{t=\tau k}$，$k = 1, 2, \ldots, K$ 个观测值在**观测间隔** τ处测量系统状态。为简单起见，我们将观测过程视为恒等函数，即 $\psi(x) = x$，但该形式体系很容易推广到更一般的函数。

高频观测。对于足够精细的观测时间网格，我们假设观测值代表连续时间下的系统状态，即我们观测到连续路径 $X_{0:T}$。由此，我们可以通过对状态增量的条件期望进行经验估计 [27–29, 44]，近似一阶克拉默斯-莫亚尔系数 [43]，从而估计漂移。类似的贝叶斯非参数方法 [45] 认为，当 d $t \to 0$时，观测值之间的转移概率服从高斯分布，因此观测值的似然为（高斯）分布（参见第A 节，式7）

$$\mathcal{L}(X_{0:T} \mid f) = \exp\left[-\frac{1}{2}\int_0^T \|f(X_t)\|^2_{\sigma^2}\mathrm{d}t + \int_0^T \langle f(X_t), X_{t+\mathrm{d}t} - X_t\rangle \mathrm{d}t\right], \tag{2}$$

and 对函数值 f（式（12））施加高斯过程先验。在式（2）中，我们引入 d the 记号 $\langle u, v\rangle \doteq u^\top \cdot \sigma^{-2} v$ 和 $\|u\|_{\sigma^2} \doteq u^\top \cdot \sigma^{-2} u$。

低频观测。随着观测间隔 τ 增大，假设相邻两次观测之间服从高斯分布的似然（式（2））在式（1）为非线性时不再成立。同样，在此情形下计算得到的状态增量 $X_{t+\tau} - X_t$ 无法准确反映底层动力学（图1 a.）。漂移的似然 $\mathrm{P}(\{\mathcal{O}_k\}_{k=1}^K|f)$ 在此类情形下呈现为路径积分

$$\mathrm{P}(\{\mathcal{O}_k\}_{k=1}^K \mid f) = \int \mathrm{P}(\{\mathcal{O}_k\}_{k=1}^K, X_{0:T} \mid f)\mathcal{D}(X_{0:T}) = \int \mathrm{P}(\{\mathcal{O}_k\}_{k=1}^K \mid X_{0:T})\mathrm{P}(X_{0:T}|f)\mathcal{D}(X_{0:T}),, \tag{3}$$

[3]严格来说，这些构造已不再是扩散桥，而是受约束扩散路径。在此，我们将扩散桥的概念加以扩展，以区别于路径增广中通常采用的扩散桥。

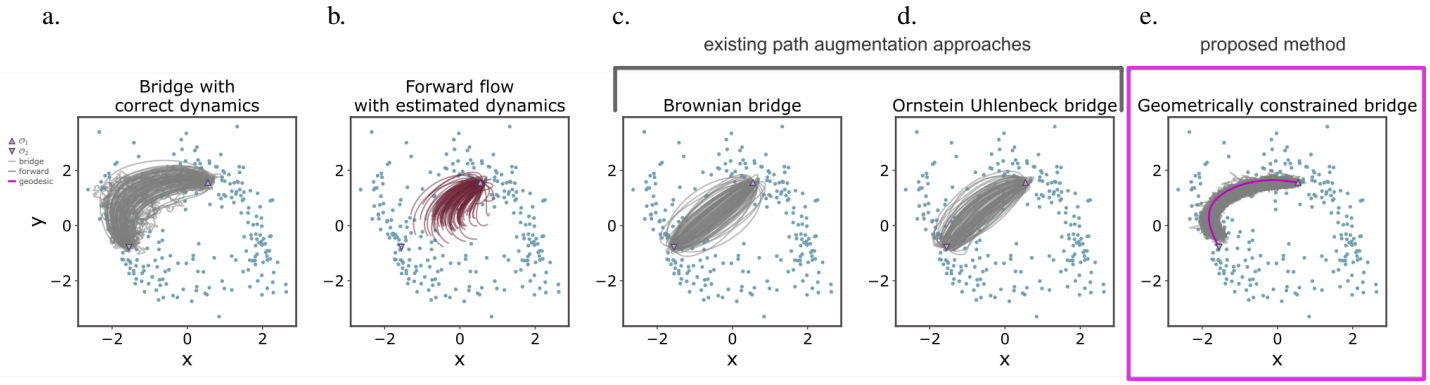


图 2：**现有路径增广策略对连续观测值之间的底层转移密度拟合较差，低估了其曲率。a.)** 遵循真实动力学时，两个连续观测值 $\mathcal{O}_1$ 和 $\mathcal{O}_2$（粉色三角形）之间的随机桥边缘密度（灰色）。**b.)** 采用高斯似然的估计动力学前向概率流（栗色）无法很好地匹配正确的转移密度，并且通常无法到达第二个观测值 $\mathcal{O}_2$（向下的粉色三角形）。常见的路径增广策略采用以下方法之一：**c.)** 布朗桥，或 **d.)** 由基于高斯似然的估计漂移局部线性化得到的奥恩斯坦-乌伦贝克（线性）桥边缘分布。这两种方法都无法很好地匹配正确的转移密度，因为它们低估了其曲率。**e.)** 所提出的几何约束路径增广通过将桥路径引向流形上连接连续观测值的测地曲线，从而更好地近似底层转移密度；该流形由观测值诱导。

其中 $\{\mathcal{O}_k\}_{k=1}^K$ 表示 K 离散时间观测值，$\mathrm{P}(X_{0:T}|f)$ 表示由式（1）中的系统所决定的先验路径概率，$\mathcal{D}(X_{0:T})$ 表示路径空间上的形式体积元，而 $\mathrm{P}(\{\mathcal{O}_k\}_{k=1}^K|X_{0:T})$ 表示给定潜在路径 $X_{0:T}$时观测值的似然。

从几何视角来看，我们可以认为，系统的不变密度可由非线性系统动力学诱导的一个（低维）流形近似。观测值本质上是该流形的样本。对于低频观测值，用于计算状态增量的欧几里得距离 $X_{t+\tau} - X_t$ 未考虑非线性动力学诱导的几何结构，因此低估了连续观测值之间转移密度的曲率（图1）。

3 方法

由于方程（3）的似然难以处理，我们将未观测到的连续路径视为潜在随机变量 $X_{0:T}$，并采用期望最大化（EM）[46] 来确定漂移函数的最大后验估计。过去已有类似的参数 [33, 47] 和非参数 [39, 45]方法用于漂移推断，主要针对高频观测场景。我们的方法受到 [39, 45] 中采用的非参数方法的启发，并有两项关键创新：

（i）我们采用一种路径增广方案，遵循**通过推断得出的非线性动力学**，该推断使用方程（2）的高斯似然（而非 [39]提出的这些动力学的局部线性近似）。（ii）重要的是，我们进一步**约束增广路径，使其在相邻观测值之间符合不变密度的几何形状**（图1 b）。

我们采用迭代算法，在每次迭代中 n 执行以下两个步骤：

(1.) 一个**期望步**，在给定漂移估计的情况下 $\hat{f}^n$ 我们为潜变量 $Q(X_{0:T}) \approx \mathrm{P}(X_{0:T}|\{\mathcal{O}\}_{k=1}^K, \hat{f}^n(x))$构造一个近似后验分布。

(2.) 一个**最大化步**，在此步骤中我们更新漂移估计。

• **路径的近似后验。（E 步）**我们将观测值之间的连续路径轨迹$X_{0:\ T}$近似为一个后验路径测度，该测度定义为自由能的最小化解

$$\mathcal{F}[Q] = \frac{1}{2}\int_0^T \int \Big[\|g(x,t) - \hat{f}(x)\|^2_{\sigma^2} + U_{\mathcal{O}}(x,t) + U_{\mathcal{G}}(x,t)\Big]\, q_t(x)\,\mathrm{d}x\,\mathrm{d}t. \tag{4}$$

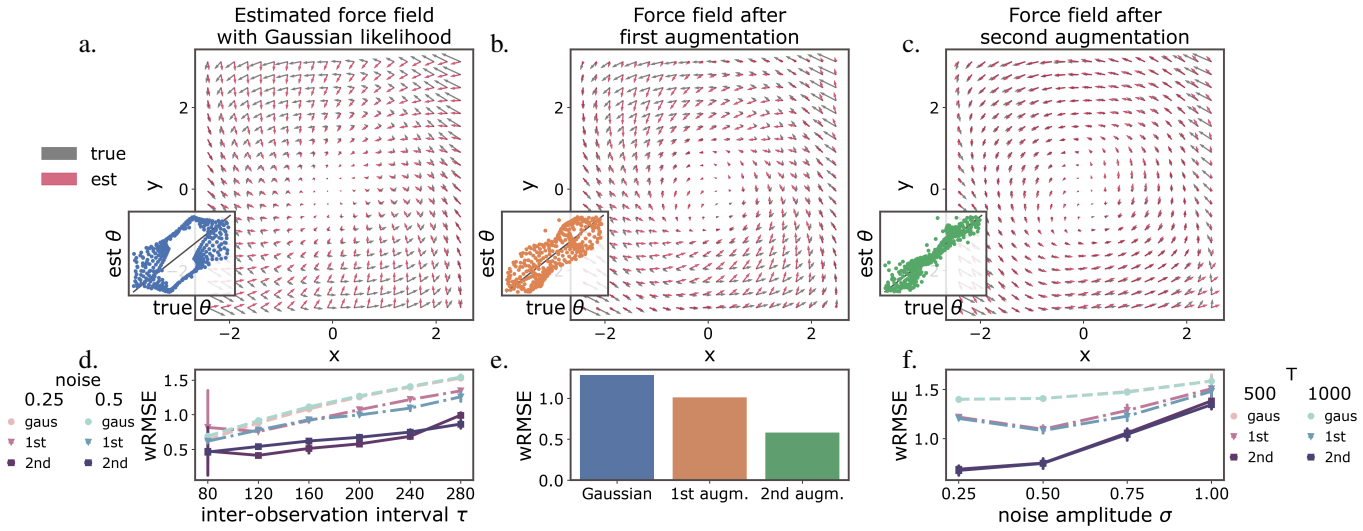


图3：**经过两次迭代后，所提出的路径增广已能很好地近似底层漂移**。采用**a.)** 高斯似然**b.)** 经过一次增广迭代后，以及**c.)** 经过第二次增广迭代后，估计得到的（红色）和真实的（灰色）力场。（**插图**）二维网格上各点的真实角度与估计角度对比。**e.)** 所示示例中，每次迭代后估计漂移的加权均方根误差（wRMSE）。用于对各网格点误差进行加权平均的权重，由观测值 $\{\mathcal{O}_k\}_{k=1}^K$ 上的核密度估计得到。**d.)** 不同噪声条件 $\sigma = \{0.25, 0.5\}$ 下，漂移估计的 wRMSE 随观测间隔 τ 的变化；漂移估计采用高斯似然（*gaus*-圆形标记）、第一次增广后（*1st*-三角形标记）和第二次增广后（*2nd*-方形标记），对应 $T = 500$。**f.)** 系统中的 wRMSE 随噪声幅度 σ 的变化；不同曲线对应不同的轨迹时长 $T = \{500, 1000\}$ 时间单位，观测间隔为 $\tau = 240$。标记的含义与 d.) 中相同。误差条表示 5 次独立实现的一个标准差。

