面向参数依赖随机动力系统的免训练分数扩散

Minglei Yang[a,*], Sicheng He[b]

[a]橡树岭国家实验室聚变能源部，美国田纳西州橡树岭[b]田纳西大学诺克斯维尔分校机械与航空航天工程系，美国田纳西州诺克斯维尔

arXiv:2602.02113v1 [stat.ML] 2 Feb 2026

摘要

模拟参数依赖的随机微分方程（SDE）面临显著的计算挑战，因为通常需要针对每个感兴趣的参数值分别进行高保真模拟。尽管机器学习方法在学习随机微分方程动力学方面取得了成功，现有方法要么需要耗费大量计算资源进行神经网络训练以估计得分函数，要么无法处理连续参数依赖性。我们提出一种免训练条件扩散模型框架，用于学习参数依赖随机微分方程的随机流映射，其中漂移和扩散系数均依赖于物理参数。该方法的关键技术创新是一种联合核加权蒙特卡罗估计量，它利用在离散参数值处采样的轨迹数据近似条件得分函数，从而能够在状态空间和连续参数域中进行插值。训练完成后，所得生成模型无需重新训练，即可生成训练范围内任意参数值对应的样本轨迹，从而显著加速参数研究、不确定性量化和实时滤波应用。我们通过三个复杂度递增的数值算例展示了所提方法的性能，结果表明，该方法能够准确近似不同参数值下的条件分布。

关键词：基于分数的扩散模型、参数相关的随机微分方程、随机流映射、生成模型、监督学习

1. 引言

随机微分方程（SDE）是描述受随机波动影响的动力系统的基本数学模型 [1, 2]。它们自然地出现在各种科学与工程应用中，包括金融建模和各种物理系统 [3, 4]。在许多此类应用中，随机微分方程的漂移和扩散系数

*声明：本手稿由 UT-Battelle, LLC 根据与美国能源部（DOE）签订的合同 DE-AC05-00OR22725 撰写。美国政府予以保留；出版商接受本文发表，即承认美国政府保留一项非独占、费用已付清、不可撤销的全球许可，可出于美国政府目的出版或复制本手稿的已发表形式，或允许他人这样做。DOE 将根据美国能源部公共访问计划，向公众提供这些联邦资助研究成果。

*通讯作者电子邮箱：yangm@ornl.gov（Minglei Yang），sicheng@utk.edu（Sicheng He）

取决于物理参数，而这些参数可能随运行条件或实验设置而变化，例如输运过程中的扩散率或等离子体物理中的碰撞频率。了解这些系统在一系列参数下的行为，对于不确定性量化（UQ）、敏感性分析以及可靠工程系统的设计至关重要 [5]。

随机微分方程的数学意义在于，它既能通过随机轨迹提供粒子描述，也能通过相应的福克–普朗克偏微分方程提供连续介质描述 [6, 7, 8]。然而，这两种表述在数值求解时都面临巨大挑战。传统的基于粒子的方法，例如欧拉–丸山格式及其高阶变体 [9, 10], ，在需要进行长时间模拟或生成大量轨迹时，计算成本可能高得难以承受。当需要开展参数研究时，计算负担还会进一步加重，因为必须针对每个感兴趣的参数值分别进行高保真模拟。与此同时，基于偏微分方程的方法 [11, 12] 则面临数值不稳定、随维数增加而扩展性不佳，以及并行化算法困难等问题。

近年来，机器学习（ML）方法已成为学习随机动力学代理模型的有前景的替代方案。模型训练完成后，无需耗费大量资源反复进行模拟，即可在新的参数值下快速推理，因此特别适用于参数研究、不确定性量化，以及与可微卡尔曼滤波器等滤波框架的集成 [13]。这些技术涵盖范围广泛，包括物理信息神经网络 [14, 15], 、高斯过程 [16, 17], 、流映射学习 [18, 19], 、多项式近似 [20], ，以及归一化流等生成模型 [21, 22, 23, 24]。

尽管取得了这些成功，生成模型的训练仍然面临挑战，例如生成对抗网络中的模式崩溃，或归一化流中的架构限制 [25, 26]。在各种机器学习方法中，基于分数的扩散模型 [27, 28, 29] 通过学习得分函数，经由反向时间过程生成样本，取得了显著成效。文献 [30] 提出了一种免训练条件扩散模型，用于学习随机流映射；该模型的精确得分函数具有闭式表达式，可通过蒙特卡罗方法进行估计。近期的扩展工作通过退出预测网络 [31] 处理了有界区域中的随机微分方程，并开发了用于摊销贝叶斯推断的多保真方法 [32]。文献 [33] 中的误差分析表明，可解析求解的得分函数能够避免近似误差累积，其误差界在 l^2 中按 $\mathcal{O}(d)$ 的尺度增长，在 l^∞中按 $\mathcal{O}(\log d)$ 的尺度增长。

在本研究中，我们将无训练条件扩散框架扩展到参数依赖的随机微分方程，以处理漂移和扩散系数同时依赖于状态和物理参数向量的情形。其关键技术创新是一种联合核加权蒙特卡洛估计量，它利用在离散参数值处采样的轨迹数据来近似条件得分函数，从而实现跨状态空间和连续参数域的插值。我们采用两阶段流程：免训练得分估计器首先通过概率流常微分方程生成带标签的训练数据，然后利用监督回归在这些数据上训练前馈神经网络。这将具有挑战性的无监督生成建模问题转化为标准的监督学习任务，同时具备训练稳定性和架构灵活性。最终得到的框架能够学习一个统一模型，在不同参数值下生成样本轨迹，从而实现

参数研究、不确定性量化和实时滤波的快速推理。

本文其余部分安排如下：第2节介绍参数依赖的随机微分方程表述和随机流映射表示。第3节介绍无训练条件扩散框架，包括生成模型表述、来自蒙特卡洛模拟的训练数据以及免训练得分估计。第4节介绍算法与实现的细节。第5节展示数值实验，最后在第6节总结全文。

2. 问题表述

我们考虑由参数依赖的随机微分方程控制的动力系统。漂移和扩散系数同时依赖于状态和参数向量。本节介绍数学框架，包括随机微分方程表述以及用于刻画单步转移的随机流映射。

2.1. 参数依赖的随机微分方程

我们考虑以下 d 维参数依赖的随机微分方程：

$$\mathrm{d}X_t = a(X_t, \mu)\,\mathrm{d}t + b(X_t, \mu)\,\mathrm{d}W_t, \quad X_0 = x_0 \in \mathbb{R}^d, \tag{1}$$

其中，$X_t \in \mathbb{R}^d$ 是状态变量，$x_0 \in \mathbb{R}^d$ 是初始条件，$\mu \in \mathcal{M} \subset \mathbb{R}^{d_\mu}$ 是属于参数域 $\mathcal{M}$ 的参数向量；$a: \mathbb{R}^d \times \mathbb{R}^{d_\mu} \to \mathbb{R}^d$ 是漂移系数，$b: \mathbb{R}^d \times \mathbb{R}^{d_\mu} \to \mathbb{R}^{d\times m}$ 是扩散系数，$W_t \in \mathbb{R}^m$ 是一个 m 维标准布朗运动。参数 μ 可根据具体应用表示各种物理量，例如扩散率或碰撞频率 [34, 35]。

假设2.1. 我们假设以下条件对所有 $\mu \in \mathcal{M}$成立 :

(i) 函数 $a(\cdot, \mu)$和 $b(\cdot, \mu)$满足全局利普希茨条件：存在$L > 0$，使得对所有 $x, y \in \mathbb{R}^d$,

$$\|a(x, \mu) - a(y, \mu)\| + \|b(x, \mu) - b(y, \mu)\|_F \le L\|x - y\|. \tag{2}$$

(ii) 函数 $a(\cdot, \mu)$和 $b(\cdot, \mu)$满足线性增长条件：存在 $K > 0$，使得对所有 $x \in \mathbb{R}^d$,

$$\|a(x, \mu)\| + \|b(x, \mu)\|_F \le K(1 + \|x\|). \tag{3}$$

(iii) 对每个固定的 x，函数 a 和 b 关于 μ 连续。

其中 $\|\cdot\|$ 表示 l^2 范数。

在假设2.1下，对于每个初始条件 x_0 和参数 μ，存在唯一强解 $\{X_t^{x_0,\mu}\}_{t\ge 0}$ 对应于式 (1) 中的随机微分方程 [1]。这确保前向演化是良定义的，且我们希望学习的条件分布由参数值唯一确定。

2.2. 随机流映射

To 为了表述学习问题，我们引入一个均匀时间网格

$$\mathcal{T} \coloneqq \{t_n : t_n = n\Delta t,\ n = 0, 1, \ldots, N_T\}, \tag{4}$$

其中 $T > 0$ 为终止时间， $N_T \in \mathbb{Z}^+$ 为时间步数， $\Delta t = T/N_T$ 为时间步长。在每个时间区间 $[t_n, t_{n+1}]$上，式（1）中的随机微分方程可以写成条件形式：

$$X_{t_{n+1}}^{t_n,x} = x + \int_{t_n}^{t_{n+1}} a(X_s^{t_n,x}, \mu)\,\mathrm{d}s + \int_{t_n}^{t_{n+1}} b(X_s^{t_n,x}, \mu)\,\mathrm{d}W_s, \tag{5}$$

其中 $X^{t_n,\ x}_{t_{n+1}}$ 表示在给定 $X_{t_n} = x$ 的条件下，时刻 t_{n+1} 的解。为简化记号，记 $X_n \coloneqq X_{t_n}$，并定义随机流映射 $F_{\Delta t}$，使得

$$X_{n+1} = X_n + F_{\Delta t}(X_n, \mu, \omega), \tag{6}$$

其中 ω 表示布朗运动在区间 $[t_n, t_{n+1}]$ 上产生的随机性。

注 2.2. 流映射 $F_{\Delta t}(x,\ \mu,\ \omega)$ 综合体现了漂移、扩散以及参数值 μ的共同作用。与确定性流映射不同，对于每个固定的 $(x,\ \mu)$，它都是一个随机变量。根据式（5），有

$$F_{\Delta t}(x, \mu, \omega) = \int_{t_n}^{t_{n+1}} a(X_s^{t_n,x}, \mu)\,\mathrm{d}s + \int_{t_n}^{t_{n+1}} b(X_s^{t_n,x}, \mu)\,\mathrm{d}W_s. \tag{7}$$

该表示明确体现了漂移项和扩散项在整个时间区间上的积分效应。

X_{n+1} 在给定 $X_n = x_n$ 和参数 μ的条件分布，记为

$$X_{n+1} \mid (X_n = x_n, \mu) \sim p(\cdot \mid x_n, \mu), \tag{8}$$

是由 μ参数化的马尔可夫过程的转移核。学习这一依赖参数的条件分布 $p(\cdot \mid x_n,\ \mu)$是本研究的核心目标。关键挑战在于构建一个能够在整个参数域 $\mathcal{M}$内准确刻画该分布的生成模型，从而无需反复进行昂贵的模拟，即可针对任意参数值μ 高效采样。

3. 免训练条件扩散框架

在定义了参数依赖的随机微分方程及其相应的流映射后，下面介绍所提出的随机流映射学习方法。该方法由四个主要部分组成：(1) 生成模型表述；(2) 来自蒙特卡洛模拟的训练数据；(3) 免训练分数估计方法；以及 (4) 用于训练生成模型的监督学习框架。该方法将 [30, 36] 中建立的框架扩展到了参数依赖的情形。

