几何约束改进稀疏观测随机动力学的推断

Dimitra Maoutsa *柏林工业大学,德国
dimitra.maoutsa@{tu-berlin.de；gmail.com}

arXiv:2304.00423v2 [stat.ME] 4 Apr 2023

摘要

在多尺度上演化的多自由度系统，其动力学通常用随机微分方程来建模。通常，这些方程的结构形式未知，而系统动力学唯一的体现是在离散时间点上的观测。尽管这类方法被广泛应用，但根据时间上稀疏的观测准确推断这些系统仍然颇具挑战。传统推断方法要么关注观测的时间结构，而忽略系统不变密度的几何特性；要么采用不变密度的几何近似，但这仅适用于保守驱动力。为克服这些局限，本文提出一种融合这两种视角的新方法。我们提出一种路径增广方案，利用数据驱动控制来考虑系统不变密度的几何特性。对增广路径进行非参数推断，能够高效识别低采样率观测系统背后的确定性力。

1 引言

从时间序列观测中揭示系统的控制方程，往往对于理解尚未得到解释的自然现象至关重要。我们的目标是找到一种与观测数据相符、并能全面增进对底层机制的现象学理解的数学表示。这需要采用恰当的表示方法，在有效简化多余自由度的同时，捕捉系统的关键性质。随机微分方程（SDE）提供了这样一种灵活的表示方法（Arnold，2014年；Lande，1976；Chandrasekhar，1943；Nelson，2004），它将系统的主导作用力表示为方程中的确定性部分，即（漂移函数） $f(\cdot): R^d \to R^d$，并将尚未解决或无关的自由度概括为作用于系统的随机力（扩散）。由此得到的演化方程形式如下

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

其中， $\sigma \in \mathcal{R}^{d\times d}$ 表示噪声幅度， W_t 表示由 d 个独立维纳过程构成的向量。该方程应按照伊藤形式体系进行解释。我们通过 $\mathcal{O}_k = \psi(X_{k\tau})$ 在不同时间点观测系统状态，其中 $X_{k\tau} \doteq X_t|_{t=\tau k}$， $k = 1, 2, \ldots\ldots, K$ 次观测以**观测间隔** τ采集，并希望根据观测 $\{\mathcal{O}_k\}_{k=0}^{K}$ 近似求出漂移函数 $f(\cdot)$。在这里，我们将考虑 $\psi(x) = x$，但该方法可推广至单调函数 $\psi(\cdot)$。

2 从时间稀疏观测中识别随机系统

对于较小的观测间隔 τ，我们认为观测到了系统状态的连续路径 $X_{0:T}$。

高频观测

在这种情况下，我们可以通过一阶克拉默斯-莫亚尔系数来识别漂移函数 $f(\cdot)$（克拉默斯，1940年；莫亚尔，1949年；塔巴尔，2019年），方法是通过经验估计状态增量的条件期望（Friedrich 等人，2000年；Ragwitz & Kantz，2001年；Boninsegna 等人，2018年；

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

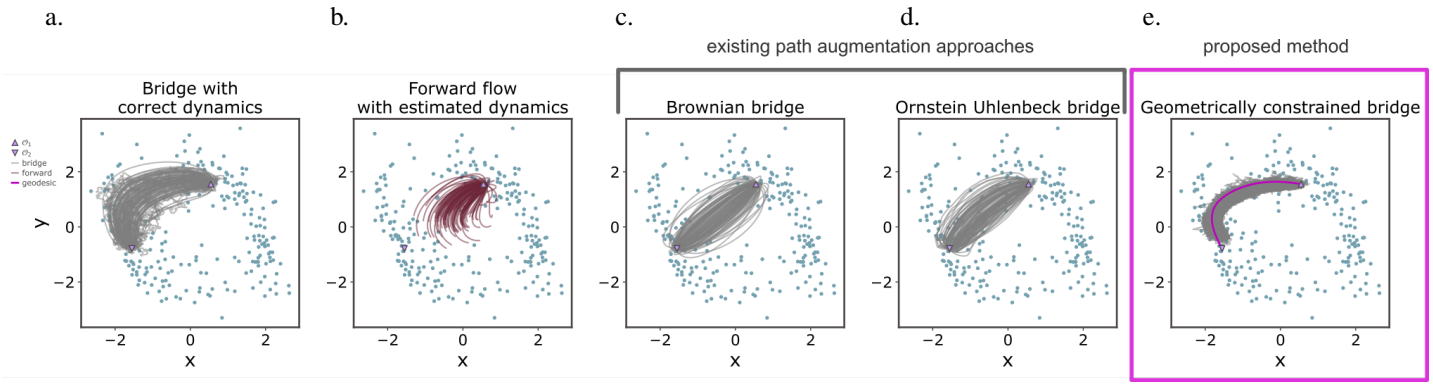


图 1：**现有的路径增广策略对连续观测之间潜在转移密度的匹配效果较差，低估了其曲率。a.)** 遵循真实动力学的两个连续观测 $\mathcal{O}_1$ 和 $\mathcal{O}_2$（粉色三角形）之间的随机桥边缘密度（灰色）。**b.)** 采用高斯似然的估计动力学所对应的前向概率流（栗色）无法很好地匹配正确的转移密度，而且往往无法到达第二个观测 $\mathcal{O}_2$（朝下的粉色三角形）。常见的路径增广策略采用以下任一种方法：**c.)**布朗桥，或**d.)** 基于高斯似然对估计漂移进行局部线性化后得到的奥恩斯坦-乌伦贝克（线性）桥边缘分布。这两种方法都无法很好地匹配正确的转移密度，因为它们低估了其曲率。**e.)** 本文提出的几何约束路径增广方法，通过将桥路径引向连接流形上连续观测的测地线，从而更好地近似潜在的转移密度。

Siegert 等人，1998年）。作为这种方法的贝叶斯非参数对应形式，（Ruttor 等人，2013年；Batz 等人，2018年）认为，对于 $\mathrm{d}t \to 0$，观测之间的转移概率服从高斯分布，因此漂移的似然为（高斯分布）（见第 A 节，公式 7）

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

为识别漂移Ruttor 等人(2013年对函数值施加高斯过程先验 f(方程12)。在方程2 中，我们引入了加权内积 $\langle u, v\rangle \doteq u^\top \cdot \sigma^{-2} v$ 和加权范数$\|u\|_{\sigma^2} \doteq u^\top \cdot \sigma^{-2} u$。

