arXiv:2302.03663v1 [cs.LG] 7 Feb 2023

SDYN-GANs：适用于一般阶随机动力学多步生成模型的对抗学习方法

Panos Stinis
高级计算、数学与数据部 太平洋西北国家实验室（*PNNL*） 华盛顿州里奇兰

Constantinos Daskalakis
电气工程与计算机科学系 麻省理工学院（*MIT*） 马萨诸塞州剑桥

Paul J. Atzberger ATZBERG@GMAIL.COM
数学系 机械工程系 加州大学圣巴巴拉分校（*UCSB*） 圣巴巴拉，加利福尼亚州 *93106*，美国
http://atzberger.org

摘要

我们提出了用于数据驱动生成建模的对抗学习方法，以刻画 n^{th} 阶随机系统的动力学。我们的方法基于生成对抗网络（GANs），其生成模型类建立在稳定的 m 步随机数值积分器之上。我们提出了不同的表述和训练方法，通过观测轨迹样本来学习随机动力学模型。我们开发了多种方法，包括采用基于最大均值差异（MMD）的判别器、使用条件分布和边缘分布的训练方案，以及学习不同时间尺度上的动态响应的方法。我们展示了如何利用这些方法为物理系统建模，从而学习力定律、阻尼系数和噪声相关参数。这些对抗学习方法能够为动态任务构建稳定的生成模型，包括长期预测，以及为随机系统开发仿真。

关键词：对抗学习、生成对抗网络（GANs）、降维、降阶建模、系统辨识、随机动力系统、随机数值方法、广义矩方法（GMM）、最大均值差异（MMD）、科学机器学习。

1. 引言

统计推断、机器学习、工程学和科学领域中的学习与建模任务，往往需要为呈现非线性随机动力学的系统推断其表征。这要求估计出的模型不仅能够解释观测到的行为，还能支持在更长

时间尺度上进行仿真或预测。要获得稳定的长期仿真和预测，需要采用数据驱动方法，从观测数据中学习结构化动力学模型。

在统计学和工程学领域，随机动力学参数与表征的估计由来已久（Peeters and De Roeck, 2001；Hurn et al., 2007；Jensen and Poulsen, 2002；Kalman, 1960；Nielsen et al., 2000）。相关方法包括推断机器人学（Olsen et al., 2002）、轨道力学（Muinonen and Bowell, 1993）、地质学（Pardo-Ig´uzquiza, 1998）、政治学（King, 1998）、金融学（Kim et al., 1998）和经济学（Cramer, 1989）中的模型与参数。开发估计量最重要的策略之一，是采用最大似然估计（MLE）（Fisher and Russell, 1922；Hald, 1999）或贝叶斯推断（BI）（Gelman et al., 1995）。对于能够轻易指定似然函数且可对其进行有效计算的系统，可以开发出具有渐近近最优数据效率的估计量（Wasserman, 2010）。然而，在实践中，为最大似然估计指定并计算似然函数通常面临重大挑战，尤其是对于随机过程（Hurn et al., 2007；Sørensen, 2004）。对于贝叶斯方法，计算和分析后验分布这一众所周知的难题又使情况更加复杂（Gelman et al., 1995；Smith and Roberts, 1993）。对于随机系统，观测数据集合的高维度，以及由系统轨迹构成的数据，也使问题进一步复杂化。

作为最大似然估计方法的替代方案，我们基于矩估计法的推广来构建对抗学习方法，采用判别器／评论器（Hansen，1982；Gretton 等，2012年；Fortet 和 Mourier，1953）。我们采用生成建模方法，开发用于学习模型和估计参数的方法；该方法通过比较候选模型的样本与观测数据来实现。此类基于样本的对抗方法的一个显著特点是，不再需要指定似然函数。这使得我们能够在更广泛的生成模型类别上进行学习。现在的关键要求变为：确定一组足够丰富的样本特征，以便判别器／评论器能够将其与观测数据进行比较，从而进行模型选择。我们基于特征核和广义矩开发方法，以区分一般类别的分布（Gretton 等，2012年；Fortet 和 Mourier，1953）。

我们介绍对抗学习方法及训练方法，用于借助一类基于m步随机数值方法的生成模型，获取 n^{th}阶随机过程的表示。针对基于二阶统计量的损失函数进行基于梯度的学习，以及轨迹观测产生的高维样本，我们开发了专门的伴随方法和自定义反向传播方法。我们的方法提供了生成对抗网络（GANs），用于学习随机系统非线性动力学的表示。我们将这些方法称为随机动力学生成对抗神经网络（SDYN-GANs）。

2. 对抗学习

我们提出对抗学习方法，旨在量化候选生成模型所生成的一组样本与训练数据之间的差异。在对抗学习中，生成模型 $G(;\theta_G)$ 与判别模型 $D(;\theta_D)$ 一同学习。这可以视为一场双人博弈。其目标是

判别器 D 尽可能减少错误，将生成器生成的一组样本与训练数据区分开来。生成器的目标则是生成质量足够高的样本，使判别模型陷入困惑，无法准确判断这些样本是否来自训练数据，参见图1。根据损失函数以及判别模型和生成模型的模型类别的不同选择，可以得到不同类型的散度和度量。这些选择会影响如何比较生成模型的分布与训练数据的分布。在实践中，人们寻求仅利用这些分布中的样本也能取得良好表现的学习方法。

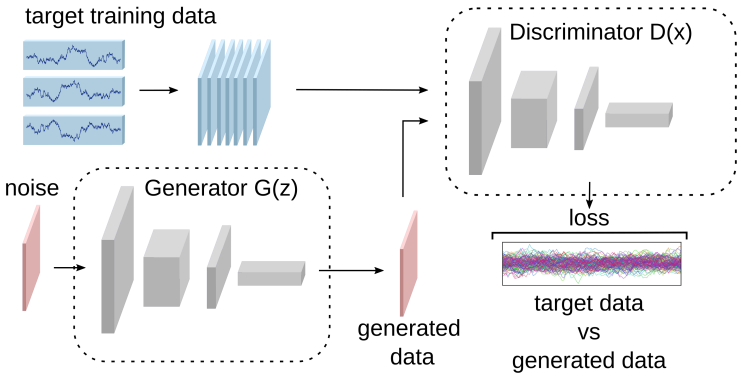

图1：**对抗学习**。生成模型 $G(\cdot;\theta_G)$ 和判别器/评论器$D(\cdot;\theta_D)$ 用于从轨迹数据的样本中学习模型。 D用于计算损失，以比较由 G 生成的样本 $\{\tilde{\mathbf{x}}_i\}_{i=1}^N$ 与训练数据中的样本 $\{\mathbf{x}_i\}_{i=1}^N$。

2.1 生成对抗网络（GANs）

在 GANs 的对抗训练中，第一个模型 G 是一个表示隐式生成模型的神经网络，记为 $G(\mathbf{Z};\theta_G)$，它从随机源 $\mathbf{Z}$生成样本。 θ_G 是该模型的参数，$\mathbf{Z}$ 表示一个在 G 下进行映射以生成样本 $\tilde{\mathbf{x}}^{[i]} = G(\mathbf{Z}^{[i]};\theta_G)$ 的随机变量。第二个模型 D 用于判断由 G生成的一组样本 $\{\tilde{\mathbf{x}}^{[i]}\}_{i=1}^N$ 与训练数据集中的样本 $\{\mathbf{x}^{[i]}\}_{i=1}^N$ 的匹配程度。设生成器的经验分布记为 $\tilde{\mu}_G$，训练数据的经验分布记为 $\tilde{\mu}_{\mathcal{D}}$。判别器/评论器模型记为 $D(\cdot;\theta_D)$。 D 的形式选择以及损失函数 ℓ 决定了如何比较两组样本的经验概率分布， $\tilde{\mu}_G$， $\tilde{\mu}_{\mathcal{D}}$。GANs 及相关统计估计量有许多种构建方式（Sriperumbudur 等人，2009；Goodfellow 等，2016）。

在生成对抗网络的原始表述中，目标函数可以表示为$\theta_D^* = \arg\min_{\theta_D} J_D(\theta_G,\theta_D)$，其中 $J_D = \mathbb{E}_{(x,y)\sim p_{\text{synth}}}[-\log\, p_{\text{D}}(y|x;\theta_D)]$，（Goodfellow 等，2014年、2016）。 $p_{\text{synth}}(x,y) = \frac{1}{2}p_{\text{data}}(x,y)+\frac{1}{2}p_{\text{model}}(x,y;\,\theta_G)$ 是通过将生成器的样本与 $\tilde{x} \sim p_{\text{model}} \sim \tilde{\mu}_G$ 混合而得到的合成数据分布的概率

训练数据 $x \sim p_{\text{data}} \sim \tilde{\mu}_{\mathcal{D}}$，（Goodfellow 等，2016）。生成数据样本的标签设为 $y = -1$，训练数据的标签设为 $y = +1$。对于给定的 θ_D，我们令 $D(x) = p_{\text{D}}(y = +1|x;\theta^D)$ 表示判别器分类器将样本 x 判定为标签 $y = +1$的概率。因此，我们有 $1 - D(x) = p_{\text{D}}(y = -1|x;\theta_D)$。采用这些选择和记号后，目标函数现在可以表示为 $J_D = -\frac{1}{2}\mathbb{E}_{x\sim p_{\text{data}}}[\log(D(x))] - \frac{1}{2}\mathbb{E}_{x\sim p_{\text{model}}(;\theta_G)}[\log(1 - D(x))]$.目标是让生成器生成高质量样本，使得即使是最优判别器 θ_D^* 也会感到困惑，并给出值 $D(x) = p_{\text{D}}(;\theta_D^*) = 1/2$。

此时最优生成模型的目标函数为 $\theta_G^* = \arg\min_{\theta_G} J_G(\theta_G,\theta_D^*(\theta_G))$。当 $J_G = -J_D$ 时，这可以被视为一个代价函数为$\ell = J_D$ 、其中 θ_D^* 和θ_G^* 分别对应最优策略的双人零和博弈。如果将目标函数完全优化至均衡状态，则对于比较生成器 $\tilde{\mu}_G$和训练数据 $\tilde{\mu}_{\mathcal{D}}$的经验分布，这将得到詹森-香农（JS）散度$JS(\tilde{\mu}_{G(;\theta_G)}\ \tilde{\mu}_{\mathcal{D}})$（Lin，1991）。最优生成器应为 $\theta_G^* = \arg\min_{\theta_G} JS(\tilde{\mu}_{G(;\theta_G)}\ \tilde{\mu}_{\mathcal{D}})$，其中 $JS(\tilde{\mu}_G,\tilde{\mu}_{\mathcal{D}}) = \frac{1}{2}D_{KL}\left(\tilde{\mu}_{\mathcal{D}}\|\frac{1}{2}(\tilde{\mu}_G + \tilde{\mu}_{\mathcal{D}})\right) + \frac{1}{2}D_{KL}\left(\tilde{\mu}_G\|\frac{1}{2}(\tilde{\mu}_G + \tilde{\mu}_{\mathcal{D}})\right)$，$D_{KL}$为库尔贝克-莱布勒散度（Cover-和 Thomas，2006；Kullback 和 Leibler，1951）。在实践中，为了缓解梯度学习方面的问题，可以通过在目标函数中使用不同的损失函数来放宽零和条件，并采用 $J_G \neq -J_D$。例如，有时会使用 $J_G = -\frac{1}{2}\mathbb{E}_{x\sim p_{\text{model}}(;\theta_G)}[\log(D(x))]$作为生成器 G的目标，以改善梯度（Goodfellow 等，2016）。此类公式可能导致其他类型的博弈，其更一般的解可表征为纳什均衡或其他稳态（Nisan 等，2007；Owen，2013；Nash，1951；Roughgarden，2010）。