项 $U_{\mathcal{O}}(x,t) \doteq -\sum_{t_k} \ln \mathrm{P}(\mathcal{O}_k|x)\delta(t-t_k)$ 迫使潜在路径经过观测值（具体而言，取决于观测过程，也可能只需接近观测值），而 $U_{\mathcal{G}}(x,t) \doteq \|\Gamma_t - x\|^2$ 则引导潜在路径趋向于连接流形上连续观测值的测地曲线 $\gamma_{t'}^k$；该流形由系统的不变密度诱导而成（第 A.1.2 节）。这里记作 $\Gamma_t \doteq \{\gamma_{t'}^k\}_{t=(k-1)\tau+t'\tau}$，其中 $\gamma_{t'}^k$ 是连接 $\mathcal{O}_k$ 和 $\mathcal{O}_{k+1}$的测地线，而 $t' \in [0,1]$。我们通过学习流形的局部度量 $\mathcal{M}$ 来确定每个区间的测地线 $\gamma_{t'}^k$（参见第 A.1.2 节和 [48]）。

根据 [49], 对于每个观测间隔 $[\mathcal{O}_k, \mathcal{O}_{k+1}]$，我们通过求解一个随机最优控制问题 [14, 40, 41] 来确定后验路径测度（式 (4) 的极小化解） u，其目标是为漂移为 $f(x)$ 的系统确定时变漂移调整 $\hat{}(x,t) := g(x,t) - f(x)$，并满足由 $U_{\mathcal{O}}(x,t)$ 确定的初始和终端约束，以及额外的路径约束 $U_{\mathcal{G}}(x,t)$。

- **漂移估计。（M步）**为了从采样的潜在路径估计漂移，我们假设函数值服从高斯过程先验，并采用类似于 [39]（详情请参见第 A.2 节）。

4 数值实验

为展示所提出方法的性能，我们针对不同噪声条件下的二维范德波尔振荡器进行了系统估计 σ，并在不同观测间隔下进行观测 τ，同时考察了不同轨迹长度 T（见第D节）。对于所考察的噪声幅度（图3f）以及每个振荡周期内观测次数超过一次的观测间隔（图3d），所提出的路径增强算法在两次迭代内，对于大多数噪声幅度（图3），都能改善基于高斯假设的朴素估计。随着噪声增大，我们的方法带来的改进会减弱（图3f），但仍不可忽略。

5 结论与讨论

我们提出了一种新方法，可从系统状态的稀疏时间观测中识别随机系统。我们提出了一种路径增广策略，该策略利用粗略漂移估计的非线性动力学，并进一步约束增广路径遵循系统不变密度的局部几何结构。我们发现，在若干噪声条件下，对于周期或准周期系统，所提出的方法能够有效恢复其底层漂移函数。

用于推断的几何约束。我们的方法调和了两类方法：一类完全依赖观测值的时间结构，另一类则近似不变密度并忽略测量的时间顺序。随着几何统计学领域的近期发展 [50,51], 以及流形假设这一概念日益受到关注 [52, 53], , 即认为多维动力系统的状态往往局限于状态空间中的低维区域，已有若干推断方法尝试结合几何视角与时间视角，以识别随机系统。在朗之万回归框架中 [54], Callaham 等人计算克拉默斯-莫亚尔系数，并如 Lade 提出的那样，通过求解系数对应的伴随福克-普朗克方程，修正低采样率造成的估计偏差 [55]。他们还通过对观测密度与所估计随机微分方程模型的平稳福克-普朗克概率密度进行矩匹配，加入正则化，从而纳入几何约束。在 [56] Tong 等人的方法中，他们考虑观测值构成的流形，以推断细胞动力学。该方法采用动态最优传输，在约束测量分布位于观测值附近的条件下对其进行插值。这种方法与我们的方法有相同的直觉，不过 Tong 等人没有使用随机微分方程来建模本质上具有随机性的细胞动力学。此外，他们没有尝试对数据的底层几何结构进行建模，而只考虑惩罚观测值与各个观测值之间距离的约束。Shnitzer 等人 [53,57] 采用扩散映射来近似后向柯尔莫哥洛夫算子的特征函数（该算子是随机 Koopman 算子的生成元 [58, 59]），并且由于特征函数遵循线性演化方程，他们使用卡尔曼滤波器演化主导算子特征谱，以纳入观测值的时间顺序。然而，他们的方法仅适用于保守系统，并假设近似算子的谱存在谱隙，因此无法处理具有连续谱的系统，例如混沌系统 [60, 61]。

随机过程的测地曲线及昂萨格-马赫卢普意义下的最概然路径。我们工作的理论基础可以追溯到 Onsager 和 Machlup 的研究 [62]，以及对扩散过程在两个预先确定的状态之间的**最概然路径(最概然路径)**的计算。早期研究采用**昂萨格-马赫卢普（OM）函数**作为拉格朗日量，推导出用状态变量和扩散过程的漂移表示**最概然路径**的表达式 [63–69]。所得拉格朗日量包含路径能量（见第B节），这也是用于确定测地线的同一目标函数（见第A.1.2节）。在我们的框架中，为了确定连续观测值之间的测地线，我们假设 $\mathcal{R}^d$ 是一个光滑流形，其配备的黎曼度量 h 由数据学习得到。然而，底层随机微分方程是在 $\mathcal{R}^d$ 上以欧几里得度量 [70]定义的。定义扩散过程时采用不同的度量会导致不同的生成元，因而每个过程对应不同的路径概率。计算改变过程定义中的黎曼度量所引起的变换，将是一个有趣的理论结果。据我们所知，文献中尚无此类结果，而且推导并不简单。度量的改变会引起过程扩散率的变化，因此无法直接使用吉尔萨诺夫变换。不过，或许可以采用类似于逆兰佩蒂变换的变换 [71]，将过程的漂移表示为一种具有乘性噪声的扩散；这种扩散会产生由观测值学习得到的度量变化（见第B节）。最后，与**昂萨格-马赫卢普（OM）**泛函的联系已表明，借助几何考量来估计未知漂移还可采用另一种方法，从而避免计算开销巨大的连续路径模拟。

局限性。 我们的方法仅适用于不变密度可由一个流形近似表示，且能够在该流形上识别测地线的系统。此外，按照构造，该方法隐含地假设不变系统的密度近似均匀采样。不过，我们预见，借助几何统计学中更先进的工具 [50, 51] 该框架或可适用于非均匀采样的不变密度。我们的实验表明，所提出的方法更适用于具有循环平衡 [72], 的系统，即稳态下存在稳态涨落概率流的系统。对于稳态下具有非涨落电流的系统，依赖保守力假设的现有方法仍能有效地恢复其状态。

致谢

我们感谢 Nina Miolane 提供详细信息，说明如何从变分自编码器近似得到的流形中计算测地线。我们还感谢 Georgios Arvanitidis 维护公开可用的代码库，其中包含本文用于构造测地线的算法；感谢 Stefan Sommer 解答有关流形上扩散过程变换的问题；并感谢 Manfred Opper 教授促使我们研究这一问题并提供初步指导。我们也要感谢 Python [73], NumPy [74],SciPy [75], Matplotlib [76], Seaborn [77], GPflow [78],pyEMD [79], 以及 PyTorch [80] 社区此前的工作，它们为本文计算部分的实现提供了便利。

本研究的实现版本将在以下代码仓库中发布：
https://github.com/dimitra-maoutsa/Geometric-path-augmentation-for-SDEs待文章提交归档出版后。

更广泛影响声明

我们提出了一种新的路径增强方法，可对观测间隔较大的随机系统进行高效推断。我们的贡献旨在强调，将快速发展的几何统计学领域中的理念纳入随机系统模型发现领域的必要性。过去，人们对确定性系统不变密度的几何与拓扑性质进行了大量研究，但对其随机对应模型的研究却并非如此，尤其是对由随机微分方程描述的系统。

我们的工作还旨在进一步指出，在增强数据量远超观测值数量的情况下，数据增强框架若能在增强过程中融入领域知识或其他类型的信息（例如此处关于系统不变密度几何结构的信息），便可能提高推断的准确性。许多采用数据增强框架的算法仅表现出**局部收敛**，例如此处采用的期望最大化算法 [81]。因此，在初始估计与真实值相差甚远的情况下，朴素的数据增强策略可能会收敛到次优解，无法反映真实值。

我们预计本研究不会产生任何直接的社会影响。不过，我们承认随机系统可能用于军事用途和金融工程；但所提出的方法并未直接建议对观测到的系统进行干预，因此不会导致不利后果。

扩散系统广泛存在于多个科学领域，例如物理学、生物学、神经科学和生态学的部分领域。我们预计，这项工作可为这些学科提供识别感兴趣系统的工具，从而使其受益。

参考文献

[1] StevanJArnold. 表型进化：持续的综合：（美国博物学家学会演讲）. 2014.[2] Russell Lande. “表型进化中的自然选择与随机遗传漂变”. In:《*Evolution*》(1976), pp. 314–334. [3] SubrahmanyanChandrasekhar. “物理学与天文学中的随机问题”。载于：《现代物理评论》15.1 (1943), p. 1.[4] Philip Nelson.《生物物理学》. WH Freeman，纽约，2004。[5] Tom Kuusela. “随机心率模型可揭示病理性心脏动力学”。载于：*Physical Review E* 69.3 (2004), p.031916.[6] Philip Sura. “南大洋和太平洋海面风的随机分析”。载于：《大气科学杂志》60.4 (2003), pp. 654–666.[7] Jose Casadiego、Dimitra Maoutsa 和 Marc Timme。“根据事件定时模式推断网络连通性”。载于：*Physical Review Letters* 121.5 (2018), p. 054101.[8] Philip Sura 和 Sarah T Gille. “在随机框架下解读风驱动的南大洋变率”。载于：《海洋研究杂志》61.3 (2003), pp. 313–334.