然而，随着观测间隔 τ 增大，连续观测之间的转移概率不能再视为高斯，因此，如果方程2中假设的两个连续观测之间的似然便不再有效，若方程1 是非线性的。

低频观测

对于此类情形，漂移的似然 $\mathrm{P}(\{\mathcal{O}_k\}_{k=1}^K|f)$ 具有如下形式的路径积分

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

其中 $\{\mathcal{O}_k\}_{k=1}^K$ 表示 K 离散时间观测，$\mathrm{P}(X_{0:T}|f)$ 表示假设动力学遵循方程1时的先验路径概率，$\mathcal{D}(X_{0:T})$ 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 $X_{0:T}$时观测的似然。

从几何角度来看，我们认为非线性系统动力学会诱导出一种不变密度，而这种密度可由一个（可能是低维的）流形近似表示。时间稀疏观测是该流形上的样本。对于低频观测，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所诱导的几何结构，因此低估了连续观测之间转移密度的曲率（图4）。

3 几何路径增广

由于式 3 的似然难以处理，我们将观测之间未观测到的连续路径视为潜在随机变量 $X_{0:T}$，并对其求得最大后验估计

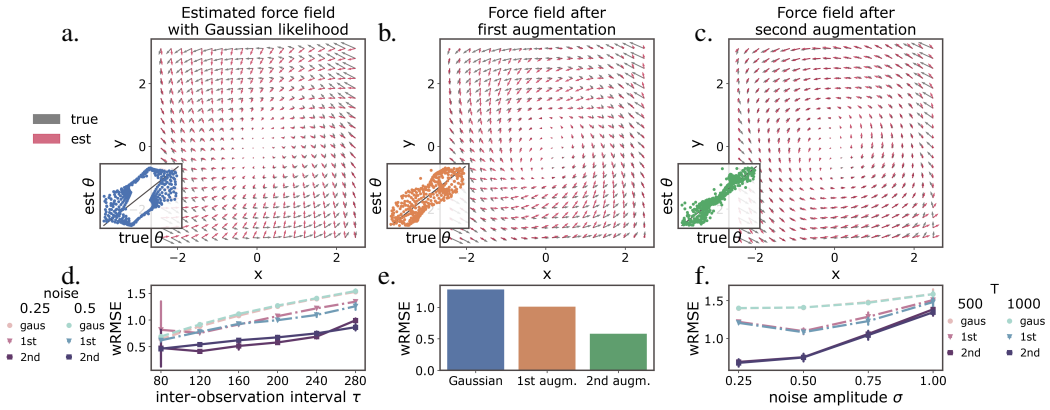


图2：**经过两次迭代后，所提出的路径增广已能较好地近似潜在漂移**。估计的（红色）和真实的（灰色）力场：**a.)** 高斯似然**b.)** 一次增广迭代后，以及**c.)** 第二次增广迭代后。（**内嵌图**）二维网格上每个点的真实角度与估计角度对比。**e.)** 所示示例中每次迭代后漂移估计的加权均方根误差（wRMSE）。每个网格点误差的平均权重根据观测数据进行核密度估计得到 $\{\mathcal{O}_k\}_{k=1}^K$。**d.)** 不同噪声条件下，wRMSE 随观测间隔 τ 的变化；分别展示使用高斯似然估计的漂移（*gaus*-圆形标记）、首次增广后（*1st*-三角形标记）以及第二次增广后（*2nd*-方形标记）的结果，对应 $T = 500$。**f.)** 系统中的 wRMSE 随噪声幅度 σ 的变化；在观测间隔为$\tau = 240$时，比较不同轨迹持续时间 $T = \{500, 1000\}$ 时间单位的结果。标记符号与 d.) 中相同。误差棒表示 5 次独立实现结果的一个标准差。

通过期望最大化（EM）算法估计漂移（Dempster等人，1977）。过去已有类似的参数化方法（Elerian 等人，2001年；Sermaidis等人，2013年）和非参数化方法（Batz 等人，2018年；Ruttor 等人，2013 年）研究漂移推断，主要针对高频观测情形。我们的方法基于（Batz 等人，2018 年；Ruttor 等人，2013 年）中讨论的非参数方法，并作出了两项重大改进：

(i) 我们采用一种路径增广方案，该方案基于**估计得到的非线性动力学**，这些动力学由采用式中高斯似然进行推断2 得到（而非采用 (Batz 等人，2018年) 中提出的这些动力学的局部线性近似）。(ii) 重要的是，我们还进一步**约束增强路径，使其在连续观测之间与不变密度的几何结构相一致**（图 4 b.）。

我们采用一种迭代算法，在每次迭代中 n 执行以下两个步骤：**(1.) E（期望）步**：给定漂移估计 f^n，构造潜变量上的近似后验分布 $Q(X_{0:T}) \approx \mathrm{P}(X_{0:T}|\{\mathcal{O}\}_{k=1}^K, f^n(x))$。

(2.) M（最大化）步：更新漂移估计。</output>

• **路径上的近似后验分布。（E步）**我们将观测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

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

项 $U_{\mathcal{O}}(x,t) \doteq -\sum_{t_k} \ln \hat{\mathrm{P}}(\mathcal{O}_k|x)\delta(t-t_k)$ 强制潜在路径经过观测（或根据观测过程而接近观测），而 $U_{\mathcal{G}}(x,t) \doteq \|\Gamma_t - x\|^2$ 则引导潜在路径趋向流形上连接相邻观测的测地曲线 $\gamma^k_{t'}\mathcal{M}$；该流形由系统的不变密度诱导（第 A.1.2 节）。这里记作 $\Gamma_t \doteq \{\gamma^k_{t'}\}_{t=(k-1)\tau+t'\tau}$，其中 $\gamma^k_{t'}$ 是连接 $\mathcal{O}_k$ 和 $\mathcal{O}_{k+1}$的测地线，且 $t' \in [0,1]$。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 $\gamma^k_{t'}$（见第 A.1.2 节和 Arvanitidis 等人（2019））。

按照 Opper (2019年) 的方法，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来确定后验路径测度（即方程 4 的极小化解） Maoutsa &

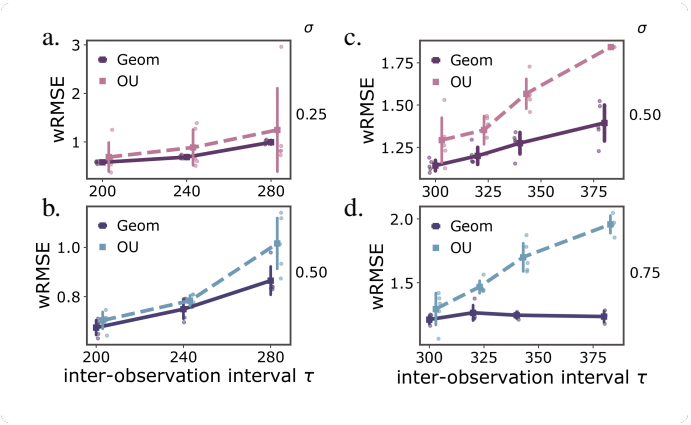


图 3：**所提出的路径增广方法与 Ornstein-Uhlenbeck 增广方法的比较**。加权均方根误差（wRMSE）随噪声幅度 σ 的变化；对于中等观测间隔，噪声幅度对应**a.)** $\sigma = 0.25$和**b.)** $\sigma = 0.50$；对于较大的观测间隔，对应**c.)** $\sigma = 0.50$和**d.)** $\sigma = 0.75$，且运动方向已知。在**c., d.** 中，观测间隔导致每个振荡周期内仅有一次观测。