生成对抗网络和其他统计估计量也可以通过基于 $J_D(\theta_G,\theta_D) = \mathbb{E}_{\tilde{\mu}_{G(;\theta_G)}}[F(X';\theta_D)] - \mathbb{E}_{\tilde{\mu}_{\mathcal{D}}}[F(X;\theta_D)]$ 以及 $J_G = -J_D \text{with} \theta_G^* = \arg\min_{\theta_G} J_G(\theta_G,\theta_D^*(\theta_G))$ 和 $\theta_D^* = \arg\min_{\theta_D} J_D(\theta_G,\theta_D)$ 的判别器/评论器来表述。从统计学角度看，这个目标函数可视为产生了一个积分概率度量（IPM）（Sripe-rumbudur 等人，2009；Muller，1997）。IPM 可表示为

$$\theta_G^* = \arg\min_{\theta_G}\max_{f\in\mathcal{F}}\left(\mathbb{E}_{\tilde{\mu}_{\mathcal{D}}}[f(X)] - \mathbb{E}_{\tilde{\mu}_{G(;\theta_G)}}[f(X')]\right)$$

$= \arg\min_{\theta_G}\max_{f\in\mathcal{F}}\ell(\theta_G;\ f)$ $= \arg\min_{\theta_G}\ell^*(\theta_G)$，其中 $\mathcal{F} = \{f \mid \exists\theta_D \text{ s.t. } f = F(;\theta_D)\}$ 以及 $\ell^*(\theta_G) = \max_{f\in\mathcal{F}}\ell(\theta_G;\ f)$。对于 $\mathcal{F} = \text{Lip-1}$，即由满足 $L \leq 1$ 的利普希茨函数构成的函数类，这会得到$\ell^*(\theta_G) = \max_{f\in\text{Lip-1}}\ell(\theta_G;\ f) = \mathcal{W}^1(\tilde{\mu}_G,\tilde{\mu}_{\mathcal{D}})$。这由 Kantorovich-Rubinstein 对偶性得出，其中 $\mathcal{W}^1$ 是经验测度 $\tilde{\mu}_G$ 和 $\tilde{\mu}_{\mathcal{D}}$ 之间的 1-瓦瑟斯坦距离（Villani，2008）。虽然瓦瑟斯坦距离为分析测度提供了广泛使用的范数，但在实际训练中，对其进行近似的计算开销可能很大，而且可能导致难以用于学习的梯度（Mallasto 等人，2019年）。

作为这些方法的替代方案，对于 IPM，我们采用一类更平滑的判别器，其 $\mathcal{F} \subset \mathcal{H}$ 由 $\mathcal{H}$ 再生核希尔伯特空间（RKHS）构成（Aronszajn，1950；Berlinet 和 Thomas-Agnan，2011）。通过选择核 $k(\cdot,\ \cdot)$，我们还可以影响比较过程中重点关注的分布特征。我们的目标函数$\mathcal{F} = \{f \in \mathcal{H} \mid \|f\|_{\mathcal{H}} \leq 1\}$为 $\theta_G^* = \arg\min_{\theta_G}\max_{f\in\mathcal{F}}\ell(\theta_G;\ f)$，其中$\ell(\theta_G;\ f) = \mathbb{E}_{\tilde{\mu}_{\mathcal{D}}}[f(X)] - \mathbb{E}_{\tilde{\mu}_{G(;\theta_G)}}[f(X')]$。

该目标函数可以通过解析方法部分求解，得到 $\ell^*(\theta_G) = \max_{f\in\mathcal{F}}\ell(\theta_G;\ f) = \|\eta_{\mathbb{P}} - \eta_{\mathbb{Q}}\|_{\mathcal{H}}$，其中 $\eta_{\mathbb{P}} = E_{X\sim\mathbb{P}}[k(\cdot,\ X)]$, $\eta_{\mathbb{Q}} = E_{X'\sim\mathbb{Q}}[k(\cdot,\ X')]$，而 $\|\cdot\|_{\mathcal{H}}$ 是再生核希尔伯特空间的范数（Sriperumbudur 等人，2010年）。这里，我们将 $\mathbb{P} = \tilde{\mu}_{\mathcal{D}}$ 视为训练数据的经验

分布，并将 $\mathbb{Q} = \tilde{\mu}_G$ 视为生成样本的经验分布。以这种方式刻画概率分布之间的差异，对应于广义矩估计法（Hansen，1982）和最大均值差异（MMD）的统计方法（Fortet 和 Mourier，1953；Smola 等，2006年）。

2.2 随DYN-GANs：随机动力学的对抗学习框架 cs

我们将介绍用于学习随机动力学生成模型的整体对抗训练方法，并在后续章节中提供更多技术细节。学习能够进行长期预测的随机动力学表示颇具挑战，因为传统估计方法未考虑时间因素，难以直接应用。这是由于轨迹数据样本维度很高，而且需要构建能够预测超出已观测动力学持续时间的稳定模型。要获得可靠的学习方法，就必须有效利用随机系统和轨迹数据的特殊结构，包括马尔可夫条件、概率依赖关系及其他属性。

尽管我们的方法适用范围更广，但此处重点关注与向量值伊藤随机过程（Oksendal，2000年）相关的生成模型学习。其动机来自具有未知特征的系统，这些系统可以通过随机微分方程（SDEs）的解来建模（Oksendal，2000年；Gardiner，1985）。从概率图模型的角度来看（Wainwright 和 Jordan，2008；Bishop，2006年；Sucar，2015年；Murphy，2012年），与轨迹观测数据相关的高维概率分布可以进行因子分解。这样便可利用依赖关系构建更局部的模型，以耦合各组成部分随时间的变化，从而大幅降低问题的维度。为获得适用于预测和长期仿真的生成模型 G ，我们基于具有可学习组成部分的稳定 m步随机数值积分器，开发了生成器类（Kloeden 和 Platen，1992）。正如我们将在后续章节中讨论的那样，我们通过研究物理系统和 $\mathbb{R}^3$中的力学问题，阐述方法的动机并考察其性能。在这种情况下，我们还能进一步细化数值离散化类别，以纳入物理原理，例如动力学中保守贡献项的近似时间可逆性、涨落-耗散平衡及其他属性（Verlet，1967；Gronbech-Jensen 和 Farago，2013；Jasuja 和 Atzberger，2022；Tabak 和 Atzberger，2015；Trask 等，2020；Lopez 和 Atzberger，2022）。

我们开发了随机动态生成对抗网络（SDYN-GANs），用于学习生成模型 G ，以采样离散化轨迹 $\mathrm{X}^{[i]} = G(\mathrm{Z}^{[i]}, \theta_G)$。$\mathrm{Z}^{[i]}$ 表示噪声源的样本，经 G 映射后生成样本 $\mathrm{X}^{[i]}$。数据样本可表示为 $\mathrm{X}^{[i]} = \left(\mathrm{X}_1^{[i]}, \ldots, \mathrm{X}_N^{[i]}\right)$，其中 $\mathrm{X}_j^{[i]} = \mathrm{X}^{[i]}(t_j)$ 是时间分量。我们学习具有 m-步随机数值积分方法形式的生成器 G ： $\mathrm{X}_j = \boldsymbol{\Psi}_{j-1}(\mathrm{X}^{j-1}, \ldots, \mathrm{X}^{j-m}, \boldsymbol{\omega}_{j-1}; \mathrm{p})$，其初始条件为 $\mathrm{X}_0 = \mathrm{x}_0(\mathrm{p}) \ldots, \mathrm{X}_{m-1} = \mathrm{x}_{m-1}(\mathrm{p})$。可训练项 $\boldsymbol{\Psi}_{j-1}$ 利用时刻 t_{j-1}、 ...、 t_{j-m} 的信息，近似随机动力学从时间步长 t_{j-1} 到 t_j 的演化。数值更新通过样本 $\boldsymbol{\omega}_{j-1}$ 进行随机化，并具有可学习参数 p。对于 $G(\mathrm{Z}; \theta_G)$，本文主要考虑 $\mathrm{Z} = \{\boldsymbol{\omega}_j\}_j$ 和$\theta_G = \mathrm{p}$ 的情形。在一般情况下，Z 还可以包含额外的噪声源以及 θ_G 额外参数。使用 m-步随机方法进行生成建模，使我们能够

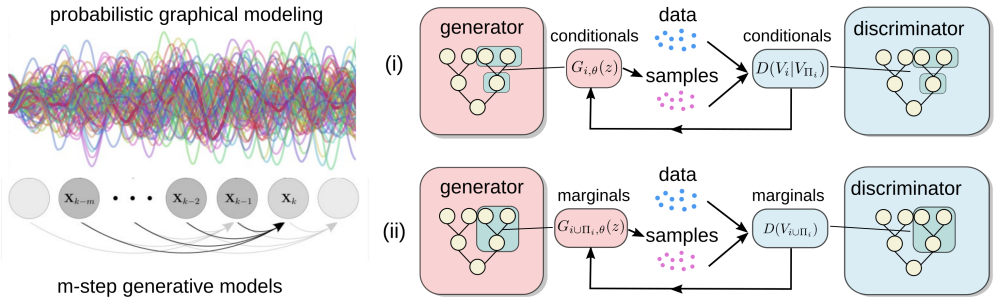

图2：**用于训练的分解**。 我们利用轨迹中构型$\mathbf{X}(t)$之间的概率依赖关系，帮助从时间序列数据（左）中学习模型。我们采用（i）条件统计量和（ii）边际统计量，设计了不同的训练方案。我们还针对不同的轨迹持续时间 $[0,\mathcal{T}]$ 以及跨时间尺度的子采样 τ （右），设计了相应的训练方法。

利用概率依赖关系对高维轨迹数据进行因子分解。我们提出的 m-步方法可视为概率图模型中结构方程的一种变体（Bishop，2006年；Murphy，2012年；Wainwright 和 Jordan，2008；Ding 等，2021年），见图2。

我们使用对抗学习来训练生成模型 G，通过判别器/评论器D 区分由 G生成的样本与训练数据中的样本。我们通过优化损失 $\theta_G^* = \arg\min_{\theta_G} \ell^*(\theta_G)$来获得生成器，其中 $\ell^*(\theta_G) = \ell(\theta_G, \theta_D^*(\theta_G))$。如第2.1节所述，这可以用许多不同的方式来表述，分别对应经验分布的不同散度和度量。我们使用损失 ℓ^*^ ^，其对应于如下形式的积分概率度量（IPM）： $\ell^*(\theta_G) = \ell^*(\mathbb{P},\mathbb{Q}) = \max_{f\in\mathcal{F}} (\mathbb{E}_{\mathbb{P}}[f(\mathbf{X})] - \mathbb{E}_{\mathbb{Q}}[f(\mathbf{X}')])$，其中 $\mathbb{P}$、 $\mathbb{Q}$ 表示进行比较的测度。我们取 $\mathbb{P} = \tilde{\mu}_{\mathcal{D}}$、$\mathbb{Q} = \tilde{\mu}_G$，并使用 $\mathcal{F} = \{f \in \mathcal{H} \mid \|f\|_{\mathcal{H}} \leq 1\}$，其中 $\mathcal{H}$是一个以 $k(\cdot,\cdot)$为核的再生核希尔伯特空间（RKHS）。通过这些选择，我们可以将生成对抗网络及其对应的积分概率度量损失表示为 $\ell^*(\theta_G) = \|\eta_{\mathbb{Q}} - \eta_{\mathbb{P}}\|_{\mathcal{H}}$，其中 $\eta_{\mathbb{P}} = \mathbb{E}_{\mathbb{P}}[k(\cdot,\mathbf{X})]$ 和$\eta_{\mathbb{Q}} = \mathbb{E}_{\mathbb{Q}}[k(\cdot,\mathbf{X})]$。这可以看作是将每个测度嵌入为再生核希尔伯特空间中的函数（Saitoh 和 Sawano，2016；Sriperumbudur 等人，2010）。它们之间差异的刻画可以表示为其嵌入表示的 $\mathcal{H}$-范数。由于 $k(\cdot,\mathbf{X})$给出了一组特征，这些特征在取期望后会生成广义矩，因此该方法对应于广义矩方法（GMM）（Hansen，1982）和最大均值差异（MMD）（Gretton 等，2012；Fortet 和 Mourier，1953；Smola 等，2006）。