[9] TD Frank、R Friedrich 和 PJ Beek。“通过漂移-扩散估计揭示的等长力产生随机序参量方程”。载于：Physical Review E 74.5 (2006), p. 051905。DOI: https://doi.org/10.1103/physreve.74.051905。[10] Anke M van Mourik、Andreas Daffertshofer 和 Peter J Beek。“人类节律性运动的确定性与随机特征”。载于：生物控制论 94.3 (2006), pp. 233–244。[11] Juraj Bergman、Dominik Schrempf、Carolin Kosiol 和 Claus Vogl。“利用前向和后向离散时间与连续时间过程进行群体遗传学推断”。载于：Journal of Theoretical Biology 439 (2018), pp. 166–180。[12] Raphael Sarfati、Jerzy Bławzdziewicz 和 Eric R Dufresne。“基于单条短布朗轨迹对力和迁移率进行最大似然估计”。载于：Soft Matter 13.11 (2017), pp. 2174–2180。[13] Georg A. Gottwald、Daan T. Crommelin 和 Christian L. E. Franzke。“随机气候理论”。载于：非线性与随机气候动力学 (2017), pp. 209–240。DOI: https: //doi.org/10.1017/9781316339251.009。[14] Dimitra Maoutsa。“从单次试验的脉冲序列观测中揭示潜在随机动力学”。Bernstein计算神经科学会议，柏林，2022。2022年9月。DOI: 10.12751/nncn.bc2022.250。[15] Nazareno Campioni、Dirk Husmeier、Juan Morales、Jennifer Gaskell 和 Colin J Torney。“从宏观尺度观测推断相互作用系统的微观尺度性质”。载于：Physical Review Research 3.4 (2021), p. 043074。[16] J Cremers 和 A Hübler。“根据实验数据构建微分方程”。载于：Zeitschrift für Naturforschung A 42.8 (1987), pp. 797–802。[17] Steven L Brunton、Joshua L Proctor 和 J Nathan Kutz。“通过稀疏识别非线性动力系统，从数据中发现控制方程”。载于：《美国国家科学院院刊》113.15 (2016), pp. 3932–3937。DOI: 10.1073/pnas. 1517384113。[18] Bryan C Daniels 和 Ilya Nemenman。“现象学动力学模型的自动化自适应推断”。载于：Nature Communications 6.1 (2015), pp. 1–8。DOI: 10.1038/ ncomms9133。[19] Kevin McGoff、Sayan Mukherjee 和 Natesh Pillai。“动力系统的统计推断：综述”。载于：Statistics Surveys 9 (2015), pp. 209–252。[20] Philipp Batz、Andreas Ruttor 和 Manfred Opper。“根据经验密度对随机微分方程的漂移进行变分估计”。载于：Journal of Statistical Mechanics: Theory and Experiment 2016.8 (2016), p. 083404。[21] Amit Singer 和 Ronald R Coifman。“结合扩散映射的非线性独立成分分析”。载于：Applied and Computational Harmonic Analysis 25.2 (2008), pp. 226–239。[22] Feliks Nüske、Péter Koltai、Lorenzo Boninsegna 和 Cecilia Clementi。“基于条件期望的有效动力学谱性质”。载于：Entropy 23.2 (2021), p. 134。DOI: https://doi.org/10.3390/e23020134。[23] Edward L Ionides、Carles Bretó 和 Aaron A King。“非线性动力系统的推断”。载于：《美国国家科学院院刊》103.49 (2006), pp. 18438–18443。DOI: https://doi.org/10.1073/pnas.0603181103。[24] Ronen Talmon 和 Ronald R Coifman。“利用经验几何对随机动力系统进行内在建模”。载于：Applied and Computational Harmonic Analysis 39.1 (2015), pp. 138–160。[25] Carmeline J Dsilva、Ronen Talmon、C William Gear、Ronald R Coifman 和 Ioannis G Kevrekidis。“多尺度快慢随机动力系统的一类数据驱动降阶方法”。载于：SIAM Journal on Applied Dynamical Systems 15.3 (2016), pp. 1327–1351。[26] Rudolf Friedrich 和 Joachim Peinke。“用福克-普朗克方程描述湍流级联”。载于：Physical Review Letters 78.5 (1997), p. 863。DOI: https://doi.org/10. 1103/PhysRevLett.78.863。[27] Silke Siegert、R Friedrich 和 J Peinke。“随机系统数据集的分析”。载于：Physics Letters A 243.5-6 (1998), pp. 275–280。DOI: 10.1016/S0375-9601(98)00283-7。

[28] Mario Ragwitz 和 Holger Kantz。《从时间序列数据推导福克-普朗克方程不可或缺的有限时间修正》。发表于：Physical Review Letters 87.25 (2001)，第 254501 页。DOI：10.1103/physrevlett.87.254501。[29] Lorenzo Boninsegna、Feliks Nüske 和 Cecilia Clementi。《随机动力学方程的稀疏学习》。发表于：The Journal of chemical physics 148.24 (2018)，第 241723 页。DOI：10.1063/1.5018409。[30] David Lamouroux 和 Klaus Lehnertz。《随机过程漂移系数和扩散系数的基于核的回归》。发表于：Physics Letters A 373.39 (2009)，第 3507–3512 页。[31] Andrew Golightly 和 Darren J Wilkinson。《对含误差观测的非线性多变量扩散模型进行贝叶斯推断》。发表于：Computational Statistics & Data Analysis 52.3 (2008)，第 1674–1693 页。[32] Omiros Papaspiliopoulos、Yvo Pokern、Gareth O Roberts 和 Andrew M Stuart。《扩散过程的非参数估计：一种微分方程方法》。发表于：Biometrika 99.3 (2012)，第 511–531 页。[33] Giorgos Sermaidis、Omiros Papaspiliopoulos、Gareth O Roberts、Alexandros Beskos 和 Paul Fearnhead。《用于扩散过程精确推断的马尔可夫链蒙特卡罗》。发表于：Scandinavian Journal of Statistics 40.2 (2013)，第 294–321 页。[34] Alexandros Beskos、Omiros Papaspiliopoulos 和 Gareth O Roberts。《扩散样本路径的回溯精确模拟及其应用》。发表于：Bernoulli 12.6 (2006)，第 1077–1098 页。[35] Siddhartha Chib、Michael K Pitt 和 Neil Shephard。《扩散驱动状态空间模型的似然推断》。发表于：Por Clasificar (2006)，第 1–33 页。[36] Cagatay Yildiz、Markus Heinonen、Jukka Intosalmi、Henrik Mannerstrom 和 Harri Lahdesmaki。《无需梯度匹配的高斯过程随机微分方程学习》。发表于：2018 IEEE 第 28 届机器学习信号处理国际研讨会 (MLSP)。IEEE。2018，第 1–6 页。[37] Rudolf Friedrich、Joachim Peinke、Muhammad Sahimi 和 M Reza Rahimi Tabar。《以随机方法探索复杂性：从生物系统到湍流》。发表于：Physics Reports 506.5 (2011)，第 87–162 页。[38] Tyrus Berry 和 John Harlim。《用于特征识别的迭代扩散映射》。发表于：Applied and Computational Harmonic Analysis 45.1 (2018)，第 84–119 页。[39] Philipp Batz、Andreas Ruttor 和 Manfred Opper。《随机微分方程的近似贝叶斯学习》。发表于：神经信息处理系统进展 26 (2013)。[40] Dimitra Maoutsa 和 Manfred Opper。《用于约束随机非线性系统的确定性粒子流》。发表于：Phys. Rev. Research 4 (4 Oct. 2022)，第 043035 页。DOI：10. 1103/PhysRevResearch.4.043035。URL：https://link.aps.org/doi/10.1103/ PhysRevResearch.4.043035。[41] Dimitra Maoutsa 和 Manfred Opper。《用于约束随机微分方程的确定性粒子流》。发表于：Machine Learning and the Physical Sciences，神经信息处理系统第 35 届会议 (NeurIPS) 研讨会，arXiv 预印本 arXiv:2110.13020 (2021)。DOI：https://doi.org/10.48550/arXiv.2110.13020。[42] Valentin De Bortoli、Arnaud Doucet、Jeremy Heng 和 James Thornton。《用得分匹配模拟扩散桥》。发表于：arXiv 预印本 arXiv:2111.07243 (2021)。[43] M Tabar。《克拉默斯–莫亚尔展开与福克–普朗克方程》。发表于：《复杂非线性动力系统的分析与基于数据的重建》。Springer，2019，第 19–29 页。[44] Rudolf Friedrich、Silke Siegert、Joachim Peinke、Marcus Siefert、Michael Lindemann、Jan Raethjen、Güntner Deuschl、Gerhard Pfister 等。《从实验数据中提取模型方程》。发表于：Physics Letters A 271.3 (2000)，第 217–222 页。[45] Andreas Ruttor、Philipp Batz 和 Manfred Opper。《随机微分方程漂移函数的近似高斯过程推断》。发表于：神经信息处理系统进展 26 (2013)。[46] Arthur P Dempster、Nan M Laird 和 Donald B Rubin。《通过期望最大化算法从不完整数据中进行最大似然估计》。发表于：Journal of the Royal Statistical Society: Series B (Methodological) 39.1 (1977)，第 1–22 页。