Opper（2022；2021）；Maoutsa（2022）旨在为漂移为 $f(x)$的系统获取时变漂移调整 $\hat{u}(x,t) := g(x,t) - f(x)$，其中初始和终端约束由 $U_{\mathcal{O}}(x,t)$ 确定，另有路径约束 $U_{\mathcal{G}}(x,t)$。

• **漂移估计。（M步）**为了从采样得到的潜在路径中估计漂移，我们假设函数值服从高斯过程先验，并采用与 Batz 等人(2018年) 类似的稀疏核近似（详见第 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子进行了系统估计 σ，并在不同观测间隔下进行观测 τ，轨迹长度也各不相同 T (参见第B节)。对于所考察的噪声幅度（图 2 f.），所提出的路径增广算法在两次迭代内即可改善基于高斯假设的朴素估计，且适用于大多数噪声幅度（图 2）。随着噪声增大，该方法带来的改进逐渐减小（图 2f.），但仍不可忽略。在低噪声条件下，测地线能更好地近似流形结构；然而，路径积分控制受到与逆噪声协方差成正比的控制代价限制。对于所有观测间隔长度，我们的框架都具有相近的精度；但在较短间隔下，改进幅度较小，因为此时采用高斯似然的估计已能较好地近似真实漂移。我们将所提出的方法与 Batz 等人(2018年)提出的增广方法（采用奥恩斯坦-乌伦贝克桥进行增广）进行了比较；在较大的观测间隔下，前者给出了更准确的估计。对于每个振荡周期仅有一次观测的观测间隔（图 3c.,d.），所提出的方法还利用了状态空间中的运动方向信息，因此效果更佳（参见第 B节）。由于以系统的不变几何为条件，所提出方法的估计方差小于 Batz 等人方法的估计方差。

5 结论与讨论

我们提出了一种新方法，通过对系统状态进行时间稀疏观测来识别随机系统。该方法将纯粹依赖观测时间结构的方法与近似不变密度几何结构的方法结合起来。我们的方法采用路径增广策略，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，使其遵循系统不变密度的局部几何结构。我们发现，对于周期或准周期系统，在多种噪声条件下，所提出的方法能够高效地恢复潜在漂移函数。只有少数推断方法尝试结合几何与时间视角来识别随机系统，例如朗之万回归（Callaham 等人，2021年）、TrajectoryNet（Tong 等人，2020年）以及Shnitzer 等人（2020年；2016）提出的扩散映射方法。然而，我们的方法与这些方法不同之处在于，它采用了潜在数据几何的测地线近似（另见第C节 未来方向）。

致谢

我们感谢 Mina Miolane 早期指导我们从基于变分自编码器的数据流形近似中提取测地线，感谢 Stefan Sommer 解答有关流形上扩散的问题，也感谢 Manfred Opper 教授促使我们着手研究这一问题。我们还感谢 Georgios Arvanitidis 维护开放代码仓库，分享其此前关于从数据流形近似测地线的研究。本研究的早期报告曾发表于 NeurIPS 2022 研讨会机器学习与物理科学（Maoutsa，2023年）。我们还承认，Python（Van Rossum& Drake Jr，1995年）、NumPy（Harris 等人，2020年）、SciPy（Virtanen 等人，2020年）、Matplotlib（Hunter，2007）、Seaborn（Waskom，2021年）、GPflow（Matthews 等人，2017年）、pyEMD（Pele& Werman，2009）和 PyTorch（Paszke 等人，2017年社区的前期工作为本研究计算部分的实现提供了帮助。

本研究的实现可在以下代码仓库中找到：
https://github.com/dimitra-maoutsa/Geometric-path-augmentation-for-SDEs。

参考文献

Ludwig Arnold。将随机微分方程视为动力系统。载于《系统理论中的实现与建模》，第489–495页。施普林格，1990年。（引用页码：*13*）

Stevan J Arnold。《表型进化：持续进行的综合研究：美国博物学家学会主席演讲》，2014年。（引用页码：*1*）

Georgios Arvanitidis、Lars Kai Hansen 和 Søren Hauberg。潜在空间奇异性：深度生成模型的曲率。*arXiv* 预印本 *arXiv:1710.11379*，2017年。doi：https://doi.org/10.48550/ arXiv.1710.11379。（引自第*14*页）

Georgios Arvanitidis、Soren Hauberg、Philipp Hennig 和 Michael Schober。在数据中学习的流形上的快速、稳健短路径。载于 第*22*届人工智能与统计国际会议，第1506–1515页。PMLR，2019年。（引自第*3* 页和*14*页）

Philipp Batz、Andreas Ruttor 和 Manfred Opper。随机微分方程的近似贝叶斯学习。《物理评论 *E*》，98(2):022109，2018 年。（引自第 *2* 、 *3* 、 *4* 页和第 *10* 页）

Alexandros Beskos、Omiros Papaspiliopoulos 和 Gareth O Roberts。扩散样本路径的回溯式精确模拟及其应用。《伯努利》，12(6):1077–1098，2006年。（引自第*10*页）

Lorenzo Boninsegna、Feliks N¨uske 和 Cecilia Clementi。随机动力学方程的稀疏学习。《化学物理学杂志》，148(24):241723，2018年。doi: 10.1063/1.5018409。（引自第*1*页）

Jared L Callaham、J-C Loiseau、Georgios Rigas 和 Steven L Brunton。使用朗之万回归进行非线性随机建模。《皇家学会会刊*A*辑》，477(2250):20210092，2021年。（引自第*4*页）

Andrew Carverhill。随机动力系统的流：遍历理论。*Stochastics:*《国际概率与随机过程杂志》，14(4):273–317，1985年。（引自第*13*页）

苏布拉马尼扬·钱德拉塞卡。物理学与天文学中的随机问题。《现代物理评论》，15(1):1,1943年。*(*引文所在页：*1)*

拉斐¨尔 Chetrite 和 Hugo Touchette。条件化与驱动过程的变分及最优控制表示。《统计力学杂志：理论与实验》，2015 (12):P12001, 2015年。*(*引文所在页：*11)*