本文所称的“免训练”特指得分函数估计：不同于通过得分匹配损失训练神经网络来近似得分函数的标准扩散模型，我们的方法推导出一个闭式表达式，该表达式可以利用轨迹数据通过蒙特卡罗方法直接进行估计。随后，通过对扩散生成的标注数据进行监督回归，学习最终的随机流映射G_θ。这一两阶段流程将具有挑战性的无监督生成建模问题转化为标准的监督学习任务。

3.1. 随机流映射的生成模型

为了近似目标分布 $p(\cdot \mid x_n,\ \mu)$，我们旨在学习一个参数化生成模型 G_θ：$\mathbb{R}^d \times \mathbb{R}^{d_\mu} \times \mathbb{R}^d \to \mathbb{R}^d$，使得

$$\hat{X}_{n+1} = X_n + G_\theta(X_n, \mu, Z), \quad Z \sim \mathcal{N}(0, I_d), \tag{9}$$

其中 $Z \in \mathbb{R}^d$ 是一个标准 d维高斯随机变量，$\theta \in \mathbb{R}^p$ 表示神经网络参数，其维度 p 由网络架构决定。神经网络 G_θ 用于近似式（ $F_{\Delta t}$ （6）中定义的随机流映射，以显式的高斯潜变量 z取代抽象随机性 ω。$\hat{}$的分布 X_{n+1} 应与真实条件分布 $p(\cdot \mid X_n, \mu)$ 对所有$(X_n, \mu) \in \mathbb{R}^d \times \mathcal{M}$均匹配。

关键挑战在于，蒙特卡洛模拟无法直接提供形式为（x_n，μ，z，x_{n+1}）的带标签训练数据：虽然我们可以观测每个参数 μ 对应的状态转移（x_n，x_{n+1}），但映射到每个特定 x_{n+1} 的潜在高斯变量 z 是未知的。我们通过使用基于分数的扩散模型生成此类带标签数据来应对这一挑战，从而构建增广数据集 $\mathcal{D}_{\text{aug}} = \{(x_n,\ \mu,\ z,\ \hat{x}_{n+1})\}$。

3.2. 蒙特卡洛模拟生成的训练数据

我们假设可以获取通过对随机微分方程（1）在多个参数值下进行蒙特卡洛模拟所生成的轨迹数据。具体而言，令 $N_\mu \in \mathbb{N}$ 表示参数值的数量，$N_s \in \mathbb{N}$ 表示每个参数对应的轨迹数。对于一组参数值 $\{\mu^{(k)}\}_{k=1}^{N_\mu} \subset \mathcal{M}$，我们对每个参数值模拟 N_s 条独立轨迹，从而得到观测数据集

$$\mathcal{D}_{\text{obs}} = \left\{ \left(x_n^{(i,k)}, \mu^{(k)}, x_{n+1}^{(i,k)} \right) : i = 1, \ldots, N_s,\ k = 1, \ldots, N_\mu,\ n = 0, \ldots, N_T - 1 \right\}. \tag{10}$$

为方便记号，我们将该数据集重新组织为一组状态转移对：

$$\mathcal{D}_{\text{obs}} = \left\{ (x_n^{(j)}, \mu^{(j)}, x_{n+1}^{(j)}) \right\}_{j=1}^{J}, \tag{11}$$

其中，$J = N_s \times N_\mu \times N_T$是训练样本对的总数，$x_n^{(j)}$和 $x_{n+1}^{(j)}$ 表示连续的状态，$\mu^{(j)}$ 是相应的参数值。核心任务是根据 $\mathcal{D}_{\text{obs}}$生成增广数据集 $\mathcal{D}_{\text{aug}}$；以下小节将介绍一种免训练扩散方法来完成这一任务。

3.3. 基于扩散的方法概述

我们采用基于分数的扩散模型来学习依赖参数的条件分布 $p(x_{n+1} \mid x_n,\ \mu)$，该分布描述了在每个固定参数值 μ下从状态 x_n 到x_{n+1} 的转移概率。核心思路是构造一个扩散过程，将该条件分布变换为标准高斯分布，然后反向执行这一过程以生成样本。该框架基于基于分数的扩散理论[37, 27], ，我们将其扩展以处理依赖参数的条件分布。

该方法分为两个阶段：

(i) 正向扩散：通过在人工时间域 $\tau \in [0,1]$ 上进行扩散过程，将条件分布 $p(\cdot \mid x_n,\ \mu)$ 转换为标准高斯分布 $\mathcal{N}(0,\ I_d)$。

(ii) 反向时间常微分方程：确定性地运行反向过程，将潜在高斯样本 $z \sim \mathcal{N}(0,\ I_d)$ 映射回来自 $p(\cdot \mid x_n,\ \mu)$ 的样本。

反向时间过程由概率流常微分方程控制：

$$\mathrm{d}Z_\tau^{x_n,\mu} = \left[f(\tau) Z_\tau^{x_n,\mu} - \frac{1}{2} g^2(\tau) S(Z_\tau^{x_n,\mu}, \tau; x_n, \mu) \right] \mathrm{d}\tau, \tag{12}$$

其中 $f(\tau) = -1/(1-\tau)$和 $g(\tau) = \sqrt{(1+\tau)/(1-\tau)}$是方差保持（VP）正向过程的漂移系数和扩散系数（推导见 附录 A ），而$S(z,\ \tau;\ x_n,\ \mu)$ 是依赖于参数的条件得分函数：

$$S(z, \tau; x_n, \mu) := \nabla_z \log Q(Z_\tau^{x_n,\mu} = z \mid x_n, \mu), \tag{13}$$

其中，$Q(Z^{x_n,}{}_\tau{}^{\mu} = z \mid x_n,\ \mu)$ 表示在给定（x_n，μ）的条件下，时刻 τ 的扩散状态边缘密度。

扩散过程作用于位移 $\Delta X_{n+1} := X_{n+1} - X_n$ 而非绝对状态 X_{n+1}。因此，常微分方程（12）建立了从$Z_1 \sim \mathcal{N}(0,\ I_d)$ 到 $Z_0 \sim p_\Delta(\cdot \mid x_n,\ \mu)$的确定性映射，其中 $p_\Delta(\cdot \mid x_n,\ \mu)$ 表示位移的条件分布。这种确定性对于生成带标签的训练数据至关重要，因为它在每个潜在高斯变量 z与其对应的位移 Δx_{n+1}之间建立了一一对应关系。前向与反向扩散过程的完整推导见附录 A。

3.4. 无需训练的分数估计

[30, 36] 的一项重要贡献是推导出条件得分函数的闭式表达式，该表达式无需训练神经网络即可直接从数据中估计。我们将这一方法扩展到依赖参数的情形。

式（13）中的得分函数可以表示为位移分布上的加权平均 p_Δ（$\Delta x \mid x_n,\ \mu$）（完整推导请参见附录 B ）：

$$S(z, \tau; x_n, \mu) = \int_{\mathbb{R}^d} \left(-\frac{z - \alpha_\tau \Delta x}{\beta_\tau^2} \right) w_\tau(z, \Delta x; x_n, \mu)\, p_\Delta(\Delta x \mid x_n, \mu)\, \mathrm{d}(\Delta x), \tag{14}$$

其中，$\alpha_\tau = 1 - \tau$和 $\beta_\tau^2 = \tau$是 VP调度缩放函数，$\Delta x = x_{n+1} - x_n$表示位移。权重函数 w_τ $(z,\ \Delta x;\ x_n,\ \mu)$ 会倾向于选择更有可能产生该扩散状态的位移值

；其显式形式见 附录 B 。这一表示说明，得分函数可以解释为对所有可能位移值对应的“局部分数” $-(z-\alpha_\tau \Delta x)/\beta_\tau^2$进行加权平均。

由于位移分布 $p_\Delta(\Delta x \mid x_n, \mu)$ 未知，我们通过蒙特卡罗估计来近似式（ S Eq. (14)）中的得分函数。如果我们拥有从 $\{\Delta x^{(\ell)}\}_{\ell=1}^N$ 精确抽取的位移样本 $p_\Delta(\cdot \mid x_n, \mu)$，原则上可以将积分替换为求和，并使用式（ w_τ Eq. (14)）中的理论权重来近似得分。然而，我们的数据集 $\mathcal{D}_{\text{obs}}$ 包含来自各种 (x_n, μ) 对的样本，这些样本不一定位于精确的查询点。

为了解决这个问题，我们从 $\mathcal{D}_{\text{obs}}$ 的联合空间中选取 (x_n, μ) 的 N 个最近邻，记为 $\{(x^{(j_\ell)}{}_n, \mu^{(j_\ell)}\ x^{(j_\ell)\ n+1})\}_{\ell=1}^N$，并计算相应的位移 $\Delta x^{(j_\ell)} = x^{(j_\ell)\ n+1} - x^{(j_\ell)n}$。随后将得分函数近似为：

$$\bar{S}(z, \tau; x_n, \mu) \coloneqq \sum_{\ell=1}^{N} \left(-\frac{z - \alpha_\tau \Delta x^{(j_\ell)}}{\beta_\tau^2} \right) \bar{w}_\tau^\ell, \tag{15}$$

其中 $\bar{w}_\tau^\ell$ 是归一化的蒙特卡罗权重，用于近似式（w_τ Eq. (14)）中的理论权重。由于邻居样本并不精确位于 (x_n, μ)，我们通过核密度估计来扩展理论权重（其中仅包含高斯转移密度 $Q(z \mid \Delta x)$），以纳入其在 x_n 和 μ：

$$\bar{w}_\tau^\ell \propto \underbrace{\exp\left(-\frac{\|z - \alpha_\tau \Delta x^{(j_\ell)}\|^2}{2\beta_\tau^2} \right)}_{\text{diffusion weight } Q(z|\Delta x^{(j_\ell)})} \cdot \underbrace{\exp\left(-\frac{\|x_n - x_n^{(j_\ell)}\|^2}{2\nu_x^2} \right)}_{\text{spatial kernel}} \cdot \underbrace{\exp\left(-\frac{\|\mu - \mu^{(j_\ell)}\|^2}{2\nu_\mu^2} \right)}_{\text{parameter kernel}}, \tag{16}$$

并进行归一化 $\sum_{\ell=1}^N \bar{w}_\tau^\ell = 1$。第一项对应于理论权重 w_τ，该权重在方程附录 B中推导得出；空间核和参数核则考虑到邻近样本源自位置接近但并不完全相同的$(x_n,\ \mu)$ 值这一事实。这里，$\nu_x > 0$和 $\nu_\mu > 0$ 分别是控制空间邻近性和参数邻近性影响的带宽参数。这种联合核加权方法能够自然地同时处理状态变量 x_n 和参数 μ 上的插值，从而实现对离散训练样本之外情况的泛化。当 τ 较小时，$\beta_\tau^2 \approx 0$会使扩散权重变得高度集中；在实践中，我们会添加一个较小的正则化项 $\delta > 0$ 至 β_τ^2 以确保数值稳定性。

4. 算法与实现

该算法首先通过求解方程 (12中的概率流常微分方程，生成带标签数据，从而建立从潜空间 $Z_1 \sim \mathcal{N}(0,\ I_d)$ 到目标分布的确定性映射：

$$\mathcal{D}_{\text{aug}} = \left\{ (x_n^{(m)}, \mu^{(m)}, z^{(m)}, \hat{x}_{n+1}^{(m)}) \right\}_{m=1}^{M}. \tag{17}$$