[47] Ola Elerian、Siddhartha Chib 和 Neil Shephard。《对离散观测的非线性扩散过程进行似然推断》。发表于：《计量经济学》69.4 (2001)，第 959–993 页。[48] Georgios Arvanitidis、Soren Hauberg、Philipp Hennig 和 Michael Schober。《在从数据中学习的流形上快速、稳健地寻找最短路径》。发表于：第 22 届国际人工智能与统计学会议。PMLR。2019，第 1506–1515 页。[49] Manfred Opper。《随机微分方程的变分推断》。发表于：《物理学年鉴》531.3 (2019)，第 1800233 页。[50] Nina Miolane、Nicolas Guigui、Alice Le Brigant、Johan Mathe、Benjamin Hou、Yann Thanwerdas、Stefan Heyder、Olivier Peltre、Niklas Koep、Hadi Zaatiti 等。《Geomstats：用于机器学习中黎曼几何的 Python 软件包》。发表于：《机器学习研究期刊》21.223 (2020)，第 1–9 页。[51] Stefan Sommer。《几何统计学的概率方法：随机过程、转移分布与纤维丛几何》。发表于：《医学图像分析中的黎曼几何统计学》。爱思唯尔，2020，第 377–416 页。DOI：https://doi.org/10.1016/ B978-0-12-814725-2.00018-2。[52] Charles Fefferman、Sanjoy Mitter 和 Hariharan Narayanan。《流形假设检验》。发表于：《美国数学会杂志》29.4 (2016)，第 983–1049 页。[53] Tal Shnitzer、Ronen Talmon 和 Jean-Jacques Slotine。《用于数据驱动动力系统分析的流形学习》。发表于：《系统与控制中的 Koopman 算子》。Springer，2020，第 359–382 页。[54] Jared L Callaham、J-C Loiseau、Georgios Rigas 和 Steven L Brunton。《利用朗之万回归进行非线性随机建模》。发表于：《英国皇家学会会刊 A 辑》 477.2250 (2021)，第 20210092 页。DOI：https://doi.org/10.1098/rspa.2021.0092。[55] Steven J Lade。《克拉默斯–莫亚尔分析中的有限采样间隔效应》。发表于：Physics Letters A 373.41 (2009)，第 3705–3709 页。[56] Alexander Tong、Jessie Huang、Guy Wolf、David Van Dijk 和 Smita Krishnaswamy。《Trajectorynet：用于建模细胞动力学的动态最优传输网络》。发表于：国际机器学习会议。PMLR。2020，第 9526–9536 页。[57] Tal Shnitzer、Ronen Talmon 和 Jean-Jacques Slotine。《结合收缩观测器的流形学习：用于数据驱动的时间序列分析》。发表于：《IEEE 信号处理汇刊》65.4 (2016)，第 904–918 页。DOI：https://doi.org/10.1109/TSP.2016.2616334。[58] Dimitrios Giannakis。《遍历动力系统的数据驱动谱分解与预测》。发表于：Applied and Computational Harmonic Analysis 47.2 (2019)，第 338–396 页。[59] Nelida ˇCrnjari´c-Žic、Senka Ma´ceši´c 和 Igor Mezi´c。《随机动力系统的 Koopman 算子谱》。发表于：《非线性科学杂志》30.5 (2020)，第 2007–2056 页。[60] Bernard O Koopman 和 J v Neumann。《具有连续谱的动力系统》。发表于：《美国国家科学院院刊》18.3 (1932)，第 255–263 页。[61] Igor Mezi´c。《动力系统的谱性质、模型降阶与分解》。发表于：Nonlinear Dynamics 41.1 (2005)，第 309–325 页。[62] Lars Onsager 和 Stefan Machlup。《涨落与不可逆过程》。发表于：《物理评论》91.6 (1953)，第 1505 页。[63] Takahiko Fujita 和 Shin-ichi Kotani。《扩散过程的昂萨格-马赫卢普函数》。发表于：《京都大学数学杂志》22.1 (1982)，第 115–130 页。[64] Y Takahashi 和 S Watanabe。《扩散过程的概率泛函（昂萨格-马赫卢普函数）》。发表于：《随机积分》。Springer，1981，第 433–463 页。[65] Detlef Dürr 和 Alexander Bach。《作为扩散过程最概然路径之拉格朗日量的昂萨格-马赫卢普函数》。发表于：《数学物理通讯》60.2 (1978)，第 153–170 页。[66] Robert Graham。《一般扩散过程的路径积分表述》。发表于：《物理学杂志 B：凝聚态》26.3 (1977)，第 281–290 页。[67] Hidemi Ito。《扩散过程拉格朗日量的概率构造及其应用》。发表于：《理论物理进展》59.3 (1978)，第 725–741 页。[68] Ruslan Leontievich Stratonovich。《关于扩散过程的概率泛函》。发表于：《数学统计与概率学选译》10 (1971)，第 273–286 页。

[69] Stefan Sommer。《流形上的各向异性分布：模板估计与最可能路径》。载于：医学影像信息处理国际会议。Springer，2015年，第193–204页。[70] Mireille Capitaine。《关于椭圆扩散过程的 Onsager-Machlup 泛函》。载于：《概率研讨会 XXXIV》（2000年），第313–328页。[71] Bernt Øksendal。《随机微分方程》。载于：《随机微分方程》。Springer，2003年，第65–84页。[72] Kazuhisa Tomita 和 Hiroyuki Tomita。《涨落的不可逆环流》。载于：《理论物理进展》51.6（1974年），第1731–1749页。[73] Guido Van Rossum 和 Fred L Drake Jr.《Python教程》。第620卷。荷兰数学与计算机科学研究中心，荷兰阿姆斯特丹，1995年。[74] Charles R. Harris、K. Jarrod Millman、Stéfan J. van der Walt、Ralf Gommers、Pauli Virtanen、David Cournapeau、Eric Wieser、Julian Taylor、Sebastian Berg、Nathaniel J. Smith、Robert Kern、Matti Picus、Stephan Hoyer、Marten H. van Kerkwijk、Matthew Brett、Allan Haldane、Jaime Fernández del Río、Mark Wiebe、Pearu Peterson、Pierre Gérard-Marchant、Kevin Sheppard、Tyler Reddy、Warren Weckesser、Hameer Abbasi、Christoph Gohlke 和 Travis E. Oliphant。《使用 NumPy 进行数组编程》。载于：《自然》585.7825（2020年9月），第357–362页。DOI：10.1038/s41586-020-2649-2。URL：https://doi.org/10.1038/s41586-020-2649-2。[75] Pauli Virtanen、Ralf Gommers、Travis E. Oliphant、Matt Haberland、Tyler Reddy、David Cournapeau、Evgeni Burovski、Pearu Peterson、Warren Weckesser、Jonathan Bright、Stéfan J. van der Walt、Matthew Brett、Joshua Wilson、K. Jarrod Millman、Nikolay Mayorov、Andrew R. J. Nelson、Eric Jones、Robert Kern、Eric Larson、C J Carey、 ̇Ilhan Polat、Yu Feng、Eric W. Moore、Jake VanderPlas、Denis Laxalde、Josef Perktold、Robert Cimrman、Ian Henriksen、E. A. Quintero、Charles R. Harris、Anne M. Archibald、Antônio H. Ribeiro、Fabian Pedregosa、Paul van Mulbregt 和 SciPy 1.0 贡献者。《SciPy 1.0：Python科学计算的基础算法》。载于：《自然·方法》17（2020年），第261–272页。DOI：10.1038/s41592-019-0686-2。[76] J. D. Hunter。《Matplotlib：二维图形环境》。载于：Computing In Science & Engineering 9.3（2007年），第90–95页。DOI：10.1109/MCSE.2007.55。[77] Michael L Waskom。《Seaborn：统计数据可视化》。载于：《开放源代码软件期刊》6.60（2021年），第3021页。[78] Alexander G. de G. Matthews、Mark van der Wilk、Tom Nickson、Keisuke Fujii、Alexis Boukouvalas、Pablo León-Villagrá、Zoubin Ghahramani 和 James Hensman。《GPflow：使用 TensorFlow 的高斯过程库》。载于：《机器学习研究期刊》18.40（2017年4月），第1–6页。URL：http://jmlr.org/papers/v18/16-537.html。[79] Ofir Pele 和 Michael Werman。《快速且稳健的推土机距离》。载于：2009 IEEE 第12届国际计算机视觉会议。IEEE，2009年9月，第460–467页。[80] Adam Paszke、Sam Gross、Soumith Chintala、GregoryChanan、Edward Yang、Zachary DeVito、Zeming Lin、Alban Desmaison、Luca Antiga 和 Adam Lerer。《PyTorch中的自动微分》。载于：（2017年）。[81] Orlando Romero、Sarthak Chatterjee 和 Sérgio Pequito。《通过离散时间李雅普诺夫稳定性理论分析期望最大化算法的收敛性》。载于：2019年美国控制会议（ACC）。IEEE，2019年，第163–168页。[82] Robert S Liptser 和 Albert N Shiryaev。《随机过程统计 II：应用》。第6卷。Springer Science & Business Media，2013年。DOI：10.1007/978-3-662-10028-8。[83] Carl Edward Rasmussen。《机器学习中的高斯过程》。载于：机器学习暑期学校。Springer-Verlag，2003年，第63–71页。[84] Christoph Honisch 和 Rudolf Friedrich。《低采样率下克拉默斯-莫亚尔系数的估计》。载于：Physical Review E 83.6（2011年），第066701页。[85] Raphaël Chetrite 和 Hugo Touchette。《条件过程与驱动过程的变分表示和最优控制表示》。载于：Journal of Statistical Mechanics: Theory and Experiment 2015.12（2015年），P12001。[86] Satya N Majumdar 和 Henri Orland。《受约束随机过程的有效朗之万方程》。载于：Journal of Statistical Mechanics: Theory and Experiment 2015.6（2015年），P06039。

[87] Manfredo Perdigao Do Carmo 和 J Flaherty Francis。《黎曼几何》。第6卷。Springer，1992年。[88] John M Lee。《黎曼流形导论》。第176卷。Springer，2018年。[89] Stephen Wiggins。《动力系统中的正规双曲不变流形》。第105卷。Springer Science & Business Media，1994年。[90] Salah-Eldin A Mohammed 和 Michael KR Scheutzow。《随机微分方程的稳定流形定理》。载于：《概率年刊》（1999年），第615–652页。DOI：https://doi.org/10.1214/aop/1022677380。[91] TV Girya 和 Igor Dmitrievich Chueshov。《随机扰动耗散动力系统的惯性流形与平稳测度》。载于：《数学文集》186.1（1995年），第29–45页。[92] Neil Fenichel 和 JK Moser。《流的不变流形的持久性与光滑性》。载于：《印第安纳大学数学杂志》21.3（1971年），第193–226页。[93] Ludwig Arnold。《作为动力系统的随机微分方程》。载于：《系统理论中的实现与建模》。Springer，1990年，第489–495页。[94] Andrew Carverhill。《随机动力系统的流：遍历理论》。载于：《随机学：概率与随机过程国际期刊》14.4（1985年），第273–317页。[95] Christian Fröhlich、Alexandra Gessner、Philipp Hennig、Bernhard Schölkopf 和 Georgios Arvanitidis。《黎曼数据流形上的贝叶斯求积》。载于：国际机器学习会议。PMLR，2021年，第3459–3468页。DOI：https://doi.org/10.48550/arXiv.2102.06645。[96] Georgios Arvanitidis、Lars Kai Hansen 和 Søren Hauberg。《潜空间奇异性：深度生成模型的曲率》。载于：arXiv预印本 arXiv:1710.11379（2017年）。DOI：https://doi.org/10.48550/arXiv.1710.11379。[97] Tooru Taniguchi 和 EGD Cohen。《非平衡稳态与涨落定理的昂萨格-马赫卢普理论》。载于：《统计物理学杂志》126.1（2007年），第1–41页。[98] Artur B Adib。《扩散动力学的随机作用量：重加权、采样与最小化》。载于：《物理化学杂志 B》112.19（2008年），第5910–5916页。[99] Robert Graham。《非线性非平衡热力学的昂萨格-马赫卢普函数》。载于：《泛函积分》。Springer，1980年，第263–280页。[100] Erlend Grong 和 Stefan Sommer。《Kunita随机微分方程的最可能流》。载于：arXiv预印本 arXiv:2209.03868（2022年）。[101] Dimitra Maoutsa、Sebastian Reich 和 Manfred Opper。《通过梯度对数密度估计求解福克-普朗克方程的相互作用粒子解》。载于：Entropy 22.8（2020年），第802页。

检查清单

1. 对所有作者.....