在实践中，为了获得更便于优化和计算梯度的损失函数，我们采用一种重新表述方法，将目标函数平方，得到 $\ell^*(\theta_G) = \left(\hat{}\ell^*(\theta_G)\right)^2 = \|\eta_{\mathbb{P}} - \eta_{\mathbb{Q}}\|_{\mathcal{H}}^2 = \mathbb{E}_{\mathbf{X}\sim\mathbb{P}}[k(\mathbf{X},\mathbf{X})] - 2\mathbb{E}_{\mathbf{X}\sim\mathbb{P},\,\mathbf{X}'\sim\mathbb{Q}}[k(\mathbf{X},\mathbf{X}')] + \mathbb{E}_{\mathbf{X}'\sim\mathbb{Q}}[k(\mathbf{X}',\mathbf{X}')]$。这与能量统计（Sz´ekely 和 Rizzo，2013）以及（Gretton 等，2012）中提出的基于样本的估计量有关。为了找到用于训练 G 的近似极小值点，我们采用随机

梯度下降（SGD）方法 $\theta_G^{n+1} = \theta_G^n - \eta_n \hat{q}$，学习率为 η_n，梯度估计量 $\hat{q} \approx \nabla_{\theta_G} \ell^*(\theta_G^n)$ 基于小批次（Robbins 和 Monro，1951；Bottou 等人，1991年；Bishop，2006年）。我们采用包含动量的变体，例如 ADAM 及其他方法（Kingma 和 Ba，2014年；Tieleman 和 Hinton，2012年；Lydia 和 Francis，2019年）。为了计算 $\hat{q}$，我们将（Gretton 等，2012）中的无偏估计量用于梯度近似。我们使用 $\hat{q} = \nabla_{\theta_G} \ell^*(\theta_G)$ 估计经验损失的梯度，其中

$$\tilde{\ell}^*(\theta_G) = \frac{1}{N(N-1)} \sum_{i \neq j}^{N} k\left(\mathbf{X}^{[i]}, \mathbf{X}^{[j]}\right) - \frac{2}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} k\left(\mathbf{X}^{[i]}, \mathbf{Y}^{[j]}\right) + \frac{1}{M(M-1)} \sum_{i \neq j}^{M} k\left(\mathbf{Y}^{[i]}, \mathbf{Y}^{[j]}\right). \tag{1}$$

小批次的样本具有分布 $\mathbf{X}^{[i]} \sim \tilde{\mu}_G$ 和 $\mathbf{Y}^{[i]} \sim \tilde{\mu}_{\mathcal{D}}$。

轨迹样本 $\mathbf{X}^{[i]} = G(\mathbf{Z}^{[i]}; \theta_G)$ 取决于 θ_G，这进一步要求采用计算梯度 $\nabla_{\theta_G} \mathbf{X}^{[i]}$贡献的方法。当 $\mathbf{p} \subset \theta_G$ 时，由于这些梯度涉及 $m-$步随机积分器，因而可能带来挑战。如果使用基于记录随机积分器前向˜评估以及 ℓ^*二阶统计量的直接反向传播，这一过程可能会变得尤其昂贵。这将导致规模大到难以承受的计算调用图。这是由于轨迹 $\mathbf{X}$ 的高维性以及生成过程中多个相互依赖的步数所导致的。小批次采样 $\{\mathbf{X}^{[i]}\}_{i=1}^{n_b}$ 和二阶统计量进一步加剧了这一问题。正如我们将在后文第 4.2 节和第 4.3 节中讨论的那样，我们可以通过开发专门利用损失函数结构的伴随方法和定制反向传播方法来解决这些问题。这样，我们便能够高效地计算 $\nabla_{\theta_G} \mathbf{X}^{[i]}$ 对 $\nabla_{\theta_G} \tilde{\ell}^*$的贡献，从而获得训练所需的梯度估计。

我们还介绍并研究了 SDYN-GANs 的训练方案，利用轨迹数据的几种不同方式来比较底层概率分布，见图2。这包括使用以下方式进行比较：(i) 整个轨迹上的分布；(ii) 轨迹片段上的边缘分布；以及 (iii) 轨迹片段上的条件分布。我们针对 n^{th} 阶随机系统开发了以下训练方法

全轨迹方法（***full-traj***）：从分布 $p_\theta(\mathbf{R})$ 中采样完整轨迹 $\mathbf{R} = (\mathbf{X}(t_0), \mathbf{X}(t_0 + \delta t) \ldots, \mathbf{X}(t_0 + (m-1)\delta t), \mathbf{X}(t_1), \mathbf{X}(t_1 + \tau) \ldots, \mathbf{X}(t_1 + (N_\tau - 1)\tau))$，对应于 N_τ 步的轨迹。

边缘方法（***marginals***）：从具有以下形式的轨迹片段 $\mathbf{R}_k$ 中采样边缘分布：$\mathbf{R}_k = (\mathbf{X}(t_k), \mathbf{X}(t_k + \delta t) \ldots, \mathbf{X}(t_k + (m-1)\delta t), \mathbf{X}(s_k), \mathbf{X}(s_k + \tau) \ldots, \mathbf{X}(s_k + (\ell - 1)\tau))$，$k \in [1, N_m]$，其边缘分布为 $\{p_{\theta,k}(\mathbf{R}_k)\}_{k=1}^{N_m}$。

条件方法（条件分布）：对轨迹片段 $\mathrm{R}_k|\mathrm{Q}_k$ 进行条件采样，条件依据为形式如下的片段 Q_k：$\mathrm{R}_k = (\mathrm{X}(s_k), \mathrm{X}(s_k + \tau) \ldots, \mathrm{X}(s_k + (\ell - 1)\tau))$ 和 $\mathrm{Q}_k = (\mathrm{X}(t_k), \mathrm{X}(t_k + \delta t) \ldots, \mathrm{X}(t_k + (m-1)\delta t))$，$k \in [1, N_m]$，其条件分布为 $\{p_{\theta,k}(\mathrm{R}_k|\mathrm{Q}_k)\}_{k=1}^{N_m}$。

在这些方法中，t_k 是当前 n^{th}阶动力系统在训练数据中所表示的时间。δt 对应于用于采样初始动力状态信息的时间尺度

。 s_k 给出了我们希望预测动力学的轨迹部分。 τ 是对演化动力学进行采样以用于训练和预测的时间尺度。实践中，我们通常会有 $n = m$；然而，即使在 $n = 1$ 的情况下，我们的公式也允许 $m > 1$ ，以便使用随机多步数值方法，例如亚当斯-巴什福斯方法（Kloeden 和 Platen，1992年；Buckwar 和 Winkler，2006年；Ewald 和 T´emam，2005年）。

在实践中，常见的选择是令 $\delta t = \Delta t$，并令 $\tau = a\Delta t$ ，以便 $a \in \mathbb{N}$ ，从而使生成模型的随机轨迹与观测数据之间的时间采样对齐。我们将 $t_k = k\Delta t$ 和 $s_k = t_k + \tau$ 作为默认选择。原则上，只要提供一种插值方法来比较训练数据与生成模型的时间样本，也可以采用更一般的选择。我们将在下文各节讨论更多技术细节。我们在第 5 节中研究 SDYN-GANs，采用不同的训练方案 (i)–(iii) 来学习物理模型参数和力定律。

3. 相关工作

此前已有一些方法被提出，用于训练生成对抗网络，以建模确定性和随机动力学，包括神经常微分方程/随机微分方程（Chen 等，2018；Kidger 等，2021），以及受物理学问题启发的方法（Xu 等，2022；de Avila Belbute-Peres 等，2018；Stinis 等，2019；Yang 等，2020，2022）。近期的生成对抗网络方法与工程领域（Lions，1971；Giles 和 Pierce，2000；Ljung，1987；C´ea，1986）和金融领域（Kaebe 等，2009；Giles 和 Glasserman，2006）发展的伴随方法和系统辨识方法密切相关。在这些生成对抗网络方法中，最大似然估计被替换为以下方法之一：使用神经网络分类器判别器，对最初的生成对抗网络 Jensen-Shannon（JS）度量进行近似（Goodfellow 等人，2014）；或使用带梯度惩罚的神经网络（WGAN-GP）（Gulrajani 等，2017），对 Wasserstein 度量进行近似（Arjovsky 等，2017）；或采用针对边缘切片的瓦瑟斯坦度量（Deshpande 等，2019）。在时间序列分析中，概率图模型也与生成对抗网络结合以增强训练，并通过建立次可加性界来刻画，见（Ding 等，2021，2020）。

众所周知，使用通用神经网络判别器训练生成对抗网络颇具挑战。这会表现为训练过程中的振荡，并伴随收敛性和稳定性方面的已知问题（Kodali 等，2017；Thomas Pinetz 和 Pock，2018）。虽然神经常微分方程/随机微分方程已被用于梯度训练（Chen 等，2018；Tzen 和 Raginsky，2019），但其梯度计算采用了近似，可能导致结果与所选离散化所表示的随机动力学不一致。这些方法也主要针对一阶系统的动力学。

在我们的方法中，我们针对一般n^{th}阶动力学开发了一套用于计算梯度的伴随方法。在训练过程中，为确保稳定性和相容性，明确考虑所采用的 m步离散化方法的选择尤为重要。我们的方法确保训练梯度与结构化 m步生成模型类所指定的损失函数相一致。尽管我们的对抗学习框架具有普适性，但我们通过重点采用 D 来减少训练中的振荡及其他问题，在这些方法中，判别器具有解析可解性。我们的 D 基于一组特征映射

这些特征映射可视为采用核方法，将概率分布编码并嵌入再生核希尔伯特空间（RKHS）中（Gretton 等，2012年；Sriperumbudur 等人，2010年；Saitoh 和 Sawano，2016）。候选模型的样本与观测数据之间的比较对应于一组与广义矩方法（GMM）（Hansen，1982）和最大均值差异（MMD）（Gretton 等，2012年；Fortet 和 Mourier，1953）相关的损失函数。

我们采用这种损失函数的选择，与此前应用于图像生成问题的 GANs-MMD 方法相关（Li et al., 2015年；Binkowski et al., 2018；Dziugaite et al., 2015年）。我们的工作针对 n^{th} 阶随机动力学的生成模型，这类模型带来了新的挑战，包括时间因素、高维轨迹样本以及高阶动力学。Abbati et al. (2019年) 开展了使用最大均值差异学习随机微分方程模型的相关工作。该工作的重点是标量值的一阶随机微分方程，并通过分解动力学来学习漂移项，同时使用高斯过程回归（GPR）（Williams 和 Rasmussen, 1996）。他们将高斯过程回归用作平滑器，以构建漂移贡献项的估计量。对于其常数协方差项，他们将比例常数视为高斯过程回归先验分布的超参数之一，并通过最大似然估计对其进行优化（Abbati et al., 2019年）。

在本研究中，我们开发了处理更一般的向量值 n^{th} 阶随机动力学的方法，并同时学习漂移和扩散的贡献。受微尺度力学问题的启发，我们展示了如何运用这些方法学习向量值二阶惯性随机动力学的表示。我们还基于 m 步随机数值方法，引入了结构化的隐式生成模型类别。