随后，通过监督学习训练神经网络 G_θ，以最小化定义如下的均方误差（MSE）损失 $\mathcal{L}(\theta)$：

$$\mathcal{L}(\theta) = \frac{1}{M} \sum_{m=1}^{M} \left\| G_\theta(x_n^{(m)}, \mu^{(m)}, z^{(m)}) - (\hat{x}_{n+1}^{(m)} - x_n^{(m)}) \right\|^2, \tag{18}$$

算法1用于参数相关流映射的免训练条件扩散

要求： 观测数据集 $\mathcal{D}_{\text{obs}} = \{(x_n^{(j)}, \mu^{(j)}, x_{n+1}^{(j)})\}_{j=1}^{J}$，带标签样本数 M，
ODE时间步 N_τ，邻居数 N
确保： 已训练的生成模型 G_θ

$\mathcal{D}_{\text{aug}} \leftarrow \emptyset$ ▷ 初始化增广数据集
</style}用于 $m = 1, \ldots, M$ **do**
  $(x_n^{(m)}, \mu^{(m)}) \sim \mathcal{D}_{\text{obs}}$ ▷采样查询点
  $z^{(m)} \sim \mathcal{N}(\mathbf{0}, I_d)$ ▷ 采样潜变量
  $Z_{N_\tau} \leftarrow z^{(m)}$ ▷初始化常微分方程
  对于 $k = N_\tau, N_\tau - 1, \ldots, 1$ **do**
    $\tau_k \leftarrow k/N_\tau, \quad \Delta\tau \leftarrow 1/N_\tau$ ▷扩散时间
    $\{(x^{(j_\ell)}_n, \mu^{(j_\ell)}, x^{(j_\ell)}_{n+1})\}_{\ell=1}^{N} \leftarrow$ 最近邻$(x_n^{(m)}, \mu^{(m)}, N, \mathcal{D}_{\text{obs}})$ ▷ 邻居位于 (x, μ)
    $\Delta x^{(j_\ell)} \leftarrow x^{(j_\ell)}_{n+1} - x_n^{(j_\ell)}$ 对于 $\ell = 1, \ldots, N$ ▷ 计算位移
    $\bar{S} \leftarrow \sum_{\ell=1}^{N} \bar{w}_{\tau_k}^{\ell} \cdot \left(- \frac{Z_k - (1-\tau_k)\,\Delta x^{(j_\ell)}}{\tau_k} \right)$ ▷ 得分 (15), 权重 (16) 使用 (x_n, μ) 核函数
    $Z_{k-1} \leftarrow Z_k - \Delta\tau \left[f(\tau_k) Z_k - \frac{1}{2} g^2(\tau_k)\, \bar{S} \right]$ ▷欧拉步 (12)
  结束循环
  $\hat{x}_{n+1}^{(m)} \leftarrow x_n^{(m)} + Z_0$ ▷ 位移 Z_0 加初始状态
  $\mathcal{D}_{\text{aug}} \leftarrow \mathcal{D}_{\text{aug}} \cup \{(x_n^{(m)}, \mu^{(m)}, z^{(m)}, \hat{x}_{n+1}^{(m)})\}$ ▷ 带标签数据 (17)
结束循环
$G_\theta \leftarrow$InitNN() ▷初始化神经网络
重复
  $\{(x_n^{(i)}, \mu^{(i)}, z^{(i)}, \hat{x}_{n+1}^{(i)})\}_i \subset \mathcal{D}_{\text{aug}}$ ▷ 采样小批量
  $\theta \leftarrow \theta - \eta \nabla_\theta \sum_i \| G_\theta(x_n^{(i)}, \mu^{(i)}, z^{(i)}) - (\hat{x}_{n+1}^{(i)} - x_n^{(i)}) \|^2$ ▷ 梯度步（18）
直到 $\mathcal{L}_{\text{val}}(\theta)$ 收敛 ▷ 提前停止
返回 G_θ

其中训练采用基于验证损失的提前停止 $\mathcal{L}_{\text{val}}(\theta)$，其计算方式与在 $\mathcal{D}_{\text{aug}}$的留出子集上计算类似。主要计算成本来自标注数据生成过程中的最近邻搜索；我们的实现使用 k-d 树，在 [38] 中高效执行邻居查询。常微分方程求解对每个带标签样本需要进行 N_τ 次得分函数评估。

5. 数值实验

本节通过三个复杂度逐渐增加的测试问题，展示数值实验以验证所提出的方法。示例1考察具有参数依赖漂移的布朗运动（$d = 1, d_\mu = 1$），其中 μ 控制漂移系数，可作为具有已知解析解的验证案例。示例2研究具有参数依赖漂移和扩散的奥恩斯坦-乌伦贝克过程（OU）（$d = 1, d_\mu = 1$），其中 μ 同时出现在两个系数中，且平稳方差对 μ呈现非单调依赖关系。例3考察带旋转的二维奥恩斯坦-乌伦贝克过程（$d = 2, d_\mu = 1$），展示了该方法在具有耦合动力学的多维状态空间中的能力。

所有实验均在 PyTorch [39]中实现，并在配备 NVIDIA GPU 的工作站上运行。神经网络 G_θ 采用具有双曲正切激活函数的全连接架构，使用 Adam 优化器 [40] 进行训练，学习率为 0.001，批次大小为 1024。训练基于验证损失采用提前停止，耐心值为 50 个训练轮次，并使用 90%/10% 的训练/验证集划分。具体超参数（网络规模、带标签样本数量 M、常微分方程步数 N_τ、邻居 N以及带宽参数ν_x, ν_μ）针对每个示例进行调节，并在各小节中报告。我们的实现使用k-d 树 [38]，以便在标签生成过程中高效执行邻居查询。

注 5.1（可复现性）。 所有数值实验的源代码可公开获取于 https://github.com/mlmathphy/Diffusion_parameter_SDE。本节给出的所有结果均可使用所提供的代码仓库完全复现。

5.1. 示例*1*：带参数依赖漂移的布朗运动

考虑一般的随机微分方程（1），其中 $d = m = 1$ 及其系数为：

$$a(x, \mu) = \mu, \quad b(x, \mu) = 1, \tag{19}$$

其中 $\mu \in [-1, 1]$ 是漂移参数。这得到如下的一维随机微分方程：

$$\mathrm{d}X_t = \mu\,\mathrm{d}t + \mathrm{d}W_t, \quad X_0 = x_0, \tag{20}$$

其中 W_t 是标准布朗运动。由于存在精确解析解，此例可用作验证案例。

5.1.1. 精确解

对于方程（20）中的随机微分方程，其精确解为

$$X_t = X_0 + \mu t + W_t. \tag{21}$$

单步条件分布 $p(X_{n+1} \mid X_n = x, \mu)$ 是高斯分布：

$$p(X_{n+1} \mid X_n = x, \mu) = \mathcal{N}\left(x + \mu\Delta t,\, \Delta t\right), \tag{22}$$

其条件均值和方差为：

$$\mathbb{E}[X_{n+1} \mid X_n = x, \mu] = x + \mu\Delta t, \quad \mathrm{Var}[X_{n+1} \mid X_n = x, \mu] = \Delta t. \tag{23}$$

条件均值同时线性依赖于初始状态 x和漂移参数μ，而方差仅取决于时间步长 Δt。

对于时间 T处具有高斯初始条件 $X_0 \sim \mathcal{N}(m_0,\ \sigma_0^2)$ 的终端分布，精确分布也是高斯分布：

$$X_T \sim \mathcal{N}\left(m_0 + \mu T,\, \sigma_0^2 + T\right). \tag{24}$$

这是因为 $X_T = X_0 + \mu T + W_T$，其中 X_0 和 W_T 是独立高斯随机变量。

5.1.2. 训练流程

我们使用精细的时间步长 $\delta t = 0.001$ 对式（20）中的随机微分方程进行模拟，以生成训练数据，采用 Euler–Maruyama格式。训练数据集由连续的状态对 $(X_n,\ X_{n+1})$ 组成，这些状态对在 $N_\mu = 21$ 均匀分布的参数值 $\mu \in [-1, 1]$处采样。对于每个参数值，我们在时间区间$[0, T]$ 内模拟 $N_s = 5000$条独立轨迹，其中 $T = 1.0$，记录时间步长 $\Delta t = 0.1$。初始条件从 $[-5, 5]$中均匀采样，共得到 $J = 1,\ 050,\ 000$ 个训练样本对。

按照算法1，我们使用免训练条件扩散模型生成 $M = 50{,}000$ 个带标签训练样本，并基于 (x, μ) 空间中的组合距离选取 $N = 2000$ 个最近邻。式 (12) 中的概率流常微分方程采用显式Euler格式求解，共使用 $N_\tau = 1{,}000$ 个时间步。高斯核权重的带宽参数分别设为 $\nu_x = 1.0$ （空间邻近度）和 $\nu_\mu = 0.5$（参数邻近度）。生成模型 $G_\theta(x, \mu, z)$ 参数化为全连接神经网络，输入维度为 3，包含三个隐藏层，每层有 128 个神经元，输出维度为 1。该网络预测缩放位移 $(\hat{x}_{n+1} - x_n) \cdot c_{\text{scale}}$，其中$c_{\text{scale}} = 3.0$；这种缩放使网络输出归一化至近似单位方差，从而提高训练稳定性并促进收敛。

5.1.3. 结果

图1展示了学习得到的条件分布与公式（22）中的精确解的对比。左图展示了两个参数值 $\mu = -0.5$ 和 $\mu = 0.5$ 下的概率密度函数 $p(X_{n+1} \mid X_n = 2, \mu)$。两种情况下的精确条件均值分别为 $2 + (-0.5)(0.1) = 1.95$ 和 $2 + (0.5)(0.1) = 2.05$，条件标准差均为 $\sqrt{\Delta t} = \sqrt{0.1} \approx 0.316$。使用核密度估计对生成的 50,000 个样本进行估计后，学习得到的分布（标记）与精确高斯分布（实线）高度吻合。右图展示了条件均值 $\mathbb{E}[X_{n+1} \mid X_n = 2, \mu]$ 随 μ 变化的关系。根据每个参数值对应的 10,000 个样本计算出的学习均值（标记），准确捕捉到了精确公式（实线）所预测的线性关系 $x + \mu\Delta t$。定量而言，条件均值的最大绝对误差为 0.006（平均绝对误差为 0.002），条件方差的最大相对误差为 2.5%（平均相对误差为 1.1%）。