Siddhartha Chib、MichaelKPitt 和 Neil Shephard。基于似然的扩散驱动状态空间模型推断。*Por Clasificar*，第 1–33 页，2006年。*(*引文所在页：*10)*

Nguyen Dinh Cong 等人 随机动力系统的拓扑动力学。牛津大学出版社，1997年。（引自第*16*页）

Arthur P Dempster、Nan M Laird 和 Donald B Rubin。通过EM算法从不完整数据中求取最大似然。《英国皇家统计学会杂志：B辑（方法论）》，39(1): 1–22，1977年。（引自第3页）Manfredo Perdigao Do Carmo 和 J Flaherty Francis。《黎曼几何》，第6卷。施普林格，1992年。（引自第13页和第14页）Ola Elerian、Siddhartha Chib 和 Neil Shephard。对离散观测非线性扩散过程的似然推断。《计量经济学》，69(4):959–993，2001年。（引自第3页）Neil Fenichel 和 JK Moser。流的不变流形的持续性与光滑性。《印第安纳大学数学杂志》，21(3):193–226，1971年。（引自第13页）Rudolf Friedrich 和 Joachim Peinke。用福克-普朗克方程描述湍流级联。《物理评论快报》，78(5):863，1997年。doi: https://doi.org/10.1103/PhysRevLett.78.863。（引自第10页）Rudolf Friedrich、Silke Siegert、Joachim Peinke、Marcus Siefert、Michael Lindemann、Jan Raethjen、G¨untner Deuschl、Gerhard Pfister 等人。从实验数据中提取模型方程。《物理快报A辑》，271(3):217–222，2000年。（引自第1页）Christian Fr¨ohlich、Alexandra Gessner、Philipp Hennig、Bernhard Sch¨olkopf 和 Georgios Arvani- tidis。在黎曼数据流形上进行贝叶斯求积。载于《国际机器学习会议》，第3459–3468页。PMLR，2021年。doi: https://doi.org/10.48550/arXiv.2102.06645。（引自第14页）TV Girya 和 Igor Dmitrievich Chueshov。随机扰动耗散动力系统的惯性流形和平稳测度。《数学文集》，186(1):29–45，1995年。（引自第13页）Andrew Golightly 和 Darren J Wilkinson。对带误差观测的非线性多元扩散模型进行贝叶斯推断。《计算统计与数据分析》，52(3):1674–1693，2008年。Charles R. Harris、K. Jarrod Millman、St´efan 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 Fern´andez del R´ıo、Mark Wiebe、Pearu Peterson、Pierre G´erard-Marchant、Kevin Sheppard、Tyler Reddy、Warren Weckesser、Hameer Abbasi、Christoph Gohlke 和 Travis E. Oliphant。使用NumPy进行数组编程。《自然》，585(7825):357–362，2020年9月。doi: 10.1038/s41586-020-2649-2。URL https://doi.org/10.1038/s41586-020-2649-2。（引自第5页）Christoph Honisch 和 Rudolf Friedrich。低采样率下克拉默斯-莫亚尔系数的估计。《物理评论E》，83(6):066701，2011年。（引自第10页）J. D. Hunter。Matplotlib：一个二维绘图环境。《计算科学与工程》，9(3):90–95，2007年。doi: 10.1109/MCSE.2007.55。Hendrik Anthony Kramers。力场中的布朗运动与化学反应扩散模型。《物理学》，7(4):284–304，1940年。（引自第1页）Steven J Lade。克拉默斯–莫亚尔分析中的有限采样间隔效应。《物理快报A辑》，373 (41):3705–3709，2009年。（引自第10页）Russell Lande。表型进化中的自然选择与随机遗传漂变。《进化》，第314–334页，1976年。（引自第1页）John M Lee。《黎曼流形导论》，第176卷。施普林格，2018年。（引自第13页）李宗毅、Nikola Kovachki、Kamyar Azizzadenesheli、Burigede Liu、Kaushik Bhattacharya、Andrew Stuart 和 Anima Anandkumar。用于参数化偏微分方程的傅里叶神经算子。国际学习表征会议，2020a年。（引自第17页）

李宗毅、Nikola Kovachki、Kamyar Azizzadenesheli、Burigede Liu、Kaushik Bhattacharya、An- drew Stuart 和 Anima Anandkumar。《神经算子：用于偏微分方程的图核网络》。arXiv 预印本 arXiv:2003.03485，2020b。（引用见第 17 页）Robert S Liptser 和 Albert N Shiryaev。《随机过程统计学 II：应用》，第 6 卷。Springer Science & Business Media，2013。doi：10.1007/978-3-662-10028-8。（引用见第 9 页）Satya N Majumdar 和 Henri Orland。《受约束随机过程的有效朗之万方程》。《统计力学杂志：理论与实验》，2015(6)：P06039，2015。（引用见第 11 页）Dimitra Maoutsa。《从单次试验脉冲序列观测中揭示潜在随机动力学》。伯恩斯坦计算神经科学会议，柏林，2022 年 9 月。（引用见第 4 页）Dimitra Maoutsa。《稀疏观测随机非线性系统推断的几何路径增广》。arXiv 预印本 arXiv:2301.08102，2023。（引用见第 5 页）Dimitra Maoutsa 和 Manfred Opper。《用于约束随机微分方程的确定性粒子流》。机器学习与物理科学，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。（引用见第 4、12、15 和 16 页）Dimitra Maoutsa 和 Manfred Opper。《用于约束随机非线性系统的确定性粒子流》。《物理评论研究》，4：043035，2022 年 10 月。doi：10.1103/PhysRevResearch.4.043035。URL：https://link.aps.org/doi/10.1103/PhysRevResearch.4.043035。（引用见第 3、12、15 和 16 页）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》。《熵》，22(8)：802，2020。（引用见第 16 页）Alexander G. de G. Matthews、Mark van der Wilk、Tom Nickson、Keisuke. Fujii、Alexis Bouk-ouvalas、Pablo Le´on-Villagr´a、Zoubin Ghahra mani 和 James Hensman。《GPflow：使用 TensorFlow 的高斯过程库》。《机器学习研究杂志》，18(40)：1–6，2017 年 4 月。URL：http://jmlr.org/papers/v18/16-537.html。（引用见第 5 页）Salah-Eldin A Mohammed 和 Michael KR Scheutzow。《随机微分方程的稳定流形定理》。《概率年刊》，第 615–652 页，1999 年。doi：https://doi.org/10.1214/aop/1022677380。（引用见第 13 页）Jose Moyal。《随机过程与统计物理》。《英国皇家统计学会杂志：B 系列（方法论）》，11(2)：150–210，1949 年。（引用见第 1 页）Philip Nelson。《生物物理学》。WH Freeman，纽约，2004 年。（引用见第 1 页）Manfred Opper。《随机微分方程的变分推断》。《物理学年鉴》，531 (3)：1800233，2019 年。（引用见第 3、11、12 和 15 页）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》。《生物计量学》，99(3)：511–531，2012 年。（引用见第 10 页）Adam Paszke、Sam Gross、Soumith Chintala、Gregory Chanan、Edward Yang、Zachary DeVito、Zeming Lin、Alban Desmaison、Luca Antiga 和 Adam Lerer。《PyTorch 中的自动微分》，2017 年。（引用见第 5 页）Ofir Pele 和 Michael Werman。《快速且稳健的推土机距离》。载于：2009 年 IEEE 第 12 届国际计算机视觉会议，第 460–467 页。IEEE，2009 年 9 月。（引用见第 5 页）Mario Ragwitz 和 坎茨。《基于时间序列数据的福克-普朗克方程不可或缺的有限时间修正》。《物理评论快报》，87(25)：254501，2001 年。doi：10.1103/physrevlett.87.254501。（引用见第 1 页）