为训练模型，我们开发了伴随方法，以获得与所采用的 m 步随机离散化具体选择相一致的梯度。我们还开发了定制的反向传播过程，以考虑损失函数的二阶统计量。鉴于轨迹数据本身具有高维特性，我们还开发了训练方法，通过按时间尺度 τ 对轨迹进行子采样来降低维度。我们的学习方法可以调整训练对观测数据作出响应的时间尺度，从而有助于实现时间滤波和粗化方法。我们还开发了不同的训练方案，其统计准则基于随机动力学的条件分布或边缘分布。总之，我们的方法提供了灵活的基于样本的方法，可用于在一般的生成模型类别上进行训练，从而学习随机系统 n^{th} 阶动力学的结构化表示。

4. 使用 SDYN-GANs 学习生成模型

SDYN-GANs 提供了数据驱动方法，用于学习能够表示随机系统 n^{th} 阶动力学的生成模型。如第 2.2 节所述，我们使用基于积分概率度量（IPM）的对抗学习损失 $\ell^*(\theta_G)$，其中 $\mathcal{F} = \{f \in \mathcal{H} \mid \|f\|_{\mathcal{H}} \leq 1\}$ 分别对应于广义矩方法（GMM）（Hansen，1982）和最大均值差异（MMD）（Gretton 等，2012；Fortet 和 Mourier，1953；Smola 等，2006）。SDYN-GANs 中的学习只需要生成器和训练数据的样本，从而为可使用的生成模型类别提供了灵活性。我们

开发了用于动态生成模型的方法 G，这些模型基于 m 步随机离散化方法来表示轨迹样本。这需要在训练期间计算高维样本的二阶统计量损失的梯度。我们利用损失 ℓ^* 的结构、概率依赖关系，并通过开发伴随方法，设计了实用的计算方法。这为计算训练 SDYN-GANs 所需的梯度提供了高效方法。

4.1 G 基于 m-步随机积分方法的生成建模

我们基于随机数值积分方法（Kloeden 和 Platen，1992；Jasuja 和 Atzberger，2022；Atzberger 等人，2007），为若干类生成模型 G 开发了相应方法。对于 n^{th} 阶随机动力系统，我们考虑如下形式的一般随机 m 步数值方法：$\mathrm{X}_j = \boldsymbol{\Psi}_{j-1}(\mathrm{X}^{j-1}, \ldots, \mathrm{X}^{j-m}, \boldsymbol{\omega}_{j-1}; \mathrm{p})$，其初始条件为 $\mathrm{X}_0 = \mathrm{x}_0(\mathrm{p}) \quad \cdots \mathrm{X}_{m-1} = \mathrm{x}_{m-1}(\mathrm{p})$。当在时刻 t_k 以初始条件 $\mathrm{X}_k = \mathrm{x}_k(\mathrm{p})$ 开始时，X^j 近似于 $\mathrm{X}(t_j)$，其中 $k \in [0, m-1]$。$\boldsymbol{\Psi}_{j-1}$ 利用时刻 $t_j, \ldots, t_{j-m}$ 的信息，近似随机动力学从时间步 t_{j-1} 到 t_j 的演化。由于动力学具有随机性，数值更新也是随机的，$\boldsymbol{\Psi}_j(\cdot, \boldsymbol{\omega}; \mathrm{p})$ 可视为概率空间 $(\Omega\ \mathcal{B}, \mu)$ 上的随机变量。这里，对于离散化系统，我们使用一组标准独立同分布高斯样本 $\boldsymbol{\omega} = \{\boldsymbol{\omega}_j\}_j \in \Omega$。p 表示数值方法的参数，其中可包括时间步长 Δt、初始条件以及所近似的底层随机系统的其他性质。上述表述涵盖隐式方法，其中 Ψ表示通过求解隐式方程得到的更新映射（Kloeden 和 Platen，1992）。对于生成模型 G，样本按 $\mathrm{X}^{[i]} = G(\mathrm{Z}^{[i]}; \theta_G)$ 的形式生成，其中我们取 $\theta_G = \mathrm{p}$ 并对 $\mathrm{Z}^{[i]} = \{\boldsymbol{\omega}_j^{[i]}\}_j$进行采样。

对于随机微分方程和常微分方程解的近似，众所周知，必须谨慎选择数值方法的更新方式，以确保收敛性。实现收敛性的一种常见策略，是将近似中的局部精度（相容性）与关于误差如何在更大范围内累积的条件（稳定性）结合起来（Iserles，2009；Burden 和 Faires，2010；Kloeden 和 Platen，1992）。我们通过将所学习的模型 $\boldsymbol{\Psi}$ 建立在稳定且准确的随机积分方法之上，来构建具备相关性质的生成模型类。

一阶随机动力系统中广泛使用的方法包括（m =1）步 Euler-Maruyuma 方法（Maruyama，1955）、Milstien 方法（Milstein，1975）和龙格-库塔方法（Kloeden 和 Platen，1992）。为了获得更高阶精度，或用于（$n > 1$）阶动力学，可以构造 m 步多步法，例如亚当斯-巴什福斯和 Adams-Moulten 方法的随机变体（Kloeden 和 Platen，1992；Buckwar 和 Winkler，2006；Ewald 和 T´emam，2005）。对于物理系统，进一步的考虑还会带来额外要求或替代方法，例如近似时间可逆性（Verlet，1967；Gronbech-Jensen 和 Farago，2013）、涨落-耗散平衡（Atzberger 等人，2007；Tabak 和 Atzberger，2015）、几何和拓扑条件（Lopez 和 Atzberger，2022；Gross 等人，2020），或采用指数时间步积分（Jasuja 和 Atzberger，2022；Atzberger 等人，2007）。我们开发用于在生成模型类上进行学习的方法，以利用所研究随机系统的稳定性、精度、物理属性及其他性质等方面的考量。

4.2 m 步生成模型的梯度方法

训练 m 步生成模型需要使用计算方程 1 中损失 $\nabla_{\theta_G}\ell^*$ 的梯度的方法。鉴于 $\nabla_{\theta_G}\mathbf{X}^{[i]}(;\theta_G)$ 的贡献，这带来了挑战，其中 $\theta_G = \mathbf{p}$，如第 2.2 节所述。训练需要计算离散化样本路径 $\{\mathbf{X}(t;\mathbf{p})\}_{t\in[0,\mathcal{T}]}$ 的梯度贡献。为此，我们将利用隐函数定理，为我们的 m 步生成模型开发一套伴随方法（Jittorntrum，1978；C´ea，1986；Strang，2007；Smith，2012）。设$\mathbf{f} = \mathbf{f}(\mathbf{X}_0,\ldots,\mathbf{X}_N;\mathbf{p})$ 为一个向量值函数，其分量为

$$[\mathbf{f}]^{[i]}_{(j,d)} = \left[\mathbf{X}^{[i]}_j - \mathbf{\Psi}_{j-1}(\mathbf{X}^{[i]}_{j-1},\ldots,\mathbf{X}^{[i]}_{j-m},\boldsymbol{\omega}_{j-1};\mathbf{p})\right]^{(d)} = 0, \tag{2}$$

其中，$\mathbf{X}^{[i]}_k = \mathbf{X}^{[i]}(t_k)$ 是 i^{th} 样本。上标索引 d 表示$\mathbf{X}_k \in \mathbb{R}^n$的维度分量。为简化记号，令 $\mathbf{x} = \{\mathbf{x}^{[i]}\}_{i=1}^M = \{(\mathbf{X}_0,\ldots,\mathbf{X}_{N\mathcal{T}})^{[i]}\}_{i=1}^M$，其中有 M 个样本，索引为 i。用 $g(\mathbf{x}\ p)$ 表示 $= \ell^*$ 损失函数，我们需要计算其关于 p 的梯度以进行训练。对于 $\boldsymbol{\omega}^{[i]}$ 的每个实现，轨迹样本由$\mathbf{f}^{[i]}(\mathbf{x};\ p)\ = 0$隐式确定。

g 的全导数由链式法则给出

$$[\nabla_p g]_j = \frac{dg}{dp_j} = \frac{\partial g}{\partial p_j} + \sum_k \frac{\partial g}{\partial x_k}\frac{\partial x_k}{\partial p_j} = [\mathbf{g}_p + \mathbf{g}_x\mathbf{x}_p]_j. \tag{3}$$

我们将导数收集为向量$\mathbf{g}_x \in \mathbb{R}^{1\times n_x}$,$\mathbf{g}_p \in \mathbb{R}^{1\times n_p}$, $\mathbf{x}_p \in \mathbb{R}^{n_x\times n_p}$，其中$\mathbf{x}\ \in\mathbb{R}^{n_x}$,$\mathbf{p}\ \in \mathbb{R}^{n_p}$。

损失 g 关于$\mathbf{x}$ 和$\mathbf{p}$ 的导数，通常比求$\mathbf{p}$ 关于$\mathbf{x}_p = \nabla_p\mathbf{x}(p)$ 的导数更容易。我们可以利用$\mathbf{f}^{[i]}(\mathbf{x},\mathbf{p}) = 0$，并假设$\mathbf{f}^{[i]}_{\mathbf{x}}$可逆，借助$\mathbf{f}^{[i]} \in \mathbb{R}^{n_f^{[i]}}$的隐函数定理求得$\mathbf{xp}$ 的导数表达式（Jittorntrum，1978年；Smith，2012年）。由此可得

$$\frac{df}{dp} = \mathbf{f}_x\mathbf{x}_p + \mathbf{f}_p = 0 \ \ \Rightarrow\ \ \mathbf{x}_p = -\mathbf{f}_x^{-1}\mathbf{f}_p = -J^{-1}\mathbf{f}_p. \tag{4}$$

我们始终假设 $n_f = n_x$，但在符号表示中将它们区分开来，以便更清楚地说明这些条件各自的作用。x 中的雅可比矩阵由 $J = \mathbf{f}_x = \nabla_x\mathbf{f}\ \in \mathbb{R}^{n_f\times n_x}$ 给出，对于映射$\mathbf{y}\ = \mathbf{f}(\mathbf{x};\mathbf{p})$，该映射由固定的 p 定义，用于将$\mathbf{x} \in \mathbb{R}^{n_x}$映射到$\mathbf{y} \in \mathbb{R}^{n_f}$。$\mathbf{x}_p = \nabla_p\mathbf{x}(p)\ \in \mathbb{R}^{n_x\times n_p}$以及$\mathbf{f}_p \in \mathbb{R}^{n_f\times n_p}$。将其代入上式可得

$$\nabla_p g = \mathbf{g}_p + \mathbf{g}_x(-\mathbf{f}_x^{-1}\mathbf{f}_p) = \mathbf{g}_p + \left(-\mathbf{f}_x^{-T}\mathbf{g}_x^T\right)^T\mathbf{f}_p = \mathbf{g}_p - \left(J^{-T}\mathbf{g}_x^T\right)^T\mathbf{f}_p = \mathbf{g}_p - \mathbf{r}^T\mathbf{f}_p. \tag{5}$$

令$\mathbf{r}\ = J^{-T}\mathbf{g}_x^T\ \in \mathbb{R}^{n_f\times 1}$。为计算梯度 $\nabla_p g$，我们按以下步骤进行：**(i)** 计算导数$\mathbf{g}_p, \mathbf{g}_x, \mathbf{f}_p$；**(ii)** 求解线性方程组 $J^T\mathbf{r}\ = \mathbf{g}_x^T$；**(iii)** 使用 $\nabla_p g = \mathbf{g}_p - \mathbf{r}^T\mathbf{f}_p$计算梯度。

这为利用解$\mathbf{r}\ \in \mathbb{R}^{n_f\times 1}$计算 g 的梯度提供了方法。对于轨迹样本，这通常比直接求解 $\mathbf{x}_p \in \mathbb{R}^{n_x\times n_p}$更高效。我们强调，构造稠密的 $n_x \times n_x$ 雅可比矩阵并对其求逆的代价极高，因此我们将采用以下方法求解线性方程：