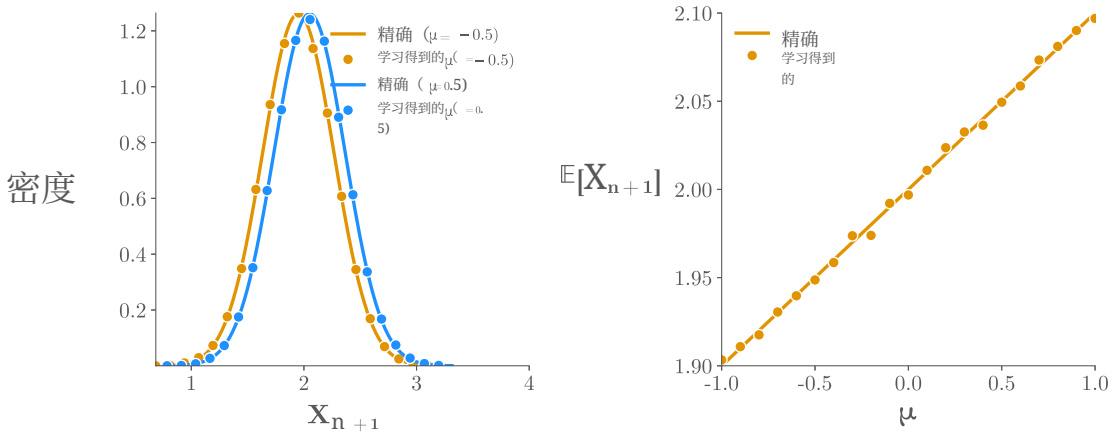

图1：示例1：条件分布 $p(X_{n+1} \mid X_n = 2,\ \mu)$。左：比较 $\mu = -0.5$和 $\mu = 0.5$的概率密度函数；实线表示精确的高斯密度（22），标记表示学习估计值。右：条件均值 $\mathbb{E}[X_{n+1} \mid X_n = 2, \mu]$ 与 μ的关系；实线表示精确公式 $x + \mu\Delta t$，标记表示学习估计值。

图2 展示了整个参数范围内条件分布 $p(X_{n+1} \mid X_n = 0, \mu)$ 的热图对比。左图展示了根据公式（22）计算出的精确分布，右图展示了学习得到的分布。随着 μ 从 -1 变化至 1，模型准确捕捉到了分布均值的变化。值得注意的是，该模型成功地对训练集中未包含的参数值进行了插值（训练集使用了 $N_\mu = 21$ 离散值），体现了该方法在连续参数空间中进行泛化的能力。

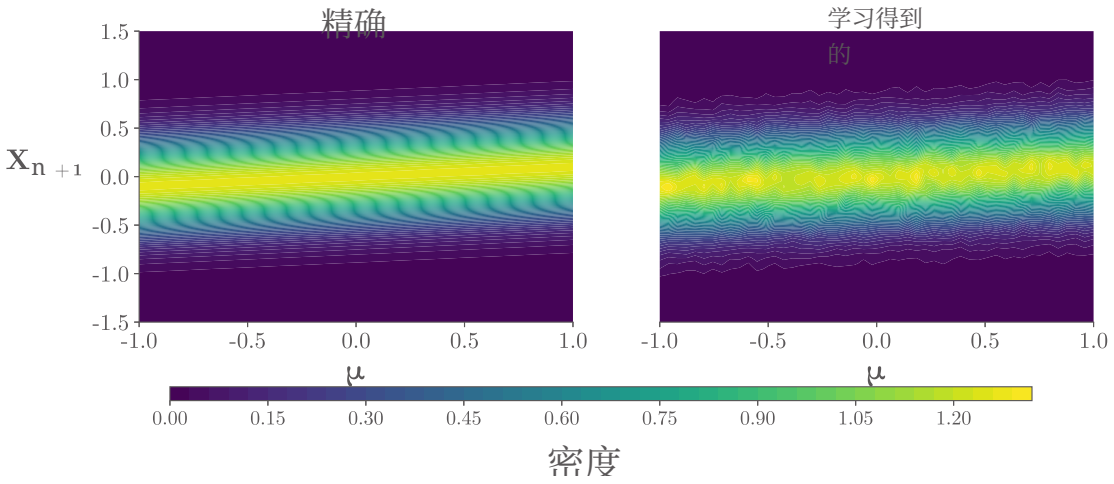

图2：示例1：条件分布 $p(X_{n+1} \mid X_n = 0,\ \mu)$ 在参数范围$\mu \in [-1, 1]$内的热图。左：由（22）得到的精确分布。右：学习得到的分布。热图使用每个5，000 个样本构建，参数取值为 μ 。学习得到的模型捕捉到了均值随 μ的线性偏移。

为展示多步预测的能力，图3 展示了从非狄拉克初始分布 $p(X_T)$ 在 $T = 1.0$ 开始时的终端分布 $X_0 \sim \mathcal{N}(0, 0.25)$。多步轨迹通过迭代应用学习得到的单步流映射生成

，经过 $N_{\text{steps}} = T/\Delta t = 10$ 步。图中比较了三种方法：(i) 使用50,000 个样本进行精确随机微分方程模拟得到的蒙特卡洛真实值（实线），(ii) 解析公式（24）（虚线），以及 (iii) 我们的学习方法，使用50,000 条生成轨迹（标记）。竖直点线表示解析均值 $m_0+\mu T$，其在 -0.5时等于 $\mu = -0.5$ （左图），在0.5 时等于 $\mu = 0.5$ （右图）。根据（24），两种情况下的终端方差均为 $\sigma_0^2+T = 0.25+1.0 = 1.25$。学习得到的分布与蒙特卡洛参考值和解析解都高度吻合，验证了学习到的流映射在长时间预测中的准确性。定量而言，两种参数值下的终端均值误差均小于0.006 ，终端标准差误差均小于0.007 。

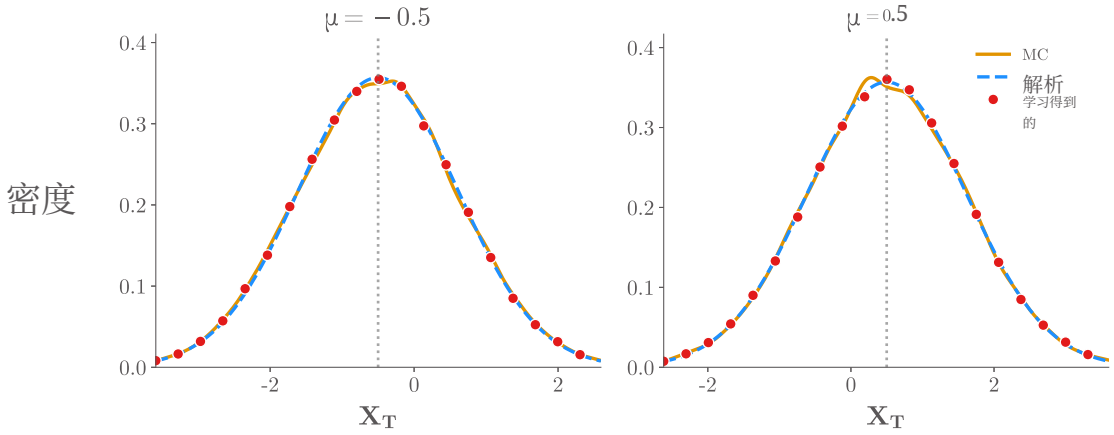

图3：示例1：终端分布 $p(X_T)$ 在 $T = 1.0$ 采用高斯初始分布 $X_0 \sim \mathcal{N}(0,\ 0。25)$，通过 10 次迭代应用学习得到的单步流映射获得。左图： $\mu = -0.5$。右图： $\mu = 0.5$。实线：蒙特卡洛真实值（50,000 个样本）；虚线：解析公式（24）；标记：学习方法（50,000 条生成轨迹）。竖直点线表示解析均值 $m_0 + \mu T$。

5.2. 示例*2*：具有参数依赖漂移和扩散的奥恩斯坦-乌伦贝克过程

为了展示当参数同时出现在漂移和扩散系数中时该方法的能力，我们考虑一般形式的随机微分方程（1），其中 $d = m = 1$ 及其系数如下：

$$a(x, \mu) = -\mu x, \quad b(x, \mu) = \sqrt{1 + \mu^2}, \tag{25}$$

其中， $\mu \in [0.5, 2]$ 同时控制均值回复率和扩散强度。这得到奥恩斯坦-乌伦贝克过程：

$$\mathrm{d}X_t = -\mu X_t \,\mathrm{d}t + \sqrt{1 + \mu^2}\,\mathrm{d}W_t, \quad X_0 = x_0. \tag{26}$$

这个例子尤其值得关注，因为平稳方差对 μ呈现非单调依赖关系。

5.2.1. 精确解

对于随机微分方程（26），精确解为

$$X_t = X_0 e^{-\mu t} + \sqrt{1+\mu^2} \int_0^t e^{-\mu(t-s)} \, \mathrm{d}W_s. \tag{27}$$

单步条件分布 $p(X_{n+1} \mid X_n = x,\ \mu)$ 为高斯分布：

$$p(X_{n+1} \mid X_n = x, \mu) = \mathcal{N}\left(x e^{-\mu \Delta t},\ \frac{1+\mu^2}{2\mu} \left(1 - e^{-2\mu \Delta t}\right) \right), \tag{28}$$

条件均值和方差为：

$$\mathbb{E}[X_{n+1} \mid X_n = x, \mu] = x e^{-\mu \Delta t}, \tag{29}$$

$$\mathrm{Var}[X_{n+1} \mid X_n = x, \mu] = \frac{1+\mu^2}{2\mu} \left(1 - e^{-2\mu \Delta t}\right). \tag{30}$$

条件均值呈指数衰减并趋于零，而方差则通过一个更复杂的表达式同时取决于 μ 和 Δt 。

平稳分布为 $\mathcal{N}(0,\ \sigma_\infty^2(\mu))$，其方差为：

$$\sigma_\infty^2(\mu) = \frac{1+\mu^2}{2\mu}. \tag{31}$$

值得注意的是，该方差关于 μ 呈非单调变化：它在 $\mu = 1$ 处达到最小值 1，并在 $\mu < 1$ 和 $\mu > 1$ 时增大。

对于时间 T 的终端分布，在高斯初始条件 $X_0 \sim \mathcal{N}(m_0,\ \sigma_0^2)$ 下，精确分布为：

$$X_T \sim \mathcal{N}\left(m_0 e^{-\mu T},\ \sigma_0^2 e^{-2\mu T} + \frac{1+\mu^2}{2\mu} \left(1 - e^{-2\mu T}\right) \right). \tag{32}$$

该表达式结合了初始条件的指数衰减与向平稳分布的趋近，并将用于验证多步预测。

5.2.2. 训练流程

我们使用Euler–Maruyama格式以精细时间步长 δt =0.001生成训练数据。数据集由位于 $N_\mu = 21$ 个均匀间隔取值$\mu \in [0.5, 2]$处的状态对(X_n, X_{n+1})组成。对于每个参数值，我们模拟 $N_s = 5000$ 条轨迹，记录时间步长为 $\Delta t = 0.1$ ，总时间为 $T = 1.0$。初始条件针对每个 μ从平稳分布(31)中采样，共计得到 $J = 1{,}050{,}000$个训练样本对。按照算法1，我们使用免训练条件扩散模型，以 $N = 1000$ 个最近邻和 $N_\tau = 500$ 个ODE时间步生成 $M = 20{,}000$个带标签样本。带宽参数为 $\nu_x = 1.0$和 $\nu_\mu = 0.3$，神经网络架构与示例1中保持相同。

5.2.3. 结果

图4展示了学习得到的条件分布与精确解(28)的比较。左图显示了 $\mu = 0.5$ 和 $\mu = 2.0$ 对应的概率密度函数 $p(X_{n+1} \mid X_n = 1, \mu)$。精确条件均值分别为 $1 \cdot e^{-0.5 \cdot 0.1} \approx 0.951$ 和 $1 \cdot e^{-2 \cdot 0.1} \approx 0.819$。均值回复效应清晰可见：较大的 μ 会导致更强地趋向于零（更小的条件均值），而扩散系数 $\sqrt{1+\mu^2}$ 也随 μ增大。学习得到的分布（标记）与精确高斯分布（实线）高度吻合。右图显示条件均值随 μ的变化，表明准确捕捉了指数衰减因子 $e^{-\mu \Delta t}$。定量而言，条件均值的最大相对误差为1.4%（均值为0.6%），条件方差的最大相对误差为2.8%（均值为1.6%）。