Carl Edward Rasmussen。《机器学习中的高斯过程》。载于：机器学习暑期学校，第 63–71 页。Springer-Verlag，2003 年。（引用见第10页）

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。神经信息处理系统进展，26，2013年。（引用页码：2，3，9，10和 15）

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。斯堪的纳维亚统计学杂志，40(2):294–321，2013年。（引用页码：3 和 10）

Tal Shnitzer、Ronen Talmon 和 Jean-Jacques Slotine。利用收缩观测器进行流形学习，实现数据驱动的时间序列分析。*IEEE* 信号处理汇刊，65(4):904–918，2016。doi: https://doi.org/10.1109/TSP.2016.2616334。（引用页码：4）

Tal Shnitzer、Ronen Talmon 和 Jean-Jacques Slotine。数据驱动动力系统分析的流形学习。载于系统与控制中的库普曼算子，第 359–382 页。施普林格，2020年。*(引自第4页)*

Silke Siegert、R Friedrich 和 J Peinke。随机系统数据集分析。《物理快报*A*辑》，243(5-6):275–280，1998年。doi: 10.1016/S0375-9601(98)00283-7。*(引自第2页)*

Jaroslav Stark、David S Broomhead、Michael Evan Davies 和 J Huke。受迫系统的延迟嵌入。II. 随机受迫。《非线性科学杂志》，13:519–577，2003年。*(引自第16页)*

M Tabar。克拉默斯–莫亚尔展开与福克–普朗克方程。载于复杂非线性动力系统的分析与数据驱动重构，第 19–29 页。施普林格，2019年。*(引自第1页)*

Ella Tamir、Martin Trapp 和 Arno Solin。带支持的输运：数据条件扩散桥。*arXiv* 预印本 *arXiv:2301.13636*，2023年。（引用页码：17）

Alexander Tong、Jessie Huang、Guy Wolf、David Van Dijk 和 Smita Krishnaswamy。Trajectory-Net：一种用于建模细胞动力学的动态最优输运网络。载于国际机器学习会议，第9526–9536页。PMLR，2020年。（引用页码：4）

Guido Van Rossum 和 Fred L Drake Jr. *Python*教程，第620卷。数学与计算机科学中心信息学，阿姆斯特丹，荷兰，1995年。（引用页码：5）

Pauli Virtanen、Ralf Gommers、Travis E. Oliphant、Matt Haberland、Tyler Reddy、David Courna- peau、Evgeni Burovski、Pearu Peterson、Warren Weckesser、Jonathan Bright、St´efan J. van der Walt、Matthew Brett、Joshua Wilson、K. Jarrod Millman、Nikolay Mayorov、Andre w R. J. Nel- son、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、Ant ˆonio H. Ribeiro、Fabian Pedregosa、Paul van Mul- bregt 和 SciPy 1.0 贡献者。《SciPy 1.0：Python科学计算的基础算法》。《自然·方法》，17：261–272，2020年。doi：10.1038/s41592-019-0686-2。（引用页码：5）

Ziyu Wang、Shuyu Cheng、Li Yueru、Jun Zhu 和 Bo Zhang。一种学习未归一化模型的Wasserstein最小速度方法。载于 人工智能与统计国际会议，第 3728–3738 页。PMLR，2020年。（引用页码：16）

Michael L Waskom。Seaborn：统计数据可视化。《开放源代码软件期刊》，6（60）：3021，2021年。（引用页码：5）

StephenWiggins. 动力系统中的正规双曲不变流形，第105卷。Springer Science & Business Media，1994年。*(引自第13页)*

附录

高频与低频观测下的漂移推断

我们考虑其演化由随机微分方程 Eq. 描述的系统。 1.

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

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

相应地，离散化维纳路径 $P_{\mathcal{W}}(X_{0:T})$（即来自无漂移过程的路径）的转移概率可以表示为

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

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，我们可以通过以下两者之间的拉东-尼科迪姆导数来表示漂移 f 的似然$P_f(X_{0:T}|f)$和 $\mathrm{P}_{\mathcal{W}}(X_{0:T})$对于时间区间 $X_{0:T}$ 内的路径 $[0, T]$(Liptser & Shiryaev， 2013 年)

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

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。该表达式是对由漂移为 f 的过程和具有相同扩散 σ的维纳过程所诱导的路径测度应用吉尔萨诺夫定理，并对连续路径 $X_{0:T}$采用欧拉-丸山离散化后得到的。

给定随机微分方程（式 7）的连续观测路径，其似然关于漂移函数具有二次型。因此，对函数值（高斯过程）采用高斯测度是这一似然的自然共轭先验。为以非参数形式识别漂移，我们假设函数值服从高斯过程先验 $f \sim \mathrm{P}_0(f) = \mathrm{GP}(m^f, k^f)$，其中 m^f和 k^f分别表示该高斯过程的均值函数和协方差函数（Ruttor 等人， 2013年）。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，需要在函数空间中计算一个概率分布，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 表示一个归一化因子，其定义为路径积分

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

其中 $\mathcal{D}f$ 表示对希尔伯特空间 f 的积分： $H_0[f] < \infty$ 。在这里，我们将函数上的先验概率表示为 $\mathrm{P}_0(f) = e^{-H_0[f]}$。在（Ruttor 等人， 2013年）中，作者证明，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 $\mathrm{d}X_t$ (Rasmussen, 2003）。更确切地说，我们将 N 系统状态观测 X_t 作为回归变量，其对应的响应变量为

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

并将高斯过程的核函数记为 $k(x,\ x')$。

若将状态观测和观测增量集合分别记为 $\mathcal{X} = \{X_t\}_{t=0}^{T-\mathrm{d}t}$ 和 $\mathcal{Y} = \{Y_t\}^{T-\mathrm{d}_{t=0}\ t}$，则漂移函数 f 的后验过程均值可表示为