   (a) 摘要和引言中提出的主要论点是否准确反映了论文的贡献和范围？ [是] 我们提供了数值结果来佐证这些论点。(b) 您是否描述了这项工作的局限性？ [是](c) 您是否讨论了这项工作可能产生的负面社会影响？ [是] 在影响声明中。(d) 您是否已阅读伦理审查指南，并确保论文符合其中的要求？ [是]

2. 如果您包含理论结果.....

   (a) 您是否陈述了所有理论结果的完整假设集合？ [不适用](b) 您是否提供了所有理论结果的完整证明？ [不适用]

3. 如果您进行了实验.....

   (a) 您是否提供了复现主要实验结果所需的代码、数据和说明（可置于补充材料中或通过 URL 提供）？ [否] 尚未提供。本研究将提交归档出版，因此目前我们不会开源代码。不过，我们的工作结合了已发表的框架 [39, 40,

48]。文章 [40, 48] 提供了包含可用实现的 GitHub代码仓库，我们使用的正是这些实现。对于 [39] ，我们按照论文正文和补充材料中的描述重新实现了漂移推断。出于归档考虑，也为了让今后阅读本文终稿的读者能够找到我们的代码，我们将提供一个尚未填充内容的代码仓库链接，待全文准备好提交后，实现在该仓库中。

(b) 您是否说明了所有训练细节（例如数据划分、超参数以及它们的选择方式）？ [是](c) 您是否报告了误差范围（例如，多次运行实验后，针对随机种子报告的误差范围）？ [不适用](d) 您是否说明了计算资源总量和所用资源类型（例如 GPU 类型、内部集群或云服务提供商）？ [是] 在配备英特尔酷睿 i7@ 1.80GHz 中央处理器的笔记本电脑上，我们的框架针对二维系统运行一次需要约 4–6 小时（具体取决于数据量）。

4. 如果您使用了现有资源（例如代码、数据、模型），或整理/发布了新资源 ...
(a) 如果您的工作使用了现有资源，是否引用了创建者？ [是](b) 您是否提及了这些资源的许可证？ [不适用](c) 您是否在补充材料中或通过 URL 提供了任何新资源？ [否]尚未。我们将在论文提交以进行归档出版后发布该框架。(d) 您是否讨论了是否以及如何获得使用或整理的数据所涉及人员的同意？ [不适用](e) 您是否讨论了所使用或整理的数据是否包含个人身份信息或冒犯性内容？ [不适用]

5. 如果您使用了众包或开展了涉及人类受试者的研究.....
(a) 是否包含向参与者提供的完整说明文本以及截图（如适用）？ [不适用](b) 是否描述了参与者可能面临的风险，并在适用时附上机构审查委员会 (IRB) 的批准链接？ [不适用](c) 是否列明了支付给参与者的预估时薪以及参与者报酬支出的总金额？ [不适用]

附录

A 高频和低频观测的漂移推断1 **3**

A.1 路径的近似后验.15A.1.1 不具有几何约束的路径近似后验. . . . . 15A.1.2 具有几何约束的路径近似后验. 17A.2 漂移函数的近似后验.19

——

B 将测地线用作几何约束的理论依据19

C 所提出的方法是否会破坏扩散过程的马尔可夫性质?
cess? **20**

D 数值实验详情 **21**

高频和低频观测的漂移推断

我们考虑其演化由随机微分方程方程（1）描述的系统。

高频观测。当系统路径 $X_{0:T}$ 在连续时间内被观测时，扩散过程在连续观测之间的无穷小转移概率服从高斯分布，即:

$$\mathrm{P}_f(X_{0:T} \mid f) \propto \exp\left(-\frac{1}{2\mathrm{d}t}\sum_t \|X_{t+\mathrm{d}t} - X_t - f(X_t)\mathrm{d}t\|^2_{\sigma^2}\right). \tag{5}$$

反过来，离散化维纳路径 $P_{\mathcal{W}}(X_{0:T})$（即无漂移过程的路径）的转移概率可表示为:

$$\mathrm{P}_{\mathcal{W}}(X_{0:T}) = \exp\left(-\frac{1}{2\mathrm{d}t}\sum_t \|X_{t+\mathrm{d}t} - X_t\|^2_{\sigma^2}\right), \tag{6}$$

其中， $\|u\|_{\sigma^2} \doteq u^\top \cdot \sigma^{-2} u$ 表示加权范数， $D \doteq \sigma^2$ 表示噪声协方差。因此，我们可以通过 $P_f(X_{0:T}|f)$与路径 $X_{0:T}$在时间区间 $[0, T]$ [82]内的 $\mathrm{P}_{\mathcal{W}}(X_{0:T})$之间的拉东–尼科迪姆导数，来表示漂移 f的似然。

$$\mathcal{L}(X_{0:T} \mid f) = \exp\left[-\frac{1}{2}\sum_t \|f(X_t)\|^2_{\sigma^2}\mathrm{d}t + \sum_t \langle f(X_t), X_{t+\mathrm{d}t} - X_t\rangle_{\sigma^2}\right], \tag{7}$$

为简洁起见，我们引入记号 $\langle u, v\rangle \doteq u^\top \cdot \sigma^{-2} v$ 表示关于逆噪声协方差 σ^{-2}的加权内积。该表达式是将吉尔萨诺夫定理应用于由具有漂移 f 的过程和维纳过程所诱导的路径测度（两者具有相同的扩散 σ），并对连续路径 $X_{0:T}$采用欧拉–丸山离散化后得到的。

对随机微分方程（式（7））连续观测路径的似然关于漂移函数呈二次型。因此，对函数值施加高斯测度（高斯过程）是该似然的一种自然共轭先验。为以非参数形式识别漂移，我们假设函数值 $f \sim \mathrm{P}_0(f) = \mathrm{GP}(m^f, k^f)$服从高斯过程先验，其中 m^f和 k^f分别表示高斯过程 [45]的均值函数和协方差函数。先验测度可写为

$$\mathrm{P}_0(f) = \exp\left[-\frac{1}{2}\int\int f(x)\left(k^f(x,x')\right)^{-1} f(x')\mathrm{d}x\mathrm{d}x'\right], \tag{8}$$

如果我们考虑一个零均值高斯过程 $m^f = 0$。

对漂移函数 f 进行贝叶斯推断，需要计算函数空间中的一个概率分布，即后验概率分布 $\mathrm{P}_f(f \mid X_{0:T})$。根据贝叶斯定理，后验分布可表示为

$$\mathrm{P}_f(f \mid X_{0:T}) = \frac{\mathrm{P}_0(f)\mathcal{L}(X_{0:T} \mid f)}{Z} \propto \mathrm{P}_0(f)\mathcal{L}(X_{0:T} \mid f), \tag{9}$$

其中 Z 表示一个定义为路径积分的归一化因子

$$Z = \int \mathrm{P}_0(f)\mathcal{L}(X_{0:T} \mid f)\mathcal{D}f, \tag{10}$$

其中 $\mathcal{D}f$ 表示在希尔伯特空间 f 上的积分： $H_0[f] < \infty$。这里，我们将函数的先验概率表示为 $\mathrm{P}_0(f) = e^{-H_0[f]}$。在 [45] 文中，作者表明，在连续时间极限下，漂移函数的非参数估计等价于高斯过程回归，其目标是确定从系统状态 X_t 到状态增量 $\mathrm{d}X_t$ [83]的映射。更准确地说，我们考虑将系统状态的观测值 N 作为回归变量 X_t，相应的响应变量为

$$Y_t = \frac{X_{t+\mathrm{d}t} - X_t}{\mathrm{d}t}, \tag{11}$$

并将高斯过程的核函数记为 $k(x, x')$。

若用 $\mathcal{X} = \{X_t\}_{t=0}^{T-\mathrm{d}t}$和 $\mathcal{Y} = \{Y_t\}^{T-}{}_{\mathrm{d}_{t=0}}{}^{t}$分别表示状态观测值和观测增量，则漂移函数 f 的后验过程均值可表示为

$$\bar{f}(x) = k^f(x, \mathcal{X})^\top \left(\mathcal{K} + \frac{\sigma^2}{\mathrm{d}t} I_N\right)^{-1} \mathcal{Y}, \tag{12}$$

这里我们滥用了符号，将 $k^f(x, \mathcal{X})$记作在点 x和 $\{\mathcal{O}_t\}_{k=1}^{K-1}$处计算核函数 k^f所得的向量。类似地， $\mathcal{K} = k^f(\mathcal{X}, \mathcal{X})$表示在所有观测值对上计算核函数所得的 $(K-1) \times (K-1)$矩阵。类似地，后验方差可写为

$$\Sigma^2(x) = k^f(x, x) - k^f(x, \mathcal{X})^\top \left(\mathcal{K} + \frac{\sigma^2}{\mathrm{d}t}\right)^{-1} k^f(x, \mathcal{X}), \tag{13}$$