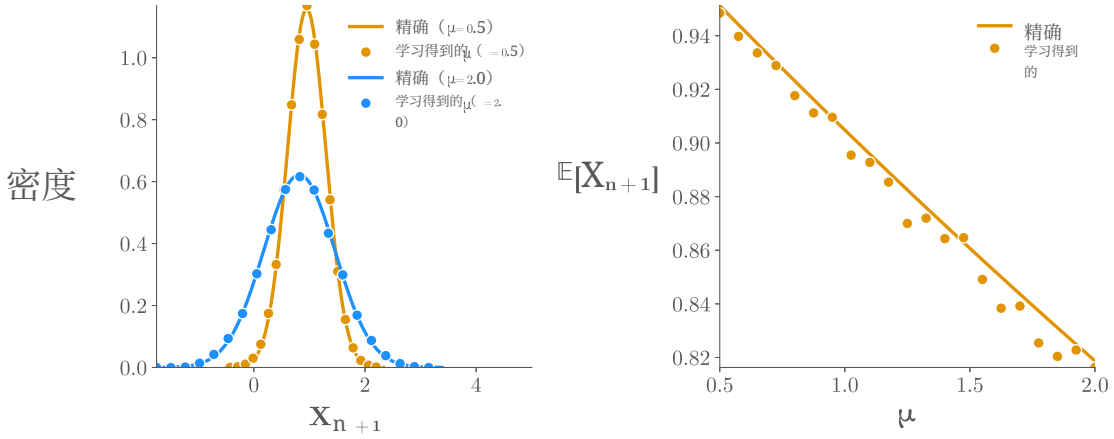

图4：示例2：条件分布 $p(X_{n+1} \mid X_n = 1, \mu)$。左：$\mu = 0.5$ 和$\mu = 2.0$的概率密度函数比较；实线表示精确的高斯密度（28），标记表示学习估计值。右：条件均值 $\mathbb{E}[X_{n+1} \mid X_n = 1, \mu]$ 与 μ的比较；实线表示精确公式 $xe^{-\mu \Delta t}$，标记表示学习估计值。

图 5 展示了从平衡点开始的条件分布 $p(X_{n+1} \mid X_n = 0, \mu)$ 的热图。单步条件方差随着μ单调增加，这可以从分布从左到右逐渐增大的分布宽度中看出。学习得到的模型在整个参数范围内准确地捕捉到了这种方差结构。

图6展示了从非平衡初始分布 $X_0 \sim \mathcal{N}(1.0, 0.25)$ 开始，在 $T = 1.0$ 时的终端分布 $p(X_T)$。使用学习到的流映射的 $N_{\text{steps}} = 10$次应用生成多步轨迹。该图比较了三种方法：蒙特卡洛真实值（实线）、解析公式（32）（虚线）以及我们提出的学习方法（标记）。竖直点线表示解析均值 $m_0 e^{-\mu T}$，对于 $\mu = 0.5$ 其值为 0.607，对于 $\mu = 2.0$其值为 0.135。可以清楚地观察到向零的均值回归，且 μ越大，回归越强。学习得到的分布与两个参考结果都高度吻合，终端均值误差低于 0.06，标准差误差低于 0.02。

图7 量化了使该示例格外有趣的非单调方差行为。左图展示了平稳方差 $\sigma_\infty^2(\mu) = (1+\mu^2)/(2\mu)$ 随 μ的变化，其在 1 处达到最小值，$\mu = 1$ （由竖直点线表示）。经过 50 个时间步的演化后，学习得到的模型准确捕捉到了这一非单调结构（50,000 个样本/ μ 值），最大相对误差为7。4%

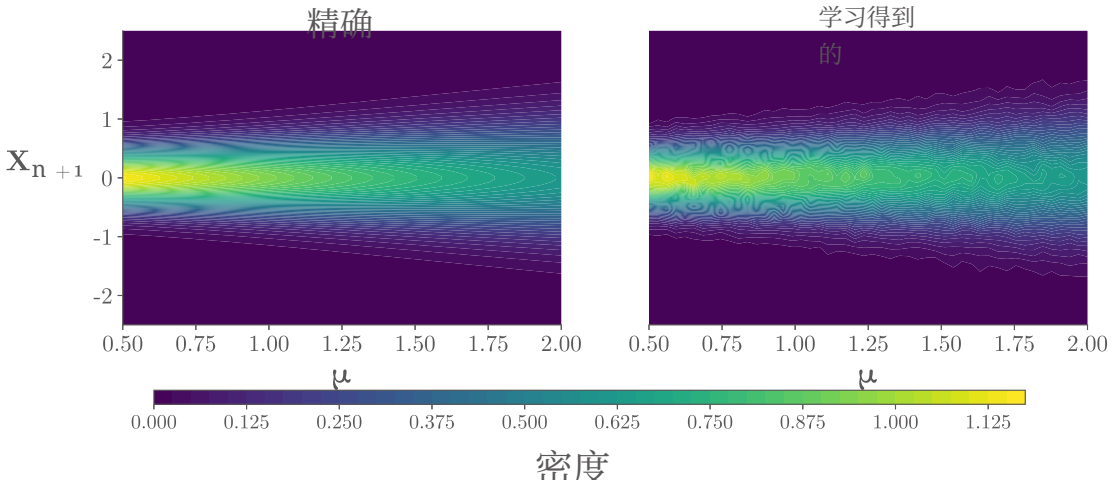

图5：示例2：条件分布 $p(X_{n+1} \mid X_n = 0,\ \mu)$ 的热图。左：精确值。右：学习得到的结果。单步条件方差随 μ 单调增加。每个 μ 值使用 20，000 个样本构建。

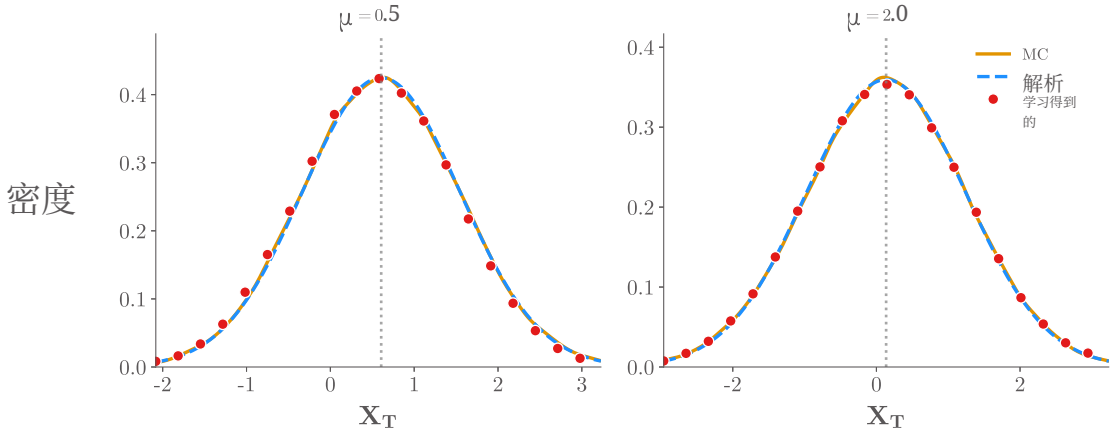

图6：示例2：终端分布 $p(X_T)$ 在 $T = 1.0$ 高斯初始分布 $X_0 \sim \mathcal{N}(1.0,\ 0.25)$，通过对学习得到的单步流映射进行 10 次迭代应用获得。实线：蒙特卡洛真实值（50，000 个样本）；虚线：解析公式（32）；标记：学习方法（50，000 条生成轨迹）。竖直点线表示解析均值 $m_0 e^{-\mu T}$。

（均值为 1.6%）。右侧面板展示了单步条件方差，其随 μ单调增加。学习得到的条件方差的最大相对误差为 3.4%（均值为 1.7%）。解析方差与学习得到的方差高度吻合，表明我们的方法能够正确捕捉参数相关动力学的均值和方差结构。

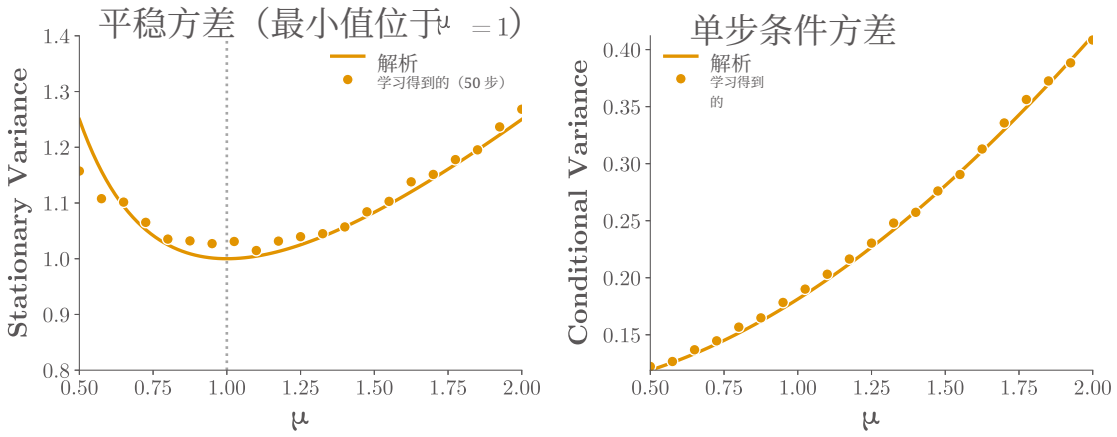

图7：示例2：方差分析。左：平稳方差 $(1+\mu^2)/(2\mu)$ 呈现非单调行为，并在 $\mu = 1$ 处达到最小值（竖直虚线）。学习得到的方差由经过 50,*000* 个时间步后的样本计算得出。右：来自 $X_n = 0$的单步条件方差，对比精确公式（28）与学习估计值（每个20,000 个样本 μ 值）。

5.3. 例*3*：带旋转的二维奥恩斯坦-乌伦贝克过程

作为展示该方法在高维情况下能力的最后一个示例，我们考虑一个带旋转的二维奥恩斯坦-乌伦贝克过程。该示例展现了真正耦合的二维动力学，其中参数同时控制轨迹的衰减率和螺旋行为。

5.3.1. 问题表述

考虑一般随机微分方程（1），其中 $d = m = 2$ 以及系数：

$$a(x,\mu) = -Ax, \quad b(x,\mu) = \sigma I_2, \tag{33}$$

其中 $\sigma = 0.5$，且漂移矩阵为：

$$A = \begin{pmatrix} \mu & -\omega \\ \omega & \mu \end{pmatrix}, \tag{34}$$

其中 $\mu \in [0.5, 2.0]$ 为衰减率参数，$\omega = 1.0$ 为固定旋转频率。由此得到二维随机微分方程：

$$\mathrm{d}X_t = -AX_t\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \quad X_0 = x_0 \in \mathbb{R}^2, \tag{35}$$