这些方法利用稀疏性，且只需计算伴随 J^T的作用。下面我们将展示如何将其应用于我们的 m步方法，并在附录A中提供更多技术细节。

我们使用伴随方法来计算基于以下形式的期望及其梯度的损失函数梯度

$$g^*(\mathbf{X}(\cdot),p) = \mathbb{E}_\omega[\phi(\mathbf{X}(\omega;p);p)] = \int \phi(\mathbf{X}(\omega;p);p)d\mu_\omega, \quad \frac{dg^*}{dp} = \mathbb{E}[\nabla_p\phi] = \mathbb{E}[\nabla_p\tilde{g}]. \tag{6}$$

我们强调， g^* 的所有参数依赖性都显式地体现在随机变量 $\mathbf{X}(\omega,\ p)$ 和 $\tilde{g}=\phi$ 中，因此参考测度 μ_ω 不依赖于 p。在实际应用中，期望将通过样本平均值 $\bar{g}=\frac{1}{M}\sum_{i=1}^{M}\tilde{g}^{[i]}$ 近似，其中 $\tilde{g}^{[i]}$ 表示 i^{th} 样本。在这种情况下，我们使用伴随方法，通过以下方式计算梯度

$$\nabla_p\bar{g}(\mathbf{x},\mathbf{p}) = \frac{1}{M}\sum_{i=1}^{M}\nabla_p\tilde{g}^{[i]}, \ \ \nabla_p\tilde{g}^{[i]} = \tilde{\mathbf{g}}_p^{[i]} - \mathbf{r}^{[i],T}\mathbf{f}_p^{[i]}, \ \ \mathbf{r}^{[i]} = J^{[i],-T}\tilde{\mathbf{g}}_x^{[i],T}, \tag{7}$$

其中 $J^{[i],\ T}=\mathbf{f}_x^{[i]\,T}$。我们在附录A中为$\mathbf{r}^{[i]}$ 提供了高效的直接方法，用于求解 m步方法中的这些线性方程组。

4.3 损失中的二阶统计量的梯度方法 ℓ^*

~方程1中的损失函数 ℓ^* 涉及计算样本的二阶统计量。这用于比较候选生成模型的样本 $\mathbf{X}^{[i]}\sim\tilde{\mu}_G$ 与训练数据的样本 $\mathbf{Y}^{[i]}\sim\tilde{\mu}_{\mathcal{D}}$。在计算梯度时，直接对前向计算进行反向传播会带来挑战，因为这可能导致计算调用图过于庞大。

~我们利用损失函数 ℓ^* 的特殊结构，提出了替代方法，以更高效地估计训练所需的梯度。由方程 1 可知， $\tilde{\ell}^*$ 涉及核计算，其中包含如下形式的项： $k(\mathbf{W}_1^{[i]},\mathbf{W}_2^{[j]};p)$，~其中 $\mathbf{W}_1,\mathbf{W}_2\in\{\mathbf{X},\mathbf{Y}\}$。为了求得梯度 $\nabla_p\ell^*$，我们需要计算三个项的贡献，每一项的一般形式为

$$\frac{d}{dp}k(\mathbf{W}_1^{[i]},\mathbf{W}_2^{[j]};p) \ \ = \ \ \partial_1k\cdot\mathbf{W}_{1,p}^{[i]} + \partial_2k\cdot\mathbf{W}_{2,p}^{[i]} + \partial_3k. \tag{8}$$

∂_qk 表示对 $k(\cdot,;p)$的 q^{th} 输入求偏导。最难计算的项是轨迹样本 $\mathrm{W}_{a,p}=\partial\mathrm{W}_a/\partial p$ 在 $\mathrm{W}_a=\mathrm{X}$ 时的偏导。在这种情况下，我们有 $\mathrm{W}_{a,p}=\mathrm{X}_p$。对于 Y，我们有 $\mathrm{Y}_p=0$，因为训练数据不依赖于 p。我们利用梯度最终会与方程 8 中的向量值项 ∂_1k,∂_2k 进行收缩这一点。因此，我们可以使用 $h(\mathrm{X}^{[i]})=\mathrm{c}\cdot\mathrm{X}^{[i]}$来计算这些项。这等价于在将 c 视为常数时计算导数，从而得到$\nabla_ph(\mathrm{X}^{[i]})=\mathrm{c}\cdot\mathrm{X}_p^{[i]}$。换言之，在求导之后的计算中，该项随后会代入数值 $\mathrm{c}=\partial_1k$ 或 $\mathrm{c}=\partial_2k$，以进行最终求值。这很有用，因为借助第 4.2 节中的伴随方法，我们可以高效地计算 $\nabla_ph(\mathrm{X}^{[i]})$。为了计算偏导 $\partial_3k=\partial k/\partial p$，我们可以在固定其他输入的同时使用反向传播。

为了计算对梯度 $\tilde{\ell}^*$的贡献，我们考虑方程 1 中的第一项 $k(\mathbf{X}^{[i]}, \mathbf{X}^{[j]}; p)$，并使用以下方式处理对样本的求和

$$\frac{d}{dp}\sum_{i\neq j} k(\mathbf{X}^{[i]}, \mathbf{X}^{[j]}; p) \quad = \quad \sum_{i_0} \mathbf{c}_1(i_0)^T \mathbf{X}_p^{[i_0]} + \sum_{j_0} \mathbf{c}_2(j_0)^T \mathbf{X}_p^{[j_0]} + \mathbf{c}_3, \tag{9}$$

其中

$$\mathbf{c}_1(i_0) = \sum_{j\neq i_0} \partial_1 k(\mathbf{X}^{[i_0]}, \mathbf{X}^{[j]}; p), \quad \mathbf{c}_2(j_0) = \sum_{i\neq j_0} \partial_2 k(\mathbf{X}^{[i]}, \mathbf{X}^{[j_0]}; p), \quad \mathbf{c}_3 = \sum_{i\neq j} \partial_3 k(\mathbf{X}^{[i]}, \mathbf{X}^{[j]}; p). \tag{10}$$

为了计算各项$\mathbf{X}_p^{[i_0]} = \frac{d}{dp}\mathbf{X}^{[i_0]}$和$\mathbf{X}_p^{[j_0]} = \frac{d}{dp}\mathbf{X}^{[j_0]}$的贡献，我们采用第 4.2 节和附录A中讨论的伴随方法。对于方程 1 中对 ℓ^* 有贡献的第二项，我们使用

$$\frac{d}{dp}\sum_{i,j} k(\mathbf{X}^{[i]}, \mathbf{Y}^{[j]}; p) = \sum_{i_0} \tilde{\mathbf{c}}_1(i_0)^T \mathbf{X}_p^{[i_0]} + \tilde{\mathbf{c}}_3, \tag{11}$$

$$\tilde{\mathbf{c}}_1(i_0) = \sum_{j} \partial_1 k(\mathbf{X}^{[i_0]}, \mathbf{Y}^{[j]}; p), \quad \tilde{\mathbf{c}}_3 = \sum_{i,j} \partial_3 k(\mathbf{X}^{[i]}, \mathbf{X}^{[j]}; p). \tag{12}$$

对于方程 1 中的第三项，p 仅通过对核的直接贡献产生影响，因为$\mathbf{Y}^{[i]}$ 不依赖于 p。我们使用以下方法计算梯度：

$$\frac{d}{dp}\sum_{i\neq j} k(\mathbf{Y}^{[i]}, \mathbf{Y}^{[j]}; p) = \hat{\mathbf{c}}_3, \quad \hat{\mathbf{c}}_3 = \sum_{i\neq j} \partial_3 k(\mathbf{Y}^{[i]}, \mathbf{Y}^{[j]}; p). \tag{13}$$

对于每一项，$\partial_j k = \partial_j k(w_1, w_2; p)$都可以通过对 k 求值并采用局部反向传播方法来计算。我们用这种方法计算$\mathbf{c}_j, \tilde{\mathbf{c}}_j, \hat{\mathbf{c}}_j$。完整梯度对所有样本求和，但在实践中我们使用随机梯度下降。这样便可使用更小的样本小批次，进一步加快计算。对于 SDYN-GANs，我们采用这些方法实现自定义反向传播方法，以对我们的 m步生成模型进行基于梯度的训练。更多细节见附录A。

5. 结果

我们展示了 SDYN-GANs 在学习二阶随机动力学生成模型方面的结果。这项工作受到微尺度机械系统及其惯性随机动力学的物理学启发。我们开发了用于学习力定律、阻尼系数和噪声相关参数的方法。第 5.2 节展示了针对与惯性奥恩斯坦-乌伦贝克（I-OU）过程相对应的微机械系统学习物理模型的结果。随后，我们在第 5.3 节研究了如何为具有惯性朗之万动力学的粒子系统学习非线性力定律。

在这些研究中，我们考察了时间尺度 τ 对轨迹采样的作用，以及基于从以下分布中采样的不同训练方案：(i) 完整轨迹，(ii) 条件分布，以及 (iii) 边缘分布。结果展示了训练 SDYN-GANs、以获得随机系统表示的几种策略。

5.1 惯性二阶随机动力学的离散化

我们始终考虑形式如下的二阶惯性随机动力学

$$m d\mathbf{V}(t) = -\gamma \mathbf{V}(t)dt + \mathbf{F}(\mathbf{X}(t))dt + \sigma d\mathbf{W}(t), \quad d\mathbf{X}(t) = \mathbf{V}(t)dt. \tag{14}$$

对于 ForSDYN-GANs，我们的生成模型使用以下 ($m = 2$) 步随机数值积分器（Gronbech-Jensen 和 Farago，2013 ）。该积分器可表示为如下形式的 Velocity-Verlet 类方案（ Verlet，1967 ）：

$$\begin{aligned}
\mathbf{X}_{n+1} &= \mathbf{X}_n + b\Delta t \mathbf{V}_n + \frac{b\Delta t^2}{2m}\mathbf{F}_n + \frac{b\Delta t}{2m}\boldsymbol{\eta}_{n+1}, \qquad a = \left(1 - \frac{\gamma \Delta t}{2m}\right)\left(1 + \frac{\gamma \Delta t}{2m}\right)^{-1}, \\
\mathbf{V}_{n+1} &= a\mathbf{V}_n + \frac{\Delta t}{2m}\left(a\mathbf{F}_n + \mathbf{F}_{n+1}\right) + \frac{b}{m}\boldsymbol{\eta}_{n+1}, \qquad b = \left(1 + \frac{\gamma \Delta t}{2m}\right)^{-1}.
\end{aligned} \tag{15}$$

我们将这种随机动力学的离散化称为 Farago 方法，已证明其具有良好的稳定性和统计性质（Gronbech-Jensen 和 Farago，2013）。热驱动通过均值为零、方差为 $\langle \boldsymbol{\eta}_n \boldsymbol{\eta}_\ell^T \rangle = \sigma^2 I \delta_{n,\ell} \Delta t$ （时间步长为 Δt）的高斯噪声 $\boldsymbol{\eta}_n$ 进行近似。该方法可表示为 2 阶多步法 $\mathbf{X}_{n+1} = \boldsymbol{\Psi}_n(\mathbf{X}_n, \mathbf{X}_{n-1}, \boldsymbol{\omega}_n; \mathbf{p})$，其中 $\boldsymbol{\Psi}_n = 2b\mathbf{X}_n - a\mathbf{X}_{n-1} + \frac{b\Delta t^2}{m}\mathbf{F}_n + \frac{b\Delta t}{2m}\left(\boldsymbol{\eta}_{n+1} + \boldsymbol{\eta}_n\right)$. 我们的生成模型始终使用这种随机数值离散化方法，来近似方程 14 中的惯性二阶随机动力学。

5.2 通过生成建模学习物理模型：惯性 Ornstein-Uhlenbeck 过程

考虑 Ornstein-Uhlenbeck（I-OU）过程（Uhlenbeck 和 Ornstein，1930），其随机动力学为

$$m d\mathbf{V}(t) = -\gamma \mathbf{V}(t)dt - K_0 \mathbf{X} dt + \mathbf{F}_0 dt + \sigma d\mathbf{W}(t), \quad d\mathbf{X}(t) = \mathbf{V}(t)dt. \tag{16}$$