其中，项 $\sigma^2/\mathrm{d}t$ 起到观测噪声的作用。

低频观测。 当观测间隔变大时（低频观测），公式（7）中的高斯似然便不再有效，因为在观测间隔较大时，转移密度不再服从高斯分布。因此，基于高斯假设的漂移估计 [26, 45] 会变得不准确。为缓解这一问题，Lade [55] 提出了一种计算漂移估计有限时间修正的方法，据我们所知，该方法主要应用于一维问题 [84]。另一方面，统计学界提出了路径增广方案，通过在观测值之间对简化系统的动力学进行采样，将观测轨迹增广为近乎连续时间的轨迹 [31–35]。然而，对于较大的观测间隔和非线性系统，用于路径增广的简化动力学与底层路径统计往往不匹配，因此这些方法的收敛速度较差，或无法识别出正确的动力学（图 2 c 和 d）。我们在此指出，使用奥恩斯坦 - 乌伦贝克桥进行路径增广，并将正确动力学的局部线性化作为漂移，能够很好地近似底层转移密度。然而，在推断过程中，真实的底层动力学未知，因此基于不准确漂移估计提出的局部线性化 [39] 在低频观测下表现不佳。

请注意，随着观测间隔 τ增大，如果系统是非线性的，或噪声依赖于状态，那么假设两个连续观测值之间服从高斯分布的高斯似然便不再成立。在此类情形下，漂移的似然可以用路径积分表示。

$$\mathrm{P}(\mathcal{O}_{1:K} \mid f) = \int \mathrm{P}(\mathcal{O}_{1:K} \mid X_{0:T})\mathrm{P}(X_{0:T} \mid f)\mathcal{D}(X_{0:T}), \tag{14}$$

其中， $\mathcal{O}_{1:K} \doteq \{\mathcal{O}_k\}_{k=1}^{K}$ 表示 K 离散时间观测值，$\mathrm{P}(X_{0:T} \mid f)$表示由漂移为 $f(x)$的扩散过程产生的先验路径概率， $\mathcal{D}(X_{0:T})$表示形式体积

元，它定义在路径空间上；$\mathrm{P}(\mathcal{O}_{1:K} \mid X_{0:T})$表示给定潜在路径 $X_{0:T}$时观测值的似然。

然而，对于非线性系统，式（14）中的路径积分难以求解，因此我们需要同时估计扩散过程的漂移和潜在状态，即近似潜在路径与漂移函数的联合后验测度 $\mathrm{P}(X_{0:T}, f \mid \mathcal{O}_{1:K})$。因此，我们将未观测到的连续路径 $X_{0:T}$ 视为潜在随机变量，并采用期望最大化（EM）算法来求得漂移函数的最大后验估计。更具体地说，我们采用迭代算法，在每次迭代 n 中交替执行以下两个步骤：

一个**期望**步骤：给定漂移的估计值 $f^n(x)$，我们构造潜变量的近似后验分布 $\hat{}Q(X_{0:T}) \approx \mathrm{P}(X_{0:T} \mid \mathcal{O}_{1:K}, f^n(x))$，并计算增广路径的期望对数似然

$$\mathfrak{L}\big(\hat{f}^n(x), Q\big) = \mathbb{E}_Q\Big[\ln \mathscr{L}\big(X_{0:T} \mid \hat{f}^n(x)\big)\Big]. \tag{15}$$

一个**最大化**步骤：通过最大化期望对数似然来更新漂移估计

$$f^{n+1}(x) = \arg\max_f \Big[\mathfrak{L}\big(f^n(x), Q\big) - \ln \mathrm{P}_0\big(f^n(x)\big)\Big]. \tag{16}$$

在式（16）中，P_0 表示函数值的高斯过程先验。

A.1 路径的近似后验。

这里，我们首先仅将各个观测值视为约束（第A.1.1节），以此构建路径的近似后验（给定观测值时路径的条件分布）。然而，这种方法会使路径增广过程中的计算量变得很大，因为相对于最初估计的漂移而言，观测值属于非典型状态。为解决这一问题，我们随后扩展这一形式体系（第A.1.2节），纳入同时考虑观测值局部几何的约束。

A.1.1 无几何约束时路径的近似后验。

$\hat{}$ 给定漂移函数（或漂移估计） $f(x)$，我们可以运用变分方法来近似在观测值 $\mathcal{O}_{1:K}$ 条件下潜在路径的后验测度。我们认为，先验过程（不考虑观测值 $\mathcal{O}_{1:K}$ 时的过程）由下式描述

$$\mathrm{P}(X_{0:T} \mid \hat{f}): \qquad \mathrm{d}X_t = \hat{f}(X_t)\mathrm{d}t + \sigma \mathrm{d}\boldsymbol{\beta}_t. \tag{17}$$

我们将定义一个以观测值为条件的近似（后验）过程。该条件过程也是一个扩散过程，其扩散项与式（17）相同，但具有经过修正的时变漂移 $g(x,\ t)$，以考虑观测值 [85, 86]。我们将近似后验测度 Q 视为由一个以观测值为条件的近似过程所诱导的后验测度 $\mathcal{O}_{1:K}$ [49]，其控制方程为

$$Q(X_{0:T}): \qquad \mathrm{d}X_t = g(X_t, t)\mathrm{d}t + \sigma \mathrm{d}\boldsymbol{\beta}_t = \Big(\hat{f}(X_t) + \sigma^2 u(X_t, t)\Big)\,\mathrm{d}t + \sigma \mathrm{d}\boldsymbol{\beta}_t. \tag{18}$$

该 有效漂移 $g(X_t, t)$ 方程（18）的解可由变分问题得到
最小化自由能

$$\mathcal{F}[Q] = \mathcal{KL}\Big(Q(X_{0:T}) || \mathrm{P}(X_{0:T} \mid \hat{f})\Big) - \sum_{k=1}^{K} E_Q[\ln \mathrm{P}(\mathcal{O}_k \mid X_{t_k})]. \tag{19}$$

应用卡梅伦-吉尔桑诺夫-马丁定理，我们可以将漂移分别为 $f(x)$ 和 $g(x,t)$ 的扩散过程所诱导的两个路径测度之间的 Kullback-Leibler 散度 ˆ 表示为

$$\mathcal{KL}\Big(Q(X_{0:T})||\mathrm{P}(X_{0:T}|\hat{f})\Big) = E_Q\left[\ln\left(\frac{dQ(X_{0:T})}{dP\left(X_{0:T}|\hat{f}\right)}\right)\right] \tag{20}$$

$$= E_Q\left[\left(-\frac{1}{2}\int_0^T \|\hat{f}(X_t) - g(X_t,t)\|^2_{\sigma^2}\mathrm{d}t + \int_0^T \frac{\hat{f}(X_t) - g(X_t,t)}{\sigma^2}\mathrm{d}\boldsymbol{\beta}_t\right)\right] \tag{21}$$

$$= E_Q\left[\left(-\frac{1}{2}\int_0^T \|\hat{f}(X_t) - g(X_t,t)\|^2_{\sigma^2}\mathrm{d}t + V_T\right)\right] \tag{22}$$

$$= \frac{1}{2}\int_0^T\int \|g(x,t) - \hat{f}(x)\|^2_{\sigma^2}\, q_t(x)\,\mathrm{d}x\,\mathrm{d}t + \mathfrak{C}, \tag{23}$$

其中，$q_t(x)$表示近似过程中 X_t 的边缘密度。在第三行中，我们引入了随机变量 $V_T = \int_0^T$ ˆ$f(X_t)-g(X_t,t)$ σ^2 $\mathrm{d}\boldsymbol{\beta}_t$。假设函数 $\ell(X_t)$ =ˆ$f(X_t)$ ~~$-g(X_t,t)$有界~~、分段连续，且属于 $L^2[0,\infty)$，则 V_T服从分布 $\mathcal{N}\left(V_T \mid 0, \int_0^T \ell^2(s)\mathrm{d}s\right)$；对于给定的 T，这将得到一个常数 $\mathfrak{C}$。因此，式（23）中的第二项与自由能的最小化无关，故将其省略。

因此，我们可以将式（19）中的自由能表示为 [49]

$$\mathcal{F}[Q] = \frac{1}{2}\int_0^T\int \Big[\|g(x,t) - \hat{f}(x)\|^2_{\sigma^2} + U(x,t)\Big]\, q_t(x)\,\mathrm{d}x\,\mathrm{d}t, \tag{24}$$

其中，项 $U(x,\ t)$计入了观测值 $U(x,\ t) = -\sum \ln \mathrm{P}(\mathcal{O}_k \mid x)\delta(t-t_k)$。$_{t_k}$

自由能泛函的最小化可视为一个随机控制问题 [49]，其目标是为漂移为 $f(x)$ 的系统确定时变漂移调整 $u(x,t) := g(x,t) - f(x)$ ˆ，使受控动力学满足观测值施加的约束。

对于精确观测的情形，即观测过程为 $\psi(x) = x$ 时，我们可以分别计算每个 $K-1$ 观测间隔的漂移调整。因此，对于连续观测之间的每个间隔，我们确定构造随机桥所需的最优控制 $u(x,t)$，使其遵循方程（17）的动力学，并以相应的观测值作为初始状态和终端状态 $\mathcal{O}_k$ 和 $\mathcal{O}_{k+1}$。

对于 $\mathcal{O}_k$ 与 $\mathcal{O}_{k+1}$ 之间观测间隔上的此类随机控制问题，其最优漂移调整可由后向方程的解得到（见 [40, 41]）

$$\frac{\partial \phi_t(x)}{\partial t} = -\mathcal{L}^\dagger_{\hat{f}}\phi_t(x) + U(x,t)\phi_t(x), \tag{25}$$

终端条件为 $\phi_T(x) = \chi(x) = \delta(x - \mathcal{O}_{k+1})$，其中 $\mathcal{L}^\dagger$ ˆf 表示方程（17）中过程的伴随福克–普朗克算子。正如 Maoutsa 等人 [40, 41] 所示，最优漂移调整 $u(x,t)$ 可以表示为两个概率流的对数梯度之差

$$u^*(x,t) = D\Big(\nabla \ln q_{T-t}(x) - \nabla \ln \rho_t(x)\Big), \tag{26}$$

其中 ρ_t 满足前向（滤波）偏微分方程（偏微分方程）

$$\frac{\partial \rho_t(x)}{\partial t} = \mathcal{L}_{\hat{f}}\rho_t(x) - U(x,t)\rho_t(x), \tag{27}$$

而 q_t 是一个时间反向偏微分方程的解，该方程取决于 $\rho_t(x$的对数梯度)

$$\frac{\partial q_t(x)}{\partial t} = -\nabla \cdot \left[\Big(\sigma^2 \nabla \ln \rho_{T-t}(x) - f(x, T-t) \Big) q_t(x) \right] + \frac{\sigma^2}{2} \nabla^2 q_t(x), \tag{28}$$