其中 W_t 是标准二维布朗运动。该矩阵结构将指数衰减（由 μ 控制）与旋转（由 ω 控制）结合起来，从而产生收敛至原点的螺旋轨迹。

5.3.2. 精确条件统计

矩阵指数 e^{-At} 可以通过解析方法计算：

$$e^{-At} = e^{-\mu t} \begin{pmatrix} \cos(\omega t) & \sin(\omega t) \\ -\sin(\omega t) & \cos(\omega t) \end{pmatrix}. \tag{36}$$

该因式分解表明，轨迹以 μ 的速率指数衰减，同时以 ω的角速度旋转。

单步条件分布为高斯分布：

$$\mathbb{E}[X_{n+1} \mid X_n = x, \mu] = e^{-A\Delta t} x, \tag{37}$$

$$\mathrm{Cov}[X_{n+1} \mid X_n = x, \mu] = \frac{\sigma^2}{2\mu} \left(1 - e^{-2\mu\Delta t}\right) I_2. \tag{38}$$

由于噪声结构具有各向同性，尽管旋转动力学不对称，协方差矩阵仍为对角矩阵，且两个维度上的方差相等。

5.3.3. 训练流程

我们通过从解析转移分布中进行精确采样来生成训练数据，设置为 $\Delta t = 0.1$、模拟时间 $T = 2.0$、在$[0.5, 2.0]$中均匀间隔分布的 $N_\mu = 16$ 个参数值，以及每个参数值对应的 $N_s = 3000$ 条轨迹，从而在 $\mathbb{R}^2$ 中得到约 960，000 个训练样本对。

按照算法1，我们使用 $M = 50,\ 000$ 个带标签样本、 $N = 2000$ 个最近邻和 $N_\tau = 500$ 个ODE时间步来生成样本。由于维度更高，我们采用更大的带宽 $\nu_x = 0.8$ 和 $\nu_\mu = 0.3$，缩放因子为 $c_{\text{scale}} = 3.0$。神经网络包含四个隐藏层，每层有256 个神经元，以适应二维动力学增加的复杂性。

5.3.4. 结果

图8 展示了三种衰减率下的样本轨迹： $\mu = 0.5$ （慢速衰减）、 $\mu = 1.0$（中速衰减）和 $\mu = 2.0$ （快速衰减）。在所有情况下，由于旋转项 ω的作用，轨迹都向内螺旋并趋近原点。学习得到的轨迹（虚线）与精确轨迹（实线）高度吻合，表明该模型捕捉到了旋转与衰减耦合的动力学。每个面板均报告平均最终半径 $\bar{r} = \frac{1}{N} \sum_{i=1}^{N} \|X_T^{(i)}\|$，该值是在终止时间计算的样本轨迹与原点之间欧几里得距离的样本平均值，涉及 $T = 5$ 条$N = 100$ 轨迹。较大的 μ 会带来更快的指数衰减，从而使 $\bar{r}$更小。精确模拟得到的 $\bar{r}_{\text{exact}}$ 与学习得到的模型得到的 $\bar{r}_{\text{learned}}$ 高度吻合，证实学习得到的模型正确捕捉了依赖参数的衰减率。

图 9 展示了从 $x_0 = (1.5, 0.5)$ 开始的三种 μ 值下的单步条件分布。顶部一行显示来自解析高斯分布的精确样本，底部一行显示学习得到的样本。学习得到的模型准确捕捉了分布的位置（均值）和宽度（协方差）；相对于起始点，均值的移动体现了其特征性的旋转。根据条件协方差公式（38），当 $\sigma^2 \Delta t$ 较小时，单步方差约为 Δt；但随着 μ 增大，方差会因因子 $1/(2\mu)$而减小。从定量结果来看，平均位置误差小于 0.024，且当 3% 时，相对方差误差小于 $\mu \leq 1$。对于 $\mu = 2$，相对方差误差

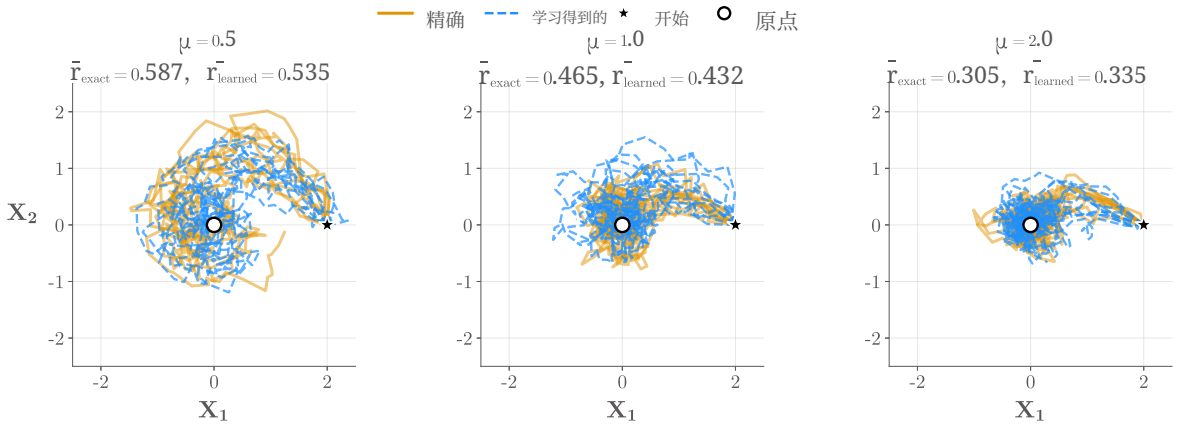

图8：例3：带旋转的二维奥恩斯坦–乌伦贝克过程从 (2，0) 开始，在 $T = 5$ 上的样本轨迹。从左至右：$\mu = 0.5$、1。0、2。0。实线：精确随机微分方程轨迹；虚线：学习得到的模型轨迹。平均最终半径 $\bar{r}$（对 100 条轨迹取平均）衡量收敛到原点的程度； μ 越大，衰减越快， $\bar{r}$ 越小。

增至 12.6%；不过，这一较大的相对误差反映的是高 μ 时较小的绝对方差（ $\mu = 2$ 时的条件方差约为 $\mu = 0.5$ 时的一半），因此相对误差对微小的绝对差异更加敏感。

图10 比较了三种起始位置下学习得到的条件均值与精确公式（37）。左图显示 X_1 分量，右图显示 X_2分量。学习得到的均值（标记）在整个参数范围内都紧密贴合精确曲线（实线），这表明旋转动力学得到了准确学习。对于所有起始位置和参数值，条件均值的最大绝对误差均小于0。05 。

图 11展示了多步行为。左图显示了 X_1 在时刻 $T = 2$ 取三个 μ 值时的边缘分布。学习得到的分布（标记）与精确的蒙特卡洛分布（实线）高度吻合。右图显示了从确定性初始条件开始的方差随时间的演化。方差从零增长并趋近于稳态值 $\sigma^2/(2\mu)$，对于0.25为 $\mu = 0.5$，对于0.125 为 $\mu = 1.0$，而对于0.0625为 $\mu = 2.0$。学习得到的方差轨迹（标记）准确跟踪了精确的解析曲线（实线），表明该模型正确捕捉了参数依赖扩散动力学。对于终端分布，均值误差低于 0.04，相对方差误差低于 8%。

该示例表明，所提出的方法能够有效扩展到具有耦合动力学的多维随机微分方程。学习得到的模型同时捕捉了旋转结构（来自 A 中的非对角项）和依赖参数的衰减率，从而生成联合二维条件分布的精确样本。

6. 结论

本文提出了一种无训练条件扩散框架，用于学习以下形式的参数依赖随机微分方程的随机流映射：$\mathrm{d}X_t = a(X_t,\ \mu)\,\mathrm{d}t + b(X_t,\ \mu)\,\mathrm{d}W_t$。其核心创新在于将 [30, 31]的免训练分数估计方法扩展至

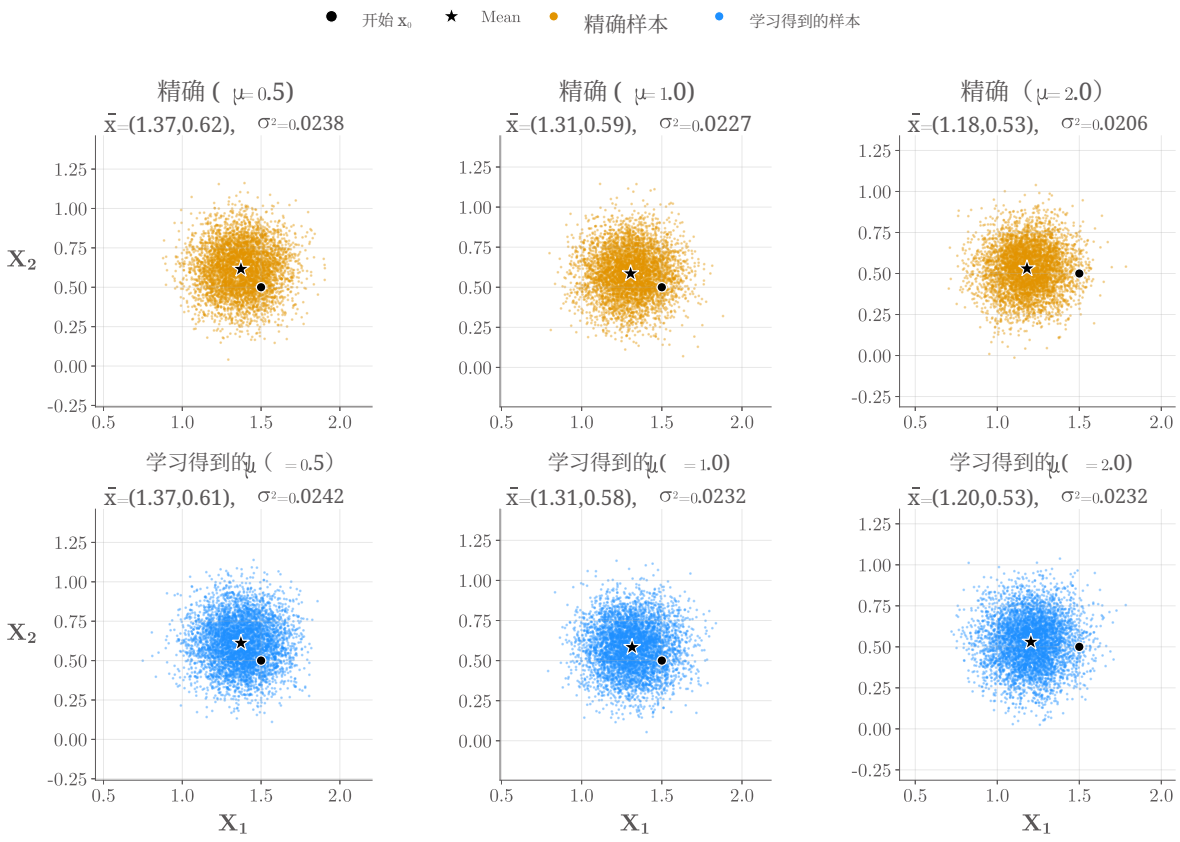

图9：例3：从(1.5,0.5)开始的单步条件分布散点图。上排：来自解析高斯分布的精确样本（37）–（38）。下排：学习模型样本（每组5,000个样本）。从左到右： $\mu = 0.5$、1.0、2.0。星号表示条件均值；黑点表示起始位置 x_0。