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

这里我们稍微滥用了记号，用 $k^f(x,\ \mathcal{X})$ 表示在点 x 和 $\{\mathcal{O}_t\}_{k=1}^{K-1}$ 处对核函数 k^f 求值得到的向量。类似地，$\mathcal{K} = k^f(\mathcal{X},\ \mathcal{X})$ 表示对所有观测对应用核函数后得到的 $(K-1) \times (K-1)$ 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（即低频观测）时，公式 7 中的高斯似然不再有效，因为观测间隔较大时，转移密度不再是高斯分布。因此，基于高斯假设的漂移估计（Friedrich & Peinke, 1997；Ruttor 等人，2013）会变得不准确。为缓解这一问题，Lade（Lade, 2009）提出了一种计算漂移估计值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（Honisch & Friedrich, 2011）。另一方面，统计学界提出了路径增广方案：通过对观测之间简化系统的动力学进行采样，将观测轨迹增广为近似连续时间轨迹（Golightly & Wilkinson, 2008；Papaspiliopoulos et al., 2012；Sermaidis 等人，2013；Beskos et al., 2006；Chib et al., 2006）。然而，对于较大的观测间隔和非线性系统，路径增广所采用的简化动力学与底层路径统计特性之间匹配较差，因此这些方法的收敛速度较慢，或无法识别出正确的动力学（图 1 c. 和 d.）。我们指出，若以正确动力学的局部线性化作为漂移，采用 Ornstein–Uhlenbeck桥进行路径增广，就能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值进行的局部线性化（Batz 等人，2018）在低频观测下效果不佳。

注意，随着观测间隔 τ 增大，如果系统是非线性的，或者噪声依赖于状态，那么假设相邻两次观测之间服从高斯似然便不再成立。在这种情况下，漂移的似然可以表示为一个路径积分

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

其中，$\mathcal{O}_{1:K} \doteq \{\mathcal{O}_k\}_{k=1}^K$ 表示离散时间观测的集合，$\mathrm{P}(X_{0:T} \mid f)$ 表示由漂移为 $f(x)$ 的扩散过程产生的先验路径概率，$\mathcal{D}(X_{0:T})$ 表示路径空间上的形式体积元，而 $\mathrm{P}(\mathcal{O}_{1:K} \mid X_{0:T})$ 表示给定潜在路径 $X_{0:T}$ 时观测的似然。

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隐状态，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，为漂移函数求得最大后验估计。具体而言，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 $\hat{}Q(X_{0:T}) \approx \mathrm{P}(X_{0:T} \mid \mathcal{O}_{1:K}, f^n(x))$，并计算增强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，在此步骤中，我们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 P_0 表示函数值上的高斯过程先验。

A.1 路径上的近似后验分布。

这里，我们首先仅将各个观测作为约束（第 A.1.1 节），来构建路径上的近似后验分布（即给定观测时路径的条件分布）。然而，这种方法会导致路径增广过程中的计算十分繁重，因为这些观测是初始估计的漂移中的非典型状态。为解决这一问题，我们随后扩展该形式化方法（第 A.1.2 节），纳入同时考虑观测局部几何的约束。

A.1.1 无几何约束时路径上的近似后验分布。

$\hat{}$ 给定漂移函数（或漂移估计） $f(x)$，我们可以运用变分技术，近似计算在观测 $\{\mathcal{O}_k\}_{k=1}^K$ 条件下潜在路径的后验测度。我们认为，先验过程（即不考虑观测 $\{\mathcal{O}_k\}_{k=1}^K$ 的过程）由以下方程描述

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

我们将定义一个以观测为条件的近似（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 $g(x,t)$，以纳入观测信息（Chetrite & Touchette，2019年；Majumdar& Orland，2015）。我们将近似后验测度 Q 视为由一个以观测为条件的近似过程所诱导的后验测度 $\mathcal{O}_{1:K}$ (Opper，2019)，其控制方程为

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

有效漂移 $g(X_t,t)$ 可由以下变分问题的解得到：最小化自由能 方程 18 的

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

通过应用卡梅伦-吉尔萨诺夫-马丁定理，我们可以将由漂移为 $f(x)$ 和 $g(x,\ t)$ 的扩散过程所诱导的两个路径测度之间的库尔贝克-莱布勒散度 ˆ 表示为

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

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

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

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

其中，$q_t(x)$表示近似过程的 X_t 边缘密度。在第三行中，我们引入了随机变量 $V_T = \int_0^T \hat{}f(X_t) - g(X_t,t)\,\sigma^2\,\mathrm{d}W_t$。假设函数 $\ell(X_t) = \hat{}f(X_t) - g(X_t,t)$ 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，这会产生一个常数 $\mathfrak{C}$。因此，公式 23 中的第二项与自由能的最小化无关，故将其省略。

因此，我们可以将公式19 的自由能表示为（Opper，2019年）

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

其中，项 $U(x,\ t)$ 用于计入观测 $U(x,\ t) = -\sum \ln \mathrm{P}(\mathcal{O}_k \mid x)\delta(t - t_k)$。$t_k$

自由能泛函的最小化可视为一个随机控制问题（Opper，2019年），其目标是确定一个时变漂移调整 $u(x,\ t) :=\hat{}\quad \hat{}g(x,\ t) - f(x)$，使漂移为 $f(x)$ 的系统满足观测施加的约束。

对于精确观测的情形，即观测过程 $\psi(x) = x$，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 $u(x,\ t)$，使其遵循式17 的动力学，并以相应观测作为初始和终止状态 $\mathcal{O}_k$ 和 $\mathcal{O}_{k+1}$。

对于这样的随机控制问题，观测间隔 $\mathcal{O}_k$ 与 $\mathcal{O}_{k+1}$ 之间的最优漂移调整可以通过求解后向方程得到（参见（Maoutsa&Opper，2022；2021））

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

终端条件为 $\phi_T(x) = \chi(x) = \delta(x - \mathcal{O}_{k+1})$，其中 $\mathcal{L}^\dagger$ ˆf 表示式 17 所述过程的伴随 Fokker-Planck 算子。Maoutsa 等人（Maoutsa & Opper，2022年；2021年）证明，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（PDE）

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

而 q_t 是一个时间反演偏微分方程的解，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 带几何约束的路径近似后验分布。

前文基于随机桥构造近似测度的方法，适用于观测在先验扩散过程的概率律下具有非零概率的情形^（式 17）。然而，当先验过程（其漂移估计值为 f）与生成观测的过程存在较大差异时，这种构造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收敛缓慢。为解决这一问题，我们在此为后验过程加入额外约束，使后验测度中的路径遵循观测的局部几何结构。接下来，我们将简要介绍黎曼几何的基础知识，随后继续讨论所提方法中的几何考量。