初始条件为 $q_0(x) \propto \rho_T(x)\chi(x)$ 。

A.1.2 具有几何约束的路径的近似后验。

前文根据随机桥构造近似测度的方法，适用于在先验扩散过程的概率律下观测值具有非零概率的情况ˆ方程（17）所示的过程。然而，当先验过程（其估计漂移为 f）与生成观测值的过程相差较大时，这种构造方法可能无法很好地近似底层路径测度，也可能导致扩散桥的数值构造收敛缓慢。为解决这一问题，我们在此为后验过程引入额外约束，促使后验测度中的路径遵循观测值的局部几何结构。下文将简要介绍黎曼几何的基础知识，随后继续探讨所提方法中的几何问题。

黎曼几何。 一个 d维 **黎曼流形** [87, 88] $(\mathcal{M}, h)$ 嵌入于 D维环境空间 $\mathcal{X} = \mathcal{R}^D$ 中，是一个光滑弯曲的 d维曲面，并配备了平滑变化的内积（黎曼）**度量** h： $x \to \langle \cdot | \cdot \rangle_x$ 定义在 $\mathcal{T}_x\mathcal{M}$ 上。每个点 $\mathcal{T}_x\mathcal{M}$ 处都定义了一个切空间 $x \in \mathcal{M}$。黎曼度量 h 在流形 $\mathcal{M}$上定义了一个规范体积测度。直观而言，它刻画了如何在流形切空间中的点之间计算局部内积 $\mathcal{M}$，因此也决定了如何计算范数，进而确定 $\mathcal{M}$上的点之间的距离。

一个**坐标图** (G, ϕ)提供了从 G 上的开集到欧几里得空间中开集的映射。如果对每个点 $x \in \mathcal{M}$ ，都存在一个局部邻域 $G \subset \mathcal{R}^d$，则流形的维数为 d 。我们可以通过以下方式表示局部坐标图(G, ϕ)上的度量 h

正定矩阵（**度量张量**） $H(x) = (h_{i,j})_{x, 0 \le i,j, \le d} = \left(\langle \frac{\partial}{\partial x_i} | \frac{\partial}{\partial x_j} \rangle_x \right)_{0 \le i, j, \le d}$ ，在每个

点 $x \in G$。

对于 $v, w \in \mathcal{T}_x\mathcal{M}$ 和 $x \in G$，它们的内积可以表示为切空间 $\mathcal{T}_x\mathcal{M}$ 上度量 h 的矩阵表示形式 $\langle v | w \rangle_x = v^\top H(x) w$，其中 $H(x) \in \mathcal{R}^{d \times d}$ 。

曲线**的长度** γ 定义为：流形上的切向量范数的积分 $[0, 1] \to \mathcal{M}$

$$\ell(\gamma_{t'}) = \int_0^1 \|\dot{\gamma}_{t'}\|_h \mathrm{d}t' = \int_0^1 \sqrt{\dot{\gamma}_{t'}^\top H(\gamma_{t'}) \dot{\gamma}_{t'}} \mathrm{d}t', \tag{29}$$

whe其中，带点的字母表示曲线的速度 $\dot{\gamma}_{t'} = \partial_{t'} \gamma_{t'}$。**测地曲线**是一条局部 y leng最短光滑曲线，它连接流形上的两个给定点。

观测值的黎曼几何。为了近似路径的后验分布，我们考虑由观测值所体现的不变密度的几何结构。为此，我们考虑这样一类系统：其动力学会产生包含系统全局吸引子的惯性不变流形，且系统轨迹集中于这些流形上 [89–94]。因此，我们假设底层系统的连续时间轨迹 $X_{0:T} \in \mathcal{R}^d$ 集中在一个维数为 m （可能）小于 d 的不变流形 $\mathcal{M} \in \mathcal{R}^{m \le d}$ 上。离散时间观测值$\mathcal{O}_k$ 因而是流形 $\mathcal{M}$ 上的样本。我们方法的核心前提是，连续两次观测之间的**未观测路径要么位于**流形上，**要么**位于流形附近$\mathcal{M}$。具体而言，我们假设未观测路径应**位于连接相邻观测值的测地线附近**，并处于 $\mathcal{M}$ 上。为此，我们提出一种路径增广框架，通过约束增广路径，使其位于所识别的相邻观测值之间的测地线附近。

然而，尽管将低维流形嵌入高维环境空间这一视角有助于我们直观理解所提出的方法，但出于计算目的，我们采用一种互补视角，该视角受到 [95] 中讨论的启发。根据这一视角，我们将整个观测空间 $\mathcal{R}^d$ 视为一个光滑的黎曼流形 $\mathcal{M} \doteq \mathcal{R}^d$，其特征由一个

黎曼度量 h 描述。观测值的非线性几何效应由度量 h 捕捉。因此，为了近似系统不变密度的几何结构，我们学习黎曼度量张量 H： $\mathcal{R}^d \to \mathcal{R}^{d\times d}$，并根据学习得到的度量计算连续观测值之间的测地线。直观地说，按照这一视角，观测值 $\{\mathcal{O}_k\}_{k=1}^K$ 会使我们在状态空间上计算距离的方式产生扭曲。

实际上，这种方法并未降低我们所处理空间的维数，而是改变了我们计算 $\mathcal{M}$上的内积，以及由此得到的距离、长度和测地曲线的方式。另一种思路是在低维流形上进行操作，但这高度依赖于对该流形维数的准确判断。例如，可以使用变分自编码器来近似观测流形，随后从解码器所实现的流形嵌入中获得黎曼度量。然而，我们对此类方法的初步结果表明，该方法需要经过大量微调，才能适应每个动力系统的特性，并且对近似流形维数的估计十分敏感。

为了学习黎曼度量并计算测地线，我们采用 Arvanitidis 等人在 [48]中提出的框架。具体而言，对于状态空间中的位置 $\mathbf{x}$，我们用观测值上计算得到的加权局部对角协方差的逆，以非参数形式近似观测所诱导的局部度量： [48]

$$H_{dd}(\mathbf{x}) = \left(\sum_{i=1}^{K} w_i(\mathbf{x}) \left(x_i^{(d)} - x^{(d)} \right)^2 + \epsilon \right)^{-1}, \tag{30}$$

权重为 $w_i(\mathbf{x}) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}\|_2^2}{2\sigma_{\mathcal{M}}^2} \right)$，其中 $x^{(d)}$ 表示向量 $\mathbf{x}$ 的第 d 维分量。参数 $\epsilon > 0$ 确保加权协方差矩阵的对角元素非零，而 $\sigma_{\mathcal{M}}$ 则刻画流形的曲率。

对于每个区间 $[\mathcal{O}_k, \mathcal{O}_{k+1}]$ 内相邻的观测值，我们将测地线 $\gamma_{t'}^k$ 定义为能量最小化曲线，即动能泛函 $\mathcal{E}(\gamma_{t'}^k) = \int_0^1 L_{\mathcal{M}}(\gamma_{t'}^k, \dot{\gamma}_{t'}^k)\mathrm{d}t'$ 的极小化曲线

$$\gamma_{t'}^{k*} = \underset{\gamma_{t'}^k, \gamma_0^k = \mathcal{O}_k, \gamma_1^k = \mathcal{O}_{k+1}}{\arg\min} \int_0^1 L_{\mathcal{M}}(\gamma_{t'}^k, \dot{\gamma}_{t'}^k)\mathrm{d}t',$$

$$\text{with} \quad \int_0^1 L_{\mathcal{M}}(\gamma_{t'}^k, \dot{\gamma}_{t'}^k)\mathrm{d}t' = \frac{1}{2}\int_0^1 \|\dot{\gamma}_{t'}^k\|_h^2, \tag{31}$$

其中 $L_{\mathcal{M}}(\gamma_{t'}^k, \dot{\gamma}_{t'}^k)$ 表示拉格朗日量。该泛函的极小化曲线与曲线长度泛函 $\ell(\gamma_{t'})$ (式（29）) 的极小化曲线相同，即测地线 [87]。

通过应用变分法，泛函 $\mathcal{E}(\gamma_t^k)$ 的极小曲线可由欧拉–拉格朗日方程求得，从而得到以下二阶微分方程系统 [87, 96]

$$\ddot{\gamma}_t^{\,k} = -\frac{1}{2} H(\gamma_t^k)^{-1} \left(2\left(I \otimes (\dot{\gamma}_t^{\,k})^\top\right) \frac{\partial \mathrm{vec}[H(\gamma_t^k)]}{\partial \gamma_t^k} \dot{\gamma}_t^{\,k} - \frac{\partial \mathrm{vec}[H(\gamma_t^k)]^\top}{\partial \gamma_t^k} \left(\dot{\gamma}_t^{\,k} \otimes \dot{\gamma}_t^{\,k} \right) \right), \tag{32}$$

满足边界条件 $\gamma_0^k = \mathcal{O}_k$ 和 $\gamma_1^k = \mathcal{O}_{k+1}$，其中 $\otimes$ 表示克罗内克积，而 $\mathrm{vec}[A]$ 表示通过堆叠矩阵 A 的列 A 将其转换为向量的向量化操作。Arvanitidis 等人 [48] 通过使用概率微分方程求解器近似求解式（32）的边值问题，获得了测地线。

扩展自由能泛函。我们将各条测地线组成的集合记为$\Gamma_t \doteq \{\gamma_{t'}^k\}_{t=(k-1)\tau+t'\tau}$，其中 $\gamma_{t'}^k$ 是连接 $\mathcal{O}_k$ 和 $\mathcal{O}_{k+1}$的测地线，而 $t' \in [0,1]$ 表示经过缩放的时间变量。除了前文所述设定（第A.1.1节）中的约束外，我们还在自由能 $U_{\mathcal{G}}(x,\ t) \doteq \|\Gamma_t - x\|^2$ 中添加一项，用以体现不变密度的局部几何，并引导潜在路径趋向连接相邻观测值的测地曲线$\gamma_{t'}^k$

$$\mathcal{F}[Q] = \frac{1}{2}\int_0^T \int \left[\|g(x,t) - \hat{f}(x)\|_{\sigma^2}^2 + U_{\mathcal{O}}(x,t) + \beta U_{\mathcal{G}}(x,t) \right] q_t(x)\, dx\, \mathrm{d}t. \tag{33}$$

这里我们将观测项记为 $U_{\mathcal{O}}(x,\ t) \doteq -\sum_{t_k} \ln \mathrm{P}(\mathcal{O}_k|x)\delta(t-t_k)$，而 β 表示一个加权常数，用于确定几何项在控制目标中的相对权重。