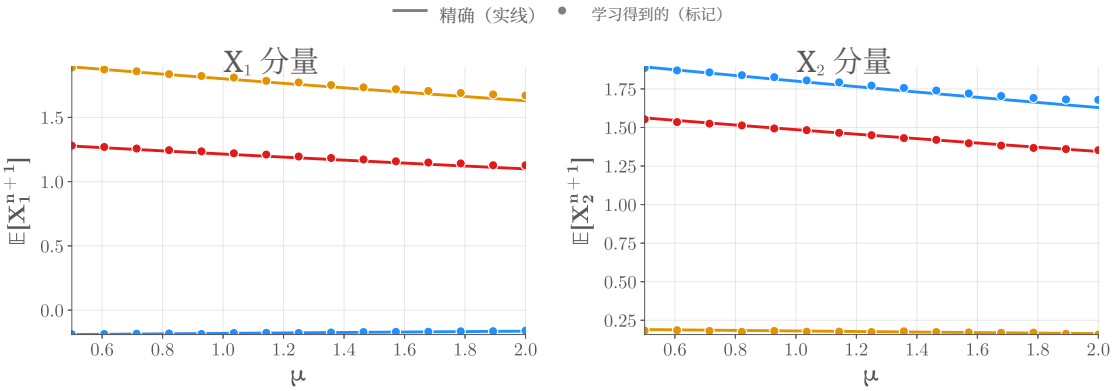

图10：例3：带旋转的二维奥恩斯坦–乌伦贝克过程的条件均值比较。左： X_1 分量。右：X_2 分量。实线：精确公式（37）。标记：学习估计值（每个点5,000 个样本）。图中显示了三个起始位置：(2,0)、(0,2)和(1.5、1.5)。

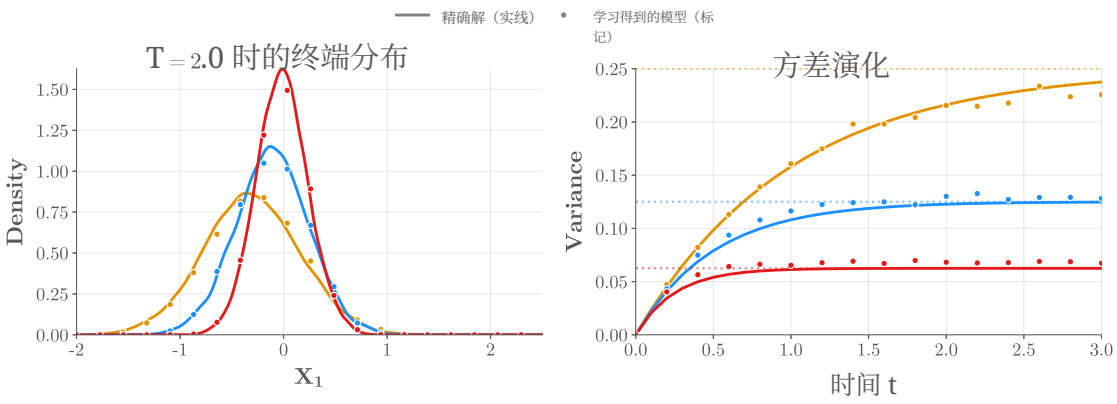

图11：例3：多步分布与方差分析。左： $\mu = 0.5$分别为 1.0、2.0 时，$T = 2$ 处 X_1 的边缘分布。右：从确定性初始条件$x_0 = (2,0)$开始的方差随时间的演化。方差从零逐渐增大，趋近平稳方差 $\sigma^2/(2\mu)$，该方差强烈依赖于 μ。实线：精确解析公式；标记：学习得到的模型。

处理参数依赖性时，条件得分函数通过经验位移的高斯核加权组合进行近似。这使得无需求解计算开销巨大的得分匹配优化问题，即可高效生成用于神经网络回归的带标签训练数据；在探索大型参数空间时，这一优势尤为明显。

我们的方法在学习参数依赖的随机动力学方面具有多项显著优势。免训练的得分估计消除了基于神经网络学习得分函数所带来的计算开销，转而采用直接利用轨迹数据的闭式蒙特卡洛估计量。通过在空间和参数两个维度上引入核加权来纳入参数依赖性，该框架能够学习一个统一模型，捕捉条件分布如何随参数 μ 变化，而无需针对每个参数值分别建立模型。两阶段方法——先通过概率流常微分方程生成带标签数据，再通过监督回归训练简单的神经网络——兼具训练稳定性和架构灵活性，这些特点难以通过端到端生成模型训练实现。

我们通过对复杂度逐渐增加的三个基准问题进行全面验证，展示了该框架的有效性。参数依赖的布朗运动（示例1）验证了基本方法，学习得到的模型在整个参数范围内，条件均值的平均绝对误差低于 0.006，条件方差的相对误差低于 3.5%。具有参数依赖扩散的奥恩斯坦-乌伦贝克过程（示例2）检验了该方法捕捉非平凡参数依赖关系的能力；在该问题中，平稳方差随均值回复率变化呈现非单调行为，而学习得到的模型以低于 3.5% 的相对方差误差复现了这一行为。带旋转的二维奥恩斯坦-乌伦贝克过程（例3）将验证扩展至多维耦合动力学；学习得到的模型既捕捉了非对角漂移项所产生的旋转结构，也捕捉了依赖参数的衰减率，条件均值误差低于 1.5%。在所有情况下，通过组合单步流映射进行多步轨迹生成，都能准确再现长时间统计量，包括终端分布和时变方差演化，这证实了

该方法具备超越单步预测进行泛化的能力。

未来工作将聚焦于将该框架扩展到高维系统，这需要为核密度估计和邻域搜索开发更高效的策略。此外，我们将研究选择带宽参数 ν_x 和 ν_μ的自适应策略；其与核密度估计中最优带宽选择的联系，可为自动调参提供理论指导。对无需训练的分数估计所引入近似误差的理论分析，将为学习到的流映射精度建立严格界限。最后，学习到的随机流映射可作为数据同化框架中的代理模型；在参数不确定的系统中，准确的参数相关动力学对于状态估计至关重要。

致谢

本材料部分基于美国能源部科学办公室先进科学计算研究办公室与聚变能源科学办公室，以及橡树岭国家实验室实验室定向研究与开发计划所支持的工作；该实验室由UT-Battelle, LLC根据合同DE-AC05-00OR22725为美国能源部运营。S. He感谢田纳西大学诺克斯维尔分校人工智能种子基金的支持。

参考文献

[1] B. Øksendal, 《随机微分方程》，载于：《随机微分方程：应用导论》，施普林格，2003年，第38–50页。[2] R. Khasminskii, 《微分方程的随机稳定性》，施普林格，2012年。[3] E. Platen、N. Bruti-Liberati, 《金融领域带跳随机微分方程的数值解》，第64卷，施普林格科学与商业媒体，2010年。[4] K. Sobczyk, 《随机微分方程：及其在物理学与工程中的应用》，第40卷，施普林格科学与商业媒体，2013年。[5] M. Grigoriu, 《随机系统：不确定性量化与传播》，施普林格科学与商业媒体，2012年。[6] V. I. Kolobov, 《等离子体和半导体中电子动力学的福克–普朗克建模》，《计算材料科学》28 (2) (2003年) 302–320。[7] A. G. Peeters、D. Strintzi, 《福克–普朗克方程及其在等离子体物理中的应用》，《物理学年鉴》520 (2-3) (2008年) 142–157。[8] M. Yang、D. del Castillo-Negrete、G. Zhang, 《一种用于α稳定莱维过程退出时间概率的高效概率方案》，arXiv预印本 arXiv:2601.09882 (2026年)。[9] C. Yuan、X. Mao, 《马尔可夫切换随机微分方程欧拉–丸山方法的收敛性》，《数学与计算机模拟》64 (2) (2004年) 223–235。

[10] M . J. P伊戈特, V. Solo, G几何欧 l尔–丸山随 hemes f机随机h性 diff微分 l

[11] F. Liu、V. Anh、I. Turner，空间分数阶福克–普朗克方程的数值解，《计算与应用数学杂志》166 (1) (2004年) 209–219。[12] P. Patie、C. Winter，多维扩散过程的首次退出时间概率：一种基于偏微分方程的方法，《计算与应用数学杂志》222 (1) (2008年) 42–53。[13] Y. Wu、S. He，DKFNet：用于场反演与机器学习的可微卡尔曼滤波器，arXiv预印本 arXiv:2509.07474 (2025)。[14] L. Yang、D. Zhang、G. E. Karniadakis，用于随机微分方程的物理信息生成对抗网络，《SIAM科学计算期刊》42 (1) (2020) A292–A317。[15] X. Chen、L. Yang、J. Duan、G. E. Karniadakis，利用福克–普朗克方程与物理信息神经网络从离散粒子观测求解逆随机问题，《SIAM科学计算期刊》43 (3) (2021) B811–B830。[16] C. Archambeau、D. Cornford、M. Opper、J. Shawe-Taylor，随机微分方程的高斯过程近似，载于：Gaussian Processes in Practice，PMLR，2007，第1–16页。[17] M. Opper，随机微分方程的变分推断，《物理学年鉴》531 (3) (2019) 1800233。[18] J. Chen、K. Wu、D. Xiu，Due：用于建模未知方程的深度学习框架与库，《SIAM评论》67 (4) (2025) 873–902。[19] Y. Chen、D. Xiu，通过流映射算子学习随机动力系统，《计算物理学期刊》508 (2024) 112984。[20] S. W. Werner、B. Peherstorfer，从数据稳定线性动力系统的样本复杂度，《计算数学基础》24 (3) (2024) 955–987。[21] M. S. Albergo、E. Vanden-Eijnden，利用随机插值构建归一化流，arXiv预印本 arXiv:2209.15571 (2022)。[22] M. Yang、P. Wang、D. del Castillo-Negrete、Y. Cao、G. Zhang，面向具有多种初始分布的随机动力系统的伪可逆归一化流，《SIAM科学计算期刊》46 (4) (2024) C508–C533。[23] J. Urain、M. Ginesi、D. Tateo、J. Peters，Imitationflow：通过归一化流学习深度稳定随机动力系统，载于：2020年IEEE/RSJ智能机器人与系统国际会议（IROS），IEEE，2020，第5231–5237页。