黎曼几何。嵌入在 D维环境空间 $\mathcal{X} = \mathcal{R}^D$ 中的 d维黎曼流形（Do Carmo 与 Flaherty Francis，1992年；Lee，2018年）（$\mathcal{M}, h$）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： $x \to \langle \cdot | \cdot \rangle_x$ 作用于 $\mathcal{T}_x\mathcal{M}$。每个点 $x \in \mathcal{M}$处都定义了一个切空间 $\mathcal{T}_x\mathcal{M}$ 。黎曼度量 h 在流形 $\mathcal{M}$上定义了一个标准体积测度。直观而言，这描述了如何在流形 $\mathcal{M}$的切空间上局部计算点之间的内积，因此也决定了如何计算 $\mathcal{M}$上点的范数，进而计算点之间的距离。

一个**坐标图** (G, ϕ)提供了从 $\mathcal{M}$ 上的开集 G 到欧几里得空间中的开集 V 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 。我们可以用正定矩阵（**度量张量**） $H(x) = (h_{i,j})_{x,\ 0 \le i,\ j,\ \le d} = \left(\langle \frac{\partial}{\partial x_i} | \frac{\partial}{\partial x_j} \rangle_x \right)_{0 \le i',\ j,\ \le d}$ 表示局部坐标图（G， ϕ）上的度量 h ，该矩阵在每个点 $x \in G$处定义。

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

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

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

其中，带点的字母表示曲线 $\dot{\gamma}_{t'} = \partial_{t'}\gamma_{t'}$的速度。**测地曲线**是连接流形上给定两点、在局部范围内长度最短的光滑曲线。

观测值的黎曼几何。为了近似路径的后验分布，我们将观测值所体现的不变密度的几何结构纳入考量。为此，我们考虑这样一些系统：其动力学会产生包含系统全局吸引子的（惯性）不变流形，并且系统轨迹集中于这些流形上（Wiggins，1994年；Mohammed & Scheutzow，1999年；Girya & Chueshov，1995年；Fenichel & Moser，1971年；Arnold，1990年；Carverhill，1985年）。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 $\mathcal{O}_k$ 因此是流形 $\mathcal{M}$上的样本。我们方法的核心前提是，相邻观测值之间未观测到的路径将位于流形 $\mathcal{M}$上或其附近。具体而言，我们假定未观测到的路径应位于连接 $\mathcal{M}$上相邻观测值的测地线附近。为此，我们提出一种路径增广框架，约束增强路径位于所识别的相邻观测值间测地线附近。

然而，尽管将低维流形嵌入高维环境空间的这一视角有助于我们理解所提出的方法，但出于计算目的，我们

采用一种互补的视角，该视角受到（Fr¨ohlich 等人，2021年）中的讨论启发。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。观测的非线性几何的影响随后由度量 h 捕获。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 $H:\ \mathcal{R}^d \to \mathcal{R}^{d\times d}$ 并根据学习到的度量计算连续观测之间的测地线。直观地说，根据这一视角，观测 $\{\mathcal{O}_k\}_{k=1}^K$ 会对我们在状态空间中计算距离的方式造成扭曲。

实际上，这种方法并未降低我们所操作空间的维度，而是改变了计算内积的方式，从而改变了其上的距离、长度和测地曲线 $\mathcal{M}$。采用低维流形进行处理是另一种思路，但这种方法在很大程度上取决于能否正确评估该流形的维度。例如，可以使用变分自编码器来近似观测流形，随后根据解码器所中介的流形嵌入来获得黎曼度量。然而，我们对此方法的初步结果表明，要使这种方法适应各动力系统的特性，需要进行大量微调，而且该方法对近似流形维度的估计十分敏感。

为学习黎曼度量并计算测地线，我们遵循 Arvanitidis 等人(2019年 提出的框架。具体而言，我们将状态空间位置 $\mathbf{x}$ 处由观测所诱导的局部度量，以非参数形式近似为根据观测计算得到的加权局部对角协方差的逆矩阵，如下所示（Arvanitidis 等人，2019年）

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

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，我们将测地线 $\gamma_{t'}^k$ 视为能量最小化曲线，即视为动能泛函 $\mathcal{E}(\gamma_{t'}^k) = \int_0^1 L_{\mathcal{M}}(\gamma_{t'}^k, \dot{\gamma}_{t'}^k)\mathrm{d}t'$ 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。该泛函的极小曲线与曲线长度泛函 $\ell(\gamma_{t'})$ (式 29)，也就是测地线的极小化曲线相同（Do Carmo 与 Flaherty Francis，1992年）。

通过应用变分法，泛函的极小曲线 $\mathcal{E}(\gamma_{t'}^k)$ 可由欧拉-拉格朗日方程求得，从而得到以下二阶微分方程组（Arvanitidis 等人，2017年；Do Carmo 与 Flaherty Francis，1992年）

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

其边界条件为 $\gamma_0^k = \mathcal{O}_k$ 和 $\gamma_1^k = \mathcal{O}_{k+1}$，其中 $\otimes$ 表示克罗内克积，而 $\mathrm{vec}[A]$ 表示通过将矩阵 A 的各列堆叠 A 成向量的向量化操作。Arvanitidis 等人（2019年）通过使用概率微分方程求解器近似求解方程32 的边值问题，从而得到测地线。

扩展自由能泛函。 我们将各条测地线的集合记为 $\Gamma_t \doteq \{\gamma_{t'}^k\}_{t=(k-1)\tau+t'\tau}$，其中 $\gamma_{t'}^k$ 是连接 $\mathcal{O}_k$ 和 $\mathcal{O}_{k+1}$的测地线，而 $t' \in [0,1]$ 表示重缩放时间变量。除了前文所述设定（第 A.1.1 节）中施加的约束外，我们还在自由能 $U_{\mathcal{G}}(x,t) \doteq \|\Gamma_t - x\|^2$ 中加入一个额外项，以刻画

不变密度的局部几何，并引导潜在路径趋向连接 γ_t^k相邻观测的测地曲线

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

这里我们用 $U_{\mathcal{O}}(x,\ t) \doteq -\sum_{t_k} \ln \mathrm{P}(\mathcal{O}_k|x)\delta(t-t_k)$ 表示观测项，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。

遵循 Opper（2019）的做法，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（Maoutsa & Opper，2022；2021）来确定后验路径测度（式 33 的极小化解）；该问题的目标是为具有漂移 $f(x)$ 的系统求得时变漂移调整 $u(x,t) :=\hat{} \hat{} g(x,t) - f(x)$，同时满足由 $U_{\mathcal{O}}(x,t)$ 定义的初始和终端约束，以及附加的路径约束 $U_{\mathcal{G}}(x,t)$。