根据 [49], ，对于每个观测间隔 $[\mathcal{O}_k, \mathcal{O}_{k+1}]$，我们通过求解随机最优控制问题 [40] 来确定后验路径测度（式 (33) 的极小化解），其目标是ˆ ˆ获得时变漂移调整 $u(x,t) := g(x,t) - f(x)$，用于漂移为 $f(x)$ 的系统；该系统的初始和终端约束由 $U_{\mathcal{O}}(x,t)$ 定义，另有路径约束 $U_{\mathcal{G}}(x,t)$。

A.2 漂移函数的近似后验分布。

对于固定的路径测度 Q，漂移 Q_f 的最优测度是一个高斯过程，表示为

$$Q_f \propto \mathrm{P}_f \exp\left(-\frac{1}{2}\int \|f(x)\|^2_{\sigma^2} A(x) - 2\langle f(x), B(x)\rangle_{\sigma^2} \mathrm{d}x\right), \tag{34}$$

with

$$A(x) \doteq \int_0^T p_t(x)\mathrm{d}t,$$

and

$$B(x) \doteq \int_0^T p_t(x) g(x,t)\mathrm{d}t,$$

其中，$p_t(x)$ 表示状态 X_t 的边缘约束密度。函数 $g(x,\ t)$ 表示有效漂移。

我们假设未知函数 f 服从高斯过程先验，即 $f \sim \mathrm{P}_0(f) = \mathrm{GP}(m^f,\ k^f)$，其中 m^f 和 k^f 分别表示高斯过程的均值和协方差函数。根据 Ruttor等人 [45], ，我们通过在一组稀疏的 S 诱导点 $\{Z_i\}_{i=1}^S$ 上优化函数值，对漂移 f 采用稀疏核近似。由此得到的漂移为

$$\hat{f}_S(x) = k^f(x,\mathcal{Z})\left(I + \Lambda\,\mathcal{K}_S\right)^{-1}\mathbf{d}, \tag{35}$$

其中，我们定义并引入了记号 $\mathcal{K}_S \doteq k^f(\mathcal{Z},\ \mathcal{Z})$

$$\Lambda = \frac{1}{\sigma^2}\mathcal{K}_S^{-1}\left(\int k^f(\mathcal{Z},x)A(x)k^f(x,\mathcal{Z})\mathrm{d}x\right)\mathcal{K}_S^{-1}. \tag{36}$$

$$\mathbf{d} = \frac{1}{\sigma^2}\mathcal{K}_S^{-1}\left(\int k^f(\mathcal{Z},x)B(x)\mathrm{d}x\right)\mathcal{K}_S^{-1}, \tag{37}$$

相应的方差同样可由下式得出

$$\Sigma_S^2(x) = k^f(x,x) - k^f(x,\mathcal{Z})\left(I + \Lambda\,\mathcal{K}_S\right)^{-1}\Lambda\,k^f(\mathcal{Z},x). \tag{38}$$

我们采用基于样本的方法来近似式（34）中的密度，这些密度由路径测度 Q 的粒子采样得到。因此，通过用样本表示密度，我们可以将密度 p_t（x）重写为以粒子位置为中心的狄拉克δ函数之和

$$p_t(x) \approx \frac{1}{N}\sum_{j=1}^{N}\delta(x - X_j(t)),$$

并将黎曼积分替换为对粒子的求和。这里，$X_j(t)$ 表示时间点 t 处第 j 个粒子的位置。

B 支持将测地线用作几何约束的理论依据

在理论物理中，扩散过程的 Onsager-Machlup 泛函一直被视为表征过程在两个预先定义状态之间最概然路径（MPP）的量。在 [62],Onsager 和 Machlup 利用扩散过程的热涨落证明，概率

路径的密度 $\gamma \in C^1\left([0,T],\ \mathcal{R}^d\right)$ 在 $\mathcal{R}^d$ 的有限区间内可表示为玻尔兹曼因子

$$\mathrm{P}(\gamma) \sim \exp\left[-\int_0^T L(\gamma(t),\dot{\gamma}(t))dt\right], \tag{39}$$

其中

$$L(\gamma(t),\dot{\gamma}(t)) = \frac{1}{2}\|\frac{\dot{\gamma}(t)-f(\gamma(t))}{\sigma}\|^2 + \frac{1}{2}\nabla\cdot f(\gamma(t)).^{4} \tag{40}$$

函数 $L(\gamma(t)\ \dot{\gamma}(t))$ 称为 **昂萨格-马赫卢普** 函数（作用量），其对时间的积分则称为昂萨格-马赫卢普作用泛函。它已被用作欧拉-拉格朗日极小化方法中的拉格朗日量，以确定扩散过程在状态空间中两个给定点之间的**最概然路径（MPP）** [66, 68]。

斯特拉托诺维奇 [68] 考察了多维扩散过程的样本处于状态空间中一条理想化光滑路径附近（即位于该路径周围无限薄的管状区域内）的概率。为计算这一概率，他构造了一个概率泛函，该泛函与作为扩散过程拉格朗日量的 Onsager-Machlup 泛函相同。Duerr 等人 [65] 考察了标量扩散过程，并利用吉尔萨诺夫变换，从路径起点状态与终点状态之间转移概率的渐近极限构造了昂萨格-马赫卢普函数。

考虑定义在配有黎曼度量 g 的黎曼流形 $(\mathcal{M},\ g)$ 上的布朗运动，昂萨格-马赫卢普泛函可表示为对拉格朗日量 [64, 99, 100]的积分

$$L(\gamma,\dot{\gamma}) = \frac{1}{2}\|\dot{\gamma}(t)\|_g^2 - \frac{1}{12}S(\gamma(t)), \tag{41}$$

其中 $\|\cdot\|_g$ 表示相对于度量 g，流形 $\mathcal{T}_X\mathcal{M}$ 的切空间上的黎曼范数，而 $S(\cdot)$ 表示流形在各点处的标量曲率。第一项是用于确定流形上测地曲线的拉格朗日量（参见式（A.1.2））

在我们提出的形式体系中，出于计算目的，我们假设整个 $\mathcal{R}^d$ 是一个光滑流形。我们可以将式（41）的第一项视为我们为计算流形$(\mathcal{R}^d,\ g)$上的测地线而优化的拉格朗日量，其中 g 是从观测值中学习得到的度量。

然而，我们观测到的系统是在 $\mathcal{R}^d$ 中定义的扩散过程，采用欧几里得度量。构建一种路径增广方案，引导增广路径趋向于相对于不同度量定义的扩散过程的测地线，这引发了对我们方法有效性的疑问。这里需要指出的是，具有一般状态依赖扩散系数$\sigma \in \mathcal{R}^{d\times m}$的扩散过程，以及 m维布朗运动，都可以被视为在流形$(\mathcal{R}^d,\ g)$上演化，并具有相应的度量 $g = \left(\sigma\sigma^\top\right)^{-1}$ [70]。因此，通过应用类似于逆兰佩蒂变换 [71] 的变换，将我们学习到的随机微分方程转换为一个因状态依赖扩散而产生所学习度量的随机微分方程，或许可以将从数据中学习到的度量与状态依赖扩散所对应的度量关联起来。此类变换的存在将为所提出的方法提供依据。我们的实证结果表明，此类变换或许是可能的。

C 所提出的方法是否会破坏扩散过程的马尔可夫性质?

所提出的路径增广似乎破坏了扩散过程的马尔可夫性质。根据式（1）中扩散的马尔可夫性质，系统状态 $X_{k\tau+\delta t}$应仅取决于状态 $X_{k\tau}$，即观测值 $\mathcal{O}_k$。所提出的增广使状态 $X_{k\tau+\delta t}$ 不仅取决于下一个观测值 $\mathcal{O}_{k+1} = X_{(k+1)\tau}$，还取决于这些观测值附近的过去和未来状态。

我们有效地构造增广路径，以计算漂移估计的似然。为计算该似然，我们需要评估连续观测值之间的转移概率。

[4]昂萨格和马赫卢普的早期工作主要关注线性过程，因此最初引入的泛函不包含散度的第二项 f，因为对于线性f而言，该项是常数。后来，为了计入轨迹熵修正，该项被加入 OM 函数中 [97, 98]

由于一般非线性系统的转移概率通常难以解析求得，因此我们必须采用数值近似。理想情况下，我们会使用桥采样器来近似转移密度，该采样器会考虑条件于连续观测值、且通过这些观测值的非线性估计随机微分方程。然而，对于较粗略的漂移估计，在估计随机微分方程的概率律下，观测值的概率为零；构造这些桥要么计算代价极高，要么根本无法完成。相反，我们在此计算“校正后”估计的似然（该校正源自不变密度），在这一估计下观测值具有非零概率，随后再基于这一“校正后”估计，在增广路径上重新估计漂移。通过考虑观测值的局部几何，我们对错误估计的漂移函数进行系统性校正，从而生成增广路径。这实际上通过路径约束推动增广过程趋向每个观测间隔的第二个观测值；该约束将增广路径推向测地线。

D 数值实验详情

我们模拟了一个二维范德波尔振荡器，其漂移函数为

$$f_1(x,y) = \mu(x - \frac{1}{3}x^3 - y) \tag{42}$$

$$f_2(x,y) = \frac{1}{\mu}x, \tag{43}$$

初始条件为 $x0 = [1.81, -1.41]$，噪声幅度为 $\sigma = \{0.25, 0.50, 0.75, 1.00\}$，总时长为 $T = \{500, 1000\}$ 时间单位。所采用的观测间隔为 $\tau = \{80, 120, 160, 200, 240, 280, 320\}$。最后一个观测间隔超过振荡器的半周期，因此每个周期仅采样一个状态。这导致了错误的估计。在这种设置下，这表明了我们能够提供估计的 τ 上限。然而，对于任何推断方法，如果观测过程每个周期仅采样一次观测，那么在没有额外假设的情况下，无法通过时间方法识别底层力场。用于模拟真实动力学和进行路径增广的离散化时间步长为 $\delta t = 0.01$。采样受控桥时，我们采用了 $N = 100$ 个粒子，使其按照相关常微分方程演化，如 [40, 101]所述。对数梯度估计器使用了 $M = 40$ 个诱导点。用于估计漂移的稀疏高斯过程基于对 $S = 300$个点进行的稀疏核近似。在所展示的模拟中，我们采用了加权参数 $\beta = 0.5$ （式（33））。这使结果适度地趋向不变密度。图 2中的示例采用了 $\beta = 1$，并且对转移密度的近似优于 $\beta = 0.5$。