The $\mathbf{X}, \mathbf{V} \in \mathbb{R}^n$。研究分子系统时会遇到相关动力学（Uhlenbeck 和 Ornstein，1930；Reichl，1997；Frenkel 和 Smit，2001；Yushchenko 和 Badalyan，2012）、微观器件（Gomez-Solano 等人，2010年；Goerlich 等人，2021年；Dudko 等人，2002年）、金融学和经济学问题（Steele，2001；Vasicek，1977；Sch¨obel 和 Zhu，1999；A¨ıt-Sahalia 和 Kimmel，2007年）以及许多其他应用（Large，2021；Wilkinson 和 Kennard，2012年；Ricciardi 和 Sacerdote，1979；Giorgini 等人，2020年）。

从力学的角度来看，这些动力学可以视为对黏性溶剂流体中的微观珠簧系统进行建模。在这种情况下，$\mathbf{X}$ 是位置，$\mathbf{V}$ 是速度，其中 $\mathbf{X}, \mathbf{V} \in \mathbb{R}^3$。 m 是珠子的质量， γ 是黏性阻尼的程度，$\sigma = \sqrt{2k_B T\gamma}$ 给出了温度T下涨落的强度，其中 k_B 表示玻尔兹曼因子（ Reichl，1997）。保守力包括刚度为 K_0 的谐振弹簧，以及恒定力 $\mathbf{F}_0 = -mg\mathbf{e}_z$，例如系数为 g且 $\mathbf{e}_z = (0,0,1)$ 的重力。这给出的势能为 $U(\mathbf{X}) = \frac{1}{2}K_0\mathbf{X}^2 - \mathbf{F}_0 \cdot \mathbf{X}$ ，由此得到保守力 $\mathbf{F}(\mathbf{X}) = -\nabla_{\mathbf{X}} U$。我们在生成模型中采用方程 16 的动力学，并通过 ($m = 2$) 步随机数值

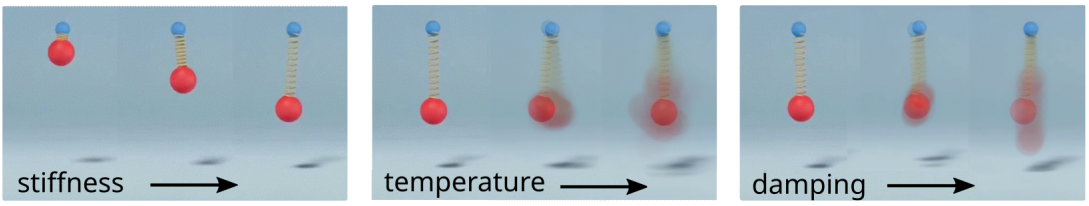

图 3：**珠簧力学：惯性奥恩斯坦-乌伦贝克过程**。我们考察三维惯性奥恩斯坦-乌伦贝克过程在不同物理条件下的随机动力学，包括 (i) 刚度降低（左）、(ii) 温度升高（中）或 (iii) 阻尼降低（右）的情形。

离散化进行近似，具体内容见第 5.1 节。我们用 θ 表示待调整的参数集合，即 $\theta = (K_0, \gamma, k_B T)$。

数据驱动任务是根据系统构型$\mathbf{X}_{[0,\mathcal{T}]} = \{\mathbf{X}(t)\}_{t\in[0,\mathcal{T}]}$的轨迹观测数据，学习生成式随机模型 $G(\cdot;\theta_G)$，使得 $G(\mathbf{Z};\theta_G) \sim \mathbf{X}^{[0,\mathcal{T}]}$。$\mathbf{Z}$ 是噪声源，此处将其视为分量独立同分布的高维向量值高斯分布。$\mathbf{V}(t)$无法直接观测。我们在图 4 中展示了训练数据的样本。训练使用离散化轨迹和形如 $(\mathbf{X}(t_1)\ldots,\mathbf{X}(t_m))$ 的轨迹片段。若不利用数据的其他结构信息，鉴于轨迹$\mathbf{X}(t)$的分布具有高维性，这项任务可能颇具挑战。

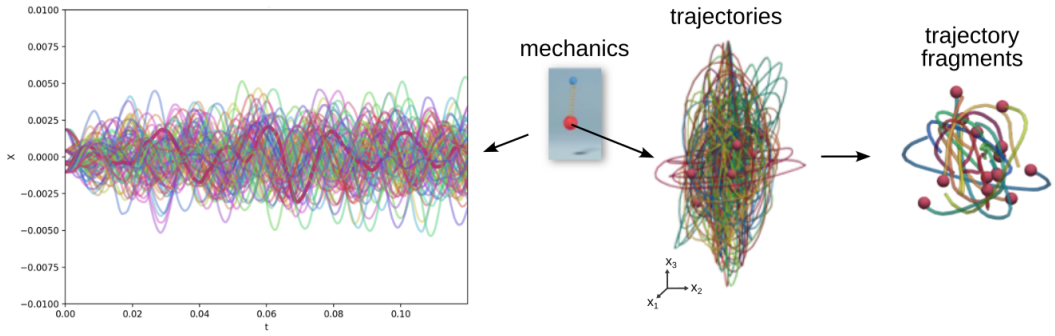

图 4：**轨迹训练数据：惯性奥恩斯坦-乌伦贝克过程**。我们展示了用于训练模型的轨迹数据样本。轨迹数据的 z 分量如（左）所示。珠簧系统及其在三维空间中的轨迹，如（中）所示。（右）展示了第 2.2 节中讨论的不同训练方案所用的轨迹片段。

SDYN-GANs 使用轨迹中的概率依赖关系以及第 2.2 节所讨论的对抗学习方法。这为学习适用于各种任务的通用模型提供了灵活的方法

即使生成模型可能不具备易于表达的概率分布。通过在我们的 SDYN-GANs 方法中使用最大均值差异，可以根据所采用的核引入进一步的归纳偏置。这可用于突出所比较经验分布的关键特征。对于 OU过程研究，我们使用具有有理二次核（RQK）的再生核希尔伯特空间（Rasmussen和Williams，2006年；Duvenaud，2014年），其定义为 $k(x,\ y) = \left(1 + \frac{\|x-y\|^2}{2\alpha\ell^2}\right)^{-\alpha}$. 其中，$\alpha$ 决定衰减速率，而 ℓ 提供用于比较的特征尺度。RQK 是一种特征核，能够对概率分布进行一般性比较（Sriperumbudur 等人，2010年）。对于基于梯度的学习而言，它的优势在于尾部不会呈指数衰减，从而在高维空间中表现更佳（Rasmussen和Williams，2006年；Duvenaud，2014年）。

图5展示了同时学习参数 $\theta = (K_0, \gamma, k_BT)$ 时的典型训练行为。我们发现，训练初期会通过增大 k_BT 来拓宽生成模型的分布，使其与训练数据样本重叠，从而降低损失。然而，当刚度 K_0 和阻尼 γ 参数接近合理值时（约在训练轮次 ~ 750），温度 k_BT 会迅速下降，使模型能够更准确地捕捉观测数据中其余的特征。在这次训练试验中，所有参数在训练轮次 $\sim 1,200$ 左右都接近正确值，见图5。

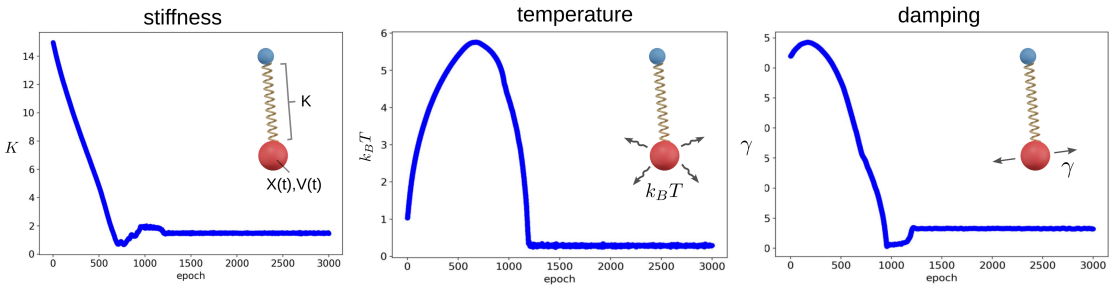

图5：**惯性奥恩斯坦-乌伦贝克过程：学习随机动力学**。我们使用 SDYN-GANs 学习随机动力学观测轨迹的生成模型，同时学习动力学中的漂移和扩散贡献。此次试验的目标值为 $K_0 = 1.5$、$\gamma = 3.2$和 $k_BT = 0.1$。我们发现，在训练初期，随机参数会变得很大，以便通过增强候选生成模型样本与观测样本之间的重叠来最小化损失。随着训练过程中刚度和阻尼参数变得更加准确，随机参数会逐渐下降并趋近目标值。

为了进一步探究 SDYN-GANs 的性能，我们采用几种不同的方法进行训练，包括改变采样系统构型 $\mathbf{X}(t)$ 时所采用的时间尺度 τ overwhich，以及隐式探测的分布类型。例如，我们在时间持续时长 $\mathcal{T}$ 内使用完整轨迹，同时考察较短时间持续时长内的条件分布或边缘分布

统计量。为此，我们采用第 2.2 节讨论的不同训练方案和时间尺度 τ。在全文中，我们使用默认的目标模型参数 $m = 0.1, K_0 = 1.5, \gamma = 3.2, k_B T = 0.1$ 以及核参数 $\alpha = 2, \ell = 0.01$。对于不同的 SDYN-GANs 方法和训练方案，我们在表 1–3 中展示所学生成模型的准确率。

我们发现，刚度和阻尼机制各自都能以约 5% 或更低的误差被学习。系统的热涨落则更难从轨迹观测中学习。这似乎是由于其他因素也会影响感知到的涨落，包括采样误差，以及为增强候选生成模型与轨迹数据之间的排列而倾向于高估热参数。我们发现，系统热涨落的准确率仅约为 50%。我们还发现，用于探测动力学时间子采样的时间延迟 τ 会显著影响学习方法的性能。

准确率 K_0 与 τ 延迟			
τ 延迟	**1.70e-02**	**1.19e-02**	**5.83e-03**
MMD-完整轨迹	$1.03e$-02 $\pm$ $1.1e$-02	$2.62e$-02 $\pm$ $2.4e$-02	$1.13e$-02 $\pm$ $7.8e$-03
MMD-条件分布	$9.19e$-03 $\pm$ $1.9e$-03	$2.22e$-02 $\pm$ $1.7e$-02	$2.66e$-02 $\pm$ $4.5e$-02
MMD-边缘分布	$1.54e$-02 $\pm$ $6.7e$-03	$8.92e$-03 $\pm$ $1.1e$-02	$4.37e$-02 $\pm$ $4.5e$-02
τ-延迟	**4.08e-03**	**2.86e-03**	**2.00e-03**
MMD-完整轨迹	$1.62e$-02 $\pm$ $1.5e$-02	$8.02e$-03 $\pm$ $5.2e$-03	$1.37e$-02 $\pm$ $7.0e$-03
MMD-条件分布	$1.03e$-01 $\pm$ $1.9e$-01	$8.52e$-01 $\pm$ $1.7e$+00	$5.68e$-01 $\pm$ $1.1e$+00
MMD-边缘分布	$4.44e$-02 $\pm$ $8.0e$-02	$1.63e$+00 $\pm$ $3.2e$+00	$1.72e$+00 $\pm$ $3.4e$+00