A.2 漂移函数上的近似后验

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

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

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函数和协方差函数。按照 Ruttor 等人 (Ruttor 等人，2013年)，我们通过在一组稀疏的 f 诱导点上优化函数值，对漂移 S 进行稀疏核近似 $\{Z_i\}_{i=1}^S$。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 $\mathcal{K}_S \doteq k^f(\mathcal{Z},\ \mathcal{Z})$

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

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

我们对式 34 中的密度采用基于样本的近似，该近似由对路径测度 Q进行粒子采样得到。因此，将密度表示为样本的形式后，可以将密度 $p_t(x)$ 改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。其中，$X_j(t)$ 表示第 j 个粒子在时间点 t的位置。

B 数值实验详情

我们模拟了一个具有如下漂移函数的二维范德波尔振子

$$f_1(x,y) = \mu(x - \frac{1}{3}x^3 - y) \tag{38}$$

$$f_2(x,y) = \frac{1}{\mu}x, \tag{39}$$

从初始条件 $x0 = [1.81, -1.41]$ 出发，并在噪声幅度 $\sigma = \{0.25, 0.50, 0.75, 1.00\}$ 下运行，总时长为 $T = \{500, 1000\}$ 时间单位。采用的观测间隔为 $\tau \in [80, 320] \times dt$。最后一个观测间隔超过了振子的半个周期，因此每个周期仅采样一个状态。这导致了错误的估计。在这种情况下，这表明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，使用时间方法无法识别底层力场。用于模拟真实动力学和路径增广的离散时间步长为 $\delta t = 0.01$。为采样受控桥，我们采用了 $N = 100$ 个粒子，使其按照相关常微分方程演化，具体如 Maoutsa& Opper (2022;2021); Maoutsa 等人 (2020) 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 $\beta = 0.5$ (方程 33)。这使系统适度地趋向不变密度。图 1 中的示例使用 $\beta = 1$构建，并能更好地近似转移密度，相较于 $\beta = 0.5$。

若每个振荡周期仅采集一次观测，要识别测地曲线，就必须对计算测地线的方法加以偏置，使其朝正确的运动方向识别测地线；在这种情况下，流形上两个连续观测之间的最短曲线并不总是正确的曲线。为此，我们为每个观测指定一个类相位变量，并且在计算测地线时，仅考虑相位值符合运动方向以及两个端点观测相位的观测。例如，对于逆时针旋转，若桥 $\phi_k = \phi(\mathcal{O}_k)$和 $\phi_{k+1} = \phi(\mathcal{O}_{k+1})$两个端点的相位分别为.....，则我们仅利用相位位于狭窄的 $[\phi_k, \phi_{k+1}]$ （若 $\phi_k < \phi_{k+1}$）内，或位于 $[\phi_k, 1] \cup [0, \phi_{k+1}]$ （若 $\phi_k > \phi_{k+1}$）内的观测来构建测地线。采用这种方法，我们实质上只考虑流形中与运动方向一致的相关部分。我们用函数 $\phi(\cdot): \mathcal{R}^D \to [0,1]$ 表示将相位变量赋给

每次观测。此处，对于范德波尔振子，我们考虑了 $\phi(\mathbf{x}) = \frac{\text{双参数反正切}(\mathbf{x}^{(2)}, \mathbf{x}^{(1)}) + \pi}{2\pi}$。为每次观测分配相位的另一种方法，是测量$\mathbf{x}^{(i)}$ 的希尔伯特变换与$\mathbf{x}^{(i)}$ 本身之间的夹角，并将数值缩放到 $[0,1]$ 区间。这里，$\mathbf{x}^{(i)}$ 表示观测的一个一维分量。

C 未来方向

随机动力系统所诱导的不变密度的拓扑和几何性质研究是一个相对未充分探索的领域，现有研究有限。Cong and Huang（Cong 等人，1997）从随机动力系统的角度研究了受噪声扰动系统的拓扑性质；在特定条件和假设下，这种研究可以转化为随机动力系统的表述。这一方向具有很大的后续研究潜力，尤其是结合针对随机动力系统设计的延迟嵌入方法（Stark 等人，2003），以估计不变流形的维数。

未来研究的一个潜在方向是直接采用黎曼朗之万动力学（Wang 等人，2020年）在近似度量上构造增广路径。另一种可能是，针对低采样率极限下的推断，未来可从算子学习视角来解决问题。若不考虑几何因素，为每个观测间隔构造增广路径，本质上就是在不同的初始和终端约束下，多次求解同一个控制问题。因此，神经网络算子学习方法

例如Li 等人（2020a；b）所使用的神经算子，能够推广到不同的初始条件以及可能不同的终端条件，因此有望通过路径增广，以更高的计算效率解决推断问题。

未来值得探索的方向可能包括采用薛定谔¨桥进行路径增广，以纳入含噪观测（这里我们隐含地假设观测中没有噪声，或只有较小的高斯噪声）。Tamir 等人(2023年) 提出了一种高效算法，可在薛定谔¨桥问题中引入路径约束。该算法可用于我们的框架，以在随机动力学推断中纳入几何归纳偏置。

D 附加图

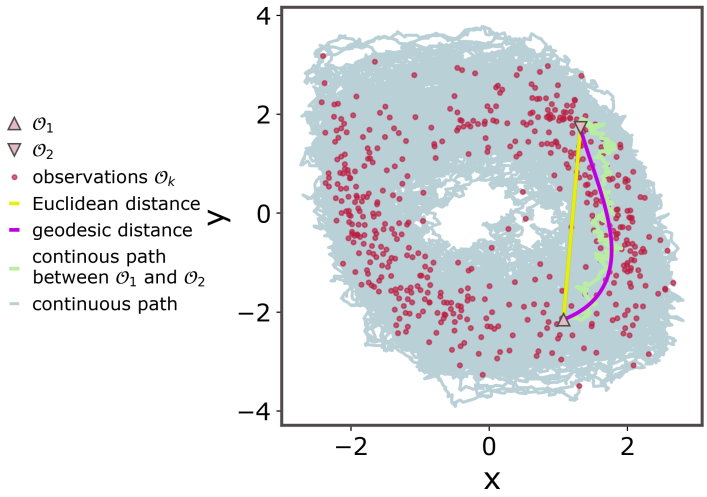


图 4：**高斯似然假设下，低频观测对应的状态增量**。欧几里得距离（黄色线）用于计算连续观测之间的状态增量，但未考虑不变密度的曲率。测地线（紫色线）能够更好地近似系统在连续观测之间的未观测状态（浅绿色线）。