[24] M. Yang、P. Wang、M. Fan、D. Lu、Y. Cao、G. Zhang，用于量化不确定性传播中替代建模的条件伪可逆归一化流，《建模与计算机器学习期刊》6 (4) (2025)。[25] T. Salimans、I. Goodfellow、W. Zaremba、V. Cheung、A. Radford、X. Chen，改进的生成对抗网络训练技术，《神经信息处理系统进展》29 (2016)。[26] I. Kobyzev、S. J. Prince、M. A. Brubaker，归一化流：当前方法的介绍与综述，《IEEE模式分析与机器智能汇刊》43 (11) (2020) 3964–3979。[27] Y. Song、J. Sohl-Dickstein、D. P. Kingma、A. Kumar、S. Ermon、B. Poole，通过随机微分方程进行基于分数的生成建模，国际学习表征会议 (2021)。[28] A. Q. Nichol、P. Dhariwal，改进的去噪扩散概率模型，载于：国际机器学习会议，PMLR，2021，第8162–8171页。[29] L. Yang、Z. Zhang、Y. Song、S. Hong、R. Xu、Y. Zhao、W. Zhang、B. Cui、M.-H. Yang，扩散模型：方法与应用的综合综述，《ACM计算调查》56 (4) (2023) 1–39。[30] Y. Liu、Y. Chen、D. Xiu、G. Zhang，用于学习随机动力系统的免训练条件扩散模型，《SIAM科学计算期刊》47 (5) (2025) C1144–C1171。[31] M. Yang、Y. Liu、D. Del-Castillo-Negrete、Y. Cao、G. Zhang，用于学习有界区域内随机动力系统流映射的生成式人工智能模型，《计算物理学期刊》(2025) 114434。[32] C. Tatsuoka、M. Yang、D. Xiu、G. Zhang，使用条件扩散模型的多保真度参数估计，arXiv预印本 arXiv:2504.01894 (2025)。[33] P. Wang、Z. Zhang、M. Yang、F. Bao、Y. Cao、G. Zhang，用于高维采样的免训练扩散模型的误差估计，arXiv预印本 arXiv:2601.19740 (2026年)。[34] Z. Wang、J. Xin、Z. Zhang，利用保结构格式计算混沌流与随机流的有效扩散率，《SIAM数值分析期刊》56 (4) (2018) 2322–2344。[35] A. V. Nguyen、D.-A. An-Vo、T. Tran-Cong、G. M. Evans，浮选中湍流效应对气泡-颗粒相互作用的随机描述综述，《国际矿物加工杂志》156 (2016) 75–86。[36] Y. Liu、M. Yang、Z. Zhang、F. Bao、Y. Cao、G. Zhang，面向密度估计的生成模型的扩散模型辅助监督学习，《建模与计算机器学习期刊》5 (1) (2024)。

[37] B. D. Anderson, 《逆时扩散方程模型》, 《随机过程及其应用》12（3）（1982）313–326。[38] J. L. Bentley, 用于关联搜索的多维二叉搜索树, 《ACM 通讯》18（9）（1975）509–517。[39] A. Paszke、S. Gross、F. Massa、A. Lerer、J. Bradbury、G. Chanan、T. Killeen、Z. Lin、N. Gimelshein、L. Antiga 等, 《PyTorch：一种命令式风格的高性能深度学习库》, 载于：《神经信息处理系统进展》, 第 32 卷, 2019 年, 第 8026–8037 页。[40] D. P. Kingma、J. Ba, 《Adam：一种随机优化方法》, arXiv 预印本 arXiv:1412.6980（2014）。

[41] J. Ho、A. Jain、P. Abbeel, 《去噪扩散概率模型》, 载于：神经信息处理系统进展, 第 33 卷, 2020 年, 第 6840–6851 页。

附录 A. 前向与反向扩散过程

本附录为第 3 节中使用的参数相关条件扩散模型提供理论基础。

附录 *A.1.* 前向扩散过程

对于固定的一对 $(x_n,\ \mu)$, 令 $\Delta X^{x_n,\ \mu}_{n+1} = X^{x_n,\ \mu}_{n+1} - x_n$ 表示位移, 其中 $X^{x_n,\ \mu}_{n+1} \sim p(\cdot \mid x_n,\ \mu)$服从条件分布。遵循 [30], , 我们在人工域 $\tau \in [0,1]$上对位移定义一个前向扩散过程：

$$\mathrm{d}Z_\tau^{x_n,\mu} = f(\tau) Z_\tau^{x_n,\mu}\,\mathrm{d}\tau + g(\tau)\,\mathrm{d}W_\tau, \quad Z_0^{x_n,\mu} = \Delta X_{n+1}^{x_n,\mu}, \tag{A.1}$$

其中 W_τ是标准布朗运动（与原始随机微分方程独立）, 而 $f(\tau)$ 和$g(\tau)$分别是前向扩散过程的漂移和扩散系数。

我们选择系数, 使前向过程具有解析上可处理的转移密度。具体而言, 我们将（A.1）的解参数化为

$$Z_\tau^{x_n,\mu} = \alpha_\tau Z_0^{x_n,\mu} + \beta_\tau \epsilon, \quad \epsilon \sim \mathcal{N}(0, I_d), \tag{A.2}$$

其中 α_τ 和 β_τ 是满足 $\alpha_0 = 1$、 $\beta_0 = 0$的确定性缩放函数。为使该参数化与（A.1）一致, 系数必须满足：

$$f(\tau) = \frac{\mathrm{d}\log\alpha_\tau}{\mathrm{d}\tau}, \quad g^2(\tau) = \frac{\mathrm{d}\beta_\tau^2}{\mathrm{d}\tau} - 2\frac{\mathrm{d}\log\alpha_\tau}{\mathrm{d}\tau}\beta_\tau^2. \tag{A.3}$$

这些关系确保 Z_τ 的边缘分布遵循所规定的缩放。在本工作中, 我们采用方差保持（VP）调度 [41]:

$$\alpha_\tau = 1 - \tau, \quad \beta_\tau^2 = \tau, \quad \tau \in [0,1]. \tag{A.4}$$

采用这一选择后, 我们有 $\alpha_1 = 0$ 和 $\beta_1 = 1$, 因此 $Z^{x_n,\ \mu}_1 \sim \mathcal{N}(0,\ I_d)$ 与初始位移分布 $p_\Delta(\cdot \mid x_n,\ \mu)$ 无关。

由（A.2）可知，给定 $Z^{x_n,\mu}_0 = \Delta x$ 时 $Z^{x_n,\mu}_\tau$ 的条件分布为高斯分布：

$$Q(Z^{x_n,\mu}_\tau \mid Z^{x_n,\mu}_0 = \Delta x) = \mathcal{N}(\alpha_\tau \Delta x, \beta^2_\tau I_d). \tag{A.5}$$

该结果可直接由（A.2）的线性性推出：给定 $Z^{x_n,\mu}_0 = \Delta x$，我们有 $Z^{x_n,\mu}_\tau = \alpha_\tau \Delta x + \beta_\tau \epsilon$ 其中 $\epsilon \sim \mathcal{N}(0, I_d)$，得到一个均值为 $\alpha_\tau \Delta x$、协方差为 $\beta^2_\tau I_d$ 的高斯分布。

附录 *A.2.* 反向时间随机微分方程与常微分方程

基于分数的扩散理论的关键结果是，前向扩散过程（A.1）存在相应的反向时间过程。从 $\tau = 1$到 $\tau = 0$ 运行反向过程，会将 $\mathcal{N}(0, I_d)$ 中的样本变回目标位移分布 $p_\Delta(\cdot \mid x_n, \mu)$中的样本。

根据 [37], ，反向时间随机微分方程为

$$\mathrm{d}Z^{x_n,\mu}_\tau = \left[f(\tau) Z^{x_n,\mu}_\tau - g^2(\tau) S(Z^{x_n,\mu}_\tau, \tau; x_n, \mu)\right] \mathrm{d}\tau + g(\tau)\, \mathrm{d}\overleftarrow{W}_\tau, \tag{A.6}$$

其中 $\overleftarrow{W}_\tau$是反向布朗运动，而 $S(z, \tau; x_n, \mu)$ 是式（13）中定义的依赖于参数的条件得分函数。注意，得分函数取决于四个量：扩散状态 z、扩散时间 τ、初始状态 x_n和参数μ；对 μ 的依赖是相较于 [30]中的标准条件扩散模型的关键扩展。

一个重要的观察来自 [27]：反向时间随机微分方程（A.6）与第 $Q(Z^{x_n,\mu}_\tau \mid x_n, \mu)$ 中介绍的确定性概率流常微分方程（12）具有相同的边缘分布。该常微分方程定义了从 $Z_1 \sim \mathcal{N}(0, I_d)$ 到$Z_0 \sim p_\Delta(\cdot \mid x_n, \mu)$ 的确定性映射，这对于生成带标签的训练数据至关重要，可确保潜变量与位移输出一一对应。

附录 B. 得分函数推导

本附录完整推导了第 3 节中使用的参数依赖条件得分函数的闭式表达式。

给定 (x_n, μ) 时 $Z^{x_n,\mu}_\tau$ 的边缘密度可通过对所有可能的初始位移值积分得到：

$$Q(Z^{x_n,\mu}_\tau = z \mid x_n, \mu) = \int_{\mathbb{R}^d} Q(Z^{x_n,\mu}_\tau = z \mid Z^{x_n,\mu}_0 = \Delta x)\, p_\Delta(\Delta x \mid x_n, \mu)\, \mathrm{d}(\Delta x). \tag{B.1}$$

代入高斯转移密度（A.5），可得

$$Q(z \mid x_n, \mu) = \int_{\mathbb{R}^d} \frac{1}{(2\pi\beta^2_\tau)^{d/2}} e^{-\frac{\|z - \alpha_\tau \Delta x\|^2}{2\beta^2_\tau}} p_\Delta(\Delta x \mid x_n, \mu)\, \mathrm{d}(\Delta x). \tag{B.2}$$

该积分表示将边缘密度表示为高斯混合，其权重由位移分布 $p_\Delta(\Delta x \mid x_n, \mu)$ 给出。

为推导得分函数，我们应用对数导数：

$$S(z, \tau; x_n, \mu) = \nabla_z \log Q(z \mid x_n, \mu) = \frac{\nabla_z Q(z \mid x_n, \mu)}{Q(z \mid x_n, \mu)}. \tag{B.3}$$

对B.2关于 z求梯度:

$$\nabla_z Q(z \mid x_n, \mu) = \int_{\mathbb{R}^d} \left(-\frac{z - \alpha_\tau \Delta x}{\beta_\tau^2} \right) Q(z \mid \Delta x)\, p_\Delta(\Delta x \mid x_n, \mu)\, \mathrm{d}(\Delta x), \tag{B.4}$$

其中 $Q(z \mid \Delta x) \coloneqq Q(Z^{x_n,\ \mu}_{\tau} = z \mid Z^{x_n,\ \mu}_{0} = \Delta x) = \mathcal{N}(z;\ \alpha_\tau \Delta x, \beta_\tau^2 I_d)$。

将（B.2）和（B.4）代入（B.3），可得精确的得分函数:

$$S(z, \tau; x_n, \mu) = \frac{\int_{\mathbb{R}^d} \left(-\frac{z - \alpha_\tau \Delta x}{\beta_\tau^2} \right) Q(z \mid \Delta x)\, p_\Delta(\Delta x \mid x_n, \mu)\, \mathrm{d}(\Delta x)}{\int_{\mathbb{R}^d} Q(z \mid \Delta x')\, p_\Delta(\Delta x' \mid x_n, \mu)\, \mathrm{d}(\Delta x')}. \tag{B.5}$$

该表达式表明，得分是位移分布上加权积分的比值。

通过定义权重函数，可以将其改写为更直观的形式:

$$w_\tau(z, \Delta x; x_n, \mu) \coloneqq \frac{Q(z \mid \Delta x)}{\int_{\mathbb{R}^d} Q(z \mid \Delta x')\, p_\Delta(\Delta x' \mid x_n, \mu)\, \mathrm{d}(\Delta x')}, \tag{B.6}$$

它满足归一化条件 $\int w_\tau\, p_\Delta(\Delta x \mid x_n, \mu)\, \mathrm{d}(\Delta x) = 1$。

采用这一记法，得分函数即为正文中的公式（14）。这种表示揭示了得分函数是所有可能位移值对应的“局部分数”的加权平均 $-(z - \alpha_\tau \Delta x)/\beta_\tau^2$，其中权重较高的位移更有可能产生当前扩散状态 z。