表 1：**准确率 K_0 与 τ-延迟：惯性奥恩斯坦-乌伦贝克过程**。SDYN-GANs 的训练采用了第 2.2 节中讨论的协议，分别基于 (i) 完整轨迹、(ii) 边缘分布和 (iii) 条件分布进行。~对于 $\phi = K_0$，刚度估计的准确率 ϕ 由相对误差 $\epsilon_{\text{rel}} = |\phi - \phi|/|\phi$|给出。训练轮次为 3000。报告了 4 次运行的标准差。所有轨迹的总持续时间为 $t_n = 1.8 \times 10^{-2}$，共计 $n = 18$ 步，且$\Delta t = 10^{-3}$。

对于时间采样，我们发现当 τ 时间尺度相对较大时，模型学习得更准确。这似乎与样本能够更好地探测轨迹数据中的时间尺度有关，在这些时间尺度上，刚度和阻尼的影响更显著。较长的轨迹持续时间似乎能为学习这些属性提供更好的信号。随着 τ 变小，我们发现针对该任务所学习模型的准确率有所下降。

我们进一步发现，所选的训练方案会对学习得到的模型准确率产生显著影响。对于当前任务，我们发现，完整轨迹方案或边缘分布方案通常表现最佳。边缘分布方案的一个显著特点是，轨迹样本包含了条件分布中出现的相同轨迹片段，尤其是那些从采样初始条件附近开始的片段。这似乎是边缘分布方案优于条件分布方案的原因之一，后者包含的长时间动力学信息较少。

准确率 k_BT 与 τ延迟的对比			
τ延迟	**1.70e-02**	**1.19e-02**	**5.83e-03**
MMD-完整轨迹	$5.36e$-01 $\pm$ $9.5e$-01	$1.71e$+00 $\pm$ $3.4e$+00	$3.62e$+00 $\pm$ $7.1e$+00
MMD-条件分布	$6.43e$-01 $\pm$ $1.1e$+00	$1.46e$+00 $\pm$ $2.6e$+00	$1.92e$+01 $\pm$ $3.8e$+01
MMD-边缘分布	$6.54e$-01 $\pm$ $1.2e$+00	$1.51e$+00 $\pm$ $2.9e$+00	$1.39e$+01 $\pm$ $2.7e$+01
τ-延迟	**4.08e-03**	**2.86e-03**	**2.00e-03**
MMD-完整轨迹	$3.47e$+00 $\pm$ $6.7e$+00	$6.43e$+00 $\pm$ $1.3e$+01	$6.90e$+00 $\pm$ $1.4e$+01
MMD-条件分布	$2.42e$+01 $\pm$ $4.7e$+01	$9.54e$+01 $\pm$ $1.7e$+02	$9.38e$+01 $\pm$ $1.7e$+02
MMD-边缘分布	$1.85e$+01 $\pm$ $3.7e$+01	$3.95e$+01 $\pm$ $7.9e$+01	$3.82e$+01 $\pm$ $7.6e$+01

表 2：**准确率 k_BT 相对于 τ-延迟：惯性奥恩斯坦-乌伦贝克过程**。SDYN-GANs 使用第 2.2 节所述的 (i) 完整轨迹、(ii) 边缘分布和 (iii) 条件分布方案进行训练。温度估计的准确率$\tilde{}$ $\tilde{}$ ϕ 对于 $\phi = k_BT$ 由相对误差 $\epsilon_{\text{rel}} = |\phi - \tilde{\phi}|/|\phi|$给出。训练轮次为 3000。报告了 4 次运行的标准差。所有轨迹的总时长为 $t_n = 1.8 \times 10^{-2}$，共 $n = 18$ 步，且$\Delta t = 10^{-3}$。

总体而言，我们发现，使用沿轨迹更多步数的完整轨迹训练方案表现更好，优于边缘分布和条件分布训练方案。边缘分布方案的表现也往往优于条件分布方案。例外的是，在少数 τ 较小的情况下，我们发现条件分布方案在估计刚度方面表现优于边缘分布方案。这或许是因为，在较短时间跨度内，与边缘分布方案中众多轨迹片段共同提供的信息相比，条件分布方案在相同初始条件附近进行时间上更局部的采样，能够提供关于刚度响应更一致且更少受干扰的信息。

准确率 γ 与 τ-延迟			
τ-延迟	**1.70e-02**	**1.19e-02**	**5.83e-03**
MMD-完整轨迹	$1.53e$-02 $\pm$ $1.5e$-02	$3.31e$-02 $\pm$ $2.5e$-02	$2.91e$-02 $\pm$ $4.3e$-02
MMD-条件分布	$2.85e$-02 $\pm$ $2.3e$-02	$6.57e$-02 $\pm$ $6.0e$-02	$2.29e$-02 $\pm$ $3.3e$-02
MMD-边缘分布	$1.67e$-02 $\pm$ $8.4e$-03	$3.34e$-02 $\pm$ $1.9e$-02	$6.20e$-02 $\pm$ $8.9e$-02
τ-延迟	**4.08e-03**	**2.86e-03**	**2.00e-03**
MMD-完整轨迹	$5.20e$-02 $\pm$ $4.7e$-02	$4.70e$-02 $\pm$ $4.6e$-02	$2.20e$-02 $\pm$ $1.8e$-02
MMD-条件分布	$2.00e$-01 $\pm$ $3.7e$-01	$1.91e$+00 $\pm$ $3.8e$+00	$1.88e$+00 $\pm$ $3.7e$+00
MMD-边缘分布	$1.90e$-01 $\pm$ $3.0e$-01	$5.54e$+00 $\pm$ $1.1e$+01	$5.47e$+00 $\pm$ $1.1e$+01

表3：**准确率 γ 与 τ延迟：惯性奥恩斯坦-乌伦贝克过程**。SDYN-GANs 使用第 2.2 节讨论的方案进行训练，这些方案分别基于 (i) 完整轨迹、(ii) 边缘分布和 (iii) 条件分布。$\tilde{}$ $\tilde{}$阻尼估计值 ϕ 的准确率 $\phi = \gamma$ 由相对误差 $\epsilon_{\text{rel}} = |\phi - \tilde{\phi}|/|\phi|$给出。训练轮次为 3000。报告了 4 次运行的标准差。所有轨迹的总时长为 $t_n = 1.8 \times 10^{-2}$，共 $n = 18$ 步，且$\Delta t = 10^{-3}$。

在较长的 τ 情形下，与条件分布相比，边缘分布往往能够更准确地同时估计刚度和阻尼。这可能是因为，与条件分布情形不同，边缘分布不要求所比较的轨迹片段的初始部分具有相同的初始条件。对于通常的短时长片段，边缘分布还包含有关长时间动力学响应的额外信息。总之，结果表明，在可能的情况下应采用完整轨迹方案，并进行相对较长时间的 τ 采样。当只能探测短时长的轨迹片段时，结果表明边缘分布较条件分布具有一定优势。哪种方案效果最佳，可能取决于任务以及所建模动力学行为对应的时间尺度。这些结果展示了使用 SDYN-GANs 为随机系统学习生成模型的若干策略。

5.3 使用 SDYN-GAN生成建模学习未知力定律

我们展示了如何利用 SDYN-GANs，根据随机系统轨迹的观测数据学习力定律**F(X)**。我们考虑如下形式的粒子系统的惯性朗之万动力学

$$md\mathbf{V}(t) = -\gamma\mathbf{V}(t)dt + \mathbf{F}(\mathbf{X};\theta_F)dt + \sqrt{2k_BT\gamma}d\mathbf{W}(t), \quad d\mathbf{X}(t) = \mathbf{V}(t)dt, \tag{17}$$

其中 X、V $\in \mathbb{R}^3$。力定律 F(X) 将使用深度神经网络（DNNs）建模（LeCun 等人，2015年；Goodfellow 等人，2016）。在这种情况下，θ_F 是待学习的 DNN 权重和偏置。这里我们考虑径向势的情形 F(X) =F(‖X‖) $=F(r)\mathbf{e}_r$，其中 $\mathbf{e}_r$ =X/‖X‖。为简单起见，我们聚焦于 m、γ、T固定的情形，但也可以将这些量与力定律一并学习。数据驱动任务是从配置轨迹 $\mathrm{X}_{[0,\mathcal{T}]} = \{\mathrm{X}(t)\}_{t\in[0,\mathcal{T}]}$（离散化为 (X($t_1$) . . . , X($t_n$))）中学习力定律的 DNN 表示 F(X; θ_F)，见图6。V(*t*) 未被直接观测。我们使用方程17中的朗之万动力学作为生成模型，并采用第5.1节讨论的 ($m = 2$) 步随机数值离散化对其进行近似。图7展示了训练数据的样本。

在我们的生成模型中，我们使用输入层-隐藏层-输出层规模为 n_{in}-n_1-n_2-n_3-n_{out} 的多层感知机深度神经网络架构，并采用 LeakyReLU（Xu 等，2015年）单元，其斜率=-0.01。在深度神经网络中，我们使用 $n_{\mathrm{out}} = n_{\mathrm{in}} = 1$，其默认值为 $n_1 = n_2 = n_3 = 100$，并为除最后一层外的所有层设置偏置。深度神经网络用于建模径向力中的 F $\mathbf{F} = F(r;\theta_F)\mathbf{e}_r$。

我们使用 SDYN-GANs，通过考察不同时间尺度 τ 下的轨迹，并采用基于以下采样方式的训练方案来学习生成模型：(i) 完整轨迹、(ii) 条件分布，以及 (iii) 边缘分布，详见第 2.2 节。为训练模型，我们模拟候选生成模型的轨迹，并将其与观测训练数据的样本进行比较。为评估所学模型的最终准确率，我们使用时变径向概率分布在 L^1-范数下的相对误差，将生成模型的样本与目标粒子系统进行比较。需要强调的是，训练过程中不使用 L^1-误差；它仅用于评估最终模型所生成随机轨迹的准确率。我们使用具有不同时间尺度 τ 和训练方案的 SDYN-GANs 进行训练，结果见表 4。

我们发现，生成模型以约 20% 的准确率学习力定律。我们发现，使用完整轨迹和边缘分布的训练方案表现

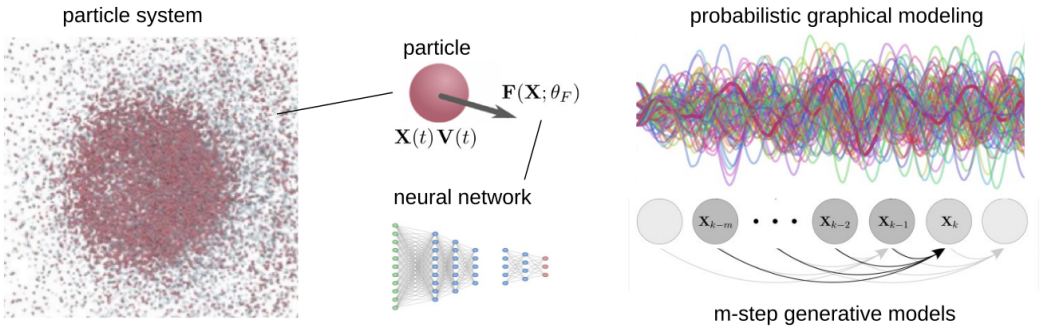

图6：**学习未知力定律**。SDYN-GANs 用于根据粒子轨迹的观测数据学习未知力定律（左）。学习到的力定律可以用神经网络或其他模型类别表示（中）。生成模型利用随机轨迹概率中的依赖关系，并采用 m 步随机数值离散化（右）。模型通过几种不同的训练方案进行学习，这些方案采用了以下统计量：(i) 完整轨迹、(ii) 条件分布和 (iii) 边缘分布，如第 2.2 节所述。

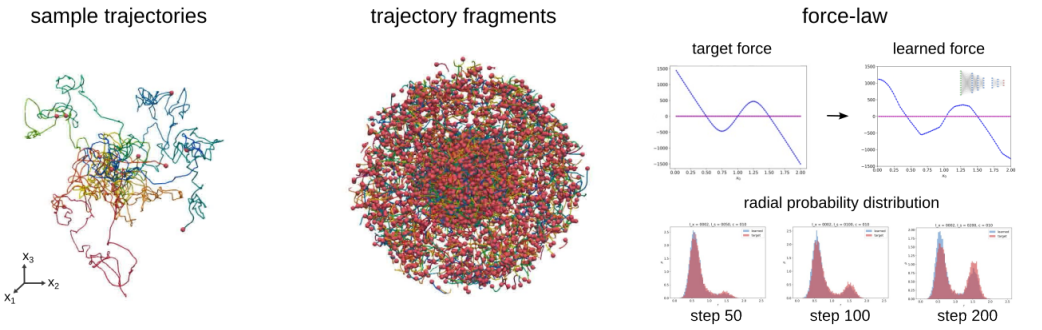

图7：**学习未知力定律：轨迹数据**。我们展示了若干粒子轨迹样本（左）。根据第 2.2 节所述的不同训练方案，利用轨迹片段样本学习力定律（中）。SDYN-GANs 用于根据粒子系统轨迹的观测数据，学习力定律的神经网络表示。我们通过比较学习模型样本的边缘时间依赖径向概率分布与目标粒子系统的相应分布，计算二者之间的相对 L^1-范数，以评估学习模型的最终准确率（右）。

最佳，见表 4。我们指出，从理论上讲，在无限时间范围极限下，Large τ 将为$\mathbf{X}$ 提供来自平衡吉布斯-玻尔兹曼分布的样本，$\rho(\mathbf{X}) = 1/Z \exp(-U(\mathbf{X})/k_B T)$。平衡分布仅取决于力定律的势能 U。有趣的是，我们发现训练过程中并不存在对

准确率 vs τ-延迟			
方法 / τ-延迟	**1.90e-02**	**1.16e-02**	**7.12e-03**
MMD-完整轨迹 ($I_s = 50$)	$1.20e\text{-}01 \pm 4.9e\text{-}02$	$1.02e\text{-}01 \pm 4.1e\text{-}02$	$8.67e\text{-}02 \pm 3.2e\text{-}02$
MMD-条件分布	$1.33e\text{-}01 \pm 3.5e\text{-}02$	$1.42e\text{-}01 \pm 6.1e\text{-}02$	$1.42e\text{-}01 \pm 4.2e\text{-}02$
MMD-边缘分布	$8.20e\text{-}02 \pm 4.5e\text{-}02$	$1.05e\text{-}01 \pm 3.4e\text{-}02$	$9.00e\text{-}02 \pm 2.3e\text{-}02$
MMD-完整轨迹 ($I_s = 100$)	$1.56e\text{-}01 \pm 8.8e\text{-}02$	$1.27e\text{-}01 \pm 4.3e\text{-}02$	$1.06e\text{-}01 \pm 2.8e\text{-}02$
MMD-条件分布	$1.39e\text{-}01 \pm 4.0e\text{-}02$	$1.68e\text{-}01 \pm 5.7e\text{-}02$	$1.33e\text{-}01 \pm 4.7e\text{-}02$
MMD-边缘分布	$9.75e\text{-}02 \pm 4.7e\text{-}02$	$9.96e\text{-}02 \pm 3.4e\text{-}02$	$9.97e\text{-}02 \pm 2.3e\text{-}02$
MMD-完整轨迹 ($I_s = 200$)	$2.22e\text{-}01 \pm 1.6e\text{-}01$	$1.69e\text{-}01 \pm 7.7e\text{-}02$	$1.51e\text{-}01 \pm 3.8e\text{-}02$
MMD-条件分布	$1.67e\text{-}01 \pm 6.2e\text{-}02$	$2.28e\text{-}01 \pm 8.3e\text{-}02$	$2.28e\text{-}01 \pm 3.7e\text{-}02$
MMD-边缘分布	$1.16e\text{-}01 \pm 3.1e\text{-}02$	$1.24e\text{-}01 \pm 2.3e\text{-}02$	$1.15e\text{-}01 \pm 2.4e\text{-}02$

表 4： **L^1-学习到的力定律的准确率**。 我们考察在时刻 $I_{step} = 50$、100、200、训练轮次 $= 5000$ 时，学习到的生成模型与目标随机过程的时变边缘概率密度的相对 L^1-误差，并对两者进行比较。报告的是 5 次运行的均值和标准差。SDYN-GANs 的训练方案基于来自以下分布的轨迹样本：（i）完整轨迹，（ii）条件分布，以及（iii）边缘分布；详情请参阅第 2.2 节。经验径向概率分布的 L^1-相对误差根据直方图计算，使用范围 $[0, L]$、$L = 2.5$ 和箱宽 $\Delta x = 0.05$。所有轨迹的总时长为 $t_{n_t} = 2.0 \times 10^{-2}$，共 $n_t = 20$ 步，$\Delta t = 10^{-3}$。

关于 τ 的结果。这些结果表明，与平衡态估计器不同，SDYN-GANs 学习方法具有灵活性，即使在非平衡情形下，也能根据相对于平衡时间尺度较短时间内的观测数据来估计力定律。

我们发现，对于这项任务，从完整轨迹和边缘分布中采样，往往比从条件分布中采样效果更好。对于较短的时间段，这似乎是因为边缘分布不仅包含条件分布中所纳入的轨迹片段的贡献，还包含来自轨迹较后部分的片段。再者，这为较长时间尺度下的动力学响应提供了额外信息，从而在此处促进了力定律的学习。训练期间，对边缘分布的匹配也可能比对条件分布的匹配要求更严格。要达成一致，生成的轨迹需要进行更全面的协调，使轨迹后续部分的统计特征与观测目标系统相匹配。

我们指出，对于条件分布和边缘分布，使用时长较短的轨迹片段有助于降低样本的维度。这似乎能改善损失函数在与观测数据进行比较时的表现。结果表明，当只能在相对较短的时间段内以片段形式探测轨迹数据时，边缘分布或许仍能为 SDYN-GANs 的训练带来一些优势。

总而言之，结果展示了使用 SDYN-GANs 从系统动力学的观测数据中学习非线性力定律及其他物理性质的几种策略。SDYN-GANs 的基于样本的方法允许使用通用的生成建模类别，而无需指定似然函数。这促进了通用生成模型的使用

，用于学习随机系统的表征。SDYN-GANs 方法也可用于非神经网络建模类别，并可与其他训练方案相结合，例如数据增强、进一步的正则化，以及由任务需求和应用领域知识所启发的其他先验信息。

6. 结论

我们提出了 SDYN-GANs，用于通过对抗学习构建表示随机动力系统的生成模型。我们方法的一项关键特点是，只需从候选生成模型中获取经验轨迹样本，即可将其与观测数据进行比较。因此，无需指定似然函数，就能使用广泛的生成模型。我们还开发了一类基于稳定的 m 步随机数值积分器的生成模型，以便学习适用于长期预测和仿真的模型。我们提出了若干训练方法，用于探测不同时间尺度上的动力学、利用概率依赖关系，以及发挥条件分布和边缘分布的不同统计性质。正如我们的结果所示，这些方法为降低训练中所用统计量的维数提供了策略，也让我们能够灵活地利用动力学的不同特征来学习生成模型。我们开发的 SDYN-GANs 方法有助于学习稳健的随机动力学模型，可用于机器学习、统计推断、动力系统辨识和科学计算中的各种任务。

致谢

作者 PJA 和 CD 获得了美国能源部高级科学计算研究 PHILMS 项目资助（DE-SC0019246）。作者 PJA 还获得了美国国家科学基金会 DMS-1616353 资助。作者 PJA 还感谢 UCSB 科学计算中心、UCSB 材料研究实验室以及美国国家科学基金会 CNS-1725797 和 NSF 材料研究科学与工程中心（DMR1121053）的支持。PS 的工作由美国能源部高级科学计算研究资助的“面向多尺度多物理问题的数学与物理信息学习机器（PhILMs）合作项目”支持。太平洋西北国家实验室由巴特尔纪念研究所根据合同 DE-AC05-76RL01830 为美国能源部运营。作者 PJA 还要感谢 Nvidia 提供的硬件资助。

附录

A. m步生成模型的梯度方程

对于我们的 m 步生成模型，在生成轨迹样本时，使用直接反向传播来记录前向计算过程，可能会产生规模大到难以承受的计算调用图。随着轨迹持续时间增加，这一问题会进一步加剧。我们推导了替代性的伴随方法，以便通过更高效的方法计算损失函数所需的梯度，用于训练。

对于我们在第 4.2 节中讨论的一般 m 步生成模型，我们采用如下形式$\mathbf{f(x,p)} = 0$。对于 $m \leq j \leq N$，有

$$[\mathbf{f}]_{(j,d)} = [\mathbf{X}_j - \mathbf{\Psi}_{j-1}(\mathbf{X}_{j-1}, \ldots, \mathbf{X}_{j-m}, \boldsymbol{\omega}_{j-1}; \mathbf{p})]^{(d)} = 0. \tag{18}$$

索引 (j,d) 表示与$\mathbf{X}_j^{(d)}$ 相关的 j^{th} 时间步长和 d^{th} 维度分量。我们采用以下记号：$\mathbf{X}_j^{(d)} = [\mathbf{X}_j]^{(d)} = \mathbf{X}_{(j,d)}$。更新映射记为$\mathbf{\Psi}_{j-1}$。对于 $j = 0, \ldots, m-1$，有

$$[\mathbf{f}]_{(0,d)} = [\mathbf{X}_0 - \mathbf{x}_0]^{(d)} = 0, \;\; [\mathbf{f}]_{(m-1,d)} = [\mathbf{X}_{m-1} - \mathbf{x}_{m-1}]^{(d)} = 0. \tag{19}$$

然后，我们可以给出 $\nabla_{X_k}\mathbf{\Psi}_k, \ldots, \nabla_{X_k}\mathbf{\Psi}_{k+m}$ 以及其他项的梯度 s.

我们推导了高效求解 $J^T\mathbf{r} = -\mathbf{f}_p$（其中 $\mathbf{r}$ 满足 $J = \mathbf{f}_x$）的方法。有关此公式的动机，详见第 4.2 节。利用方程 18–19 中 $J = \mathbf{f}_x$ 的结构，我们可以得到以下关于$\mathbf{r}$各分量的递推方程。对于 $0 \le k \le N-m$，我们有

$$r_{(j,d)}f_{(j,d),(k,d')} = r_{(k,d)}\delta_{d,d'} - \left[\nabla_{X_{(k,d')}}\Psi_{(k,d)}\right] r_{(k+1,d)} \cdots - \left[\nabla_{X_{(k,d')}}\Psi_{(k+m,d)}\right] r_{(k+m,d)} = \frac{\partial\phi}{\partial X_{(k,d')}}. \tag{20}$$

对于 $k = N-\ell,\ 1 \le \ell \le m-1$，我们有

$$r_{(j,d)}f_{(j,d),(k,d')} = r_{(k,d)}\delta_{d,d'} - \left[\nabla_{X_{(k,d')}}\Psi_{(k,d)}\right] r_{(k+1,d)} \cdots - \left[\nabla_{X_{(k,d')}}\Psi_{(k+\ell,d)}\right] r_{(k+\ell,d)} = \frac{\partial\phi}{\partial X_{(k,d')}}. \tag{21}$$

对于 $k = N$，我们有

$$r_{(j,d)}f_{(j,d),(k,d')} = r_{(N,d)}\delta_{d,d'} = \frac{\partial\phi}{\partial X_{(N,d')}}. \tag{22}$$

这给出了一个 m^{th} 阶递推关系，我们可以用它将值从 $N, N-1, \ldots, 0$ 逐步传递，从而得到 $\mathbf{r}$ 的各分量。

为求得 $\mathbf{f}_p$，对于 $m \le j \le N$，我们有

$$[\mathbf{f}_p]_{(j,d)} = -\nabla_p\mathbf{\Psi}_{(j-1,d)}. \tag{23}$$

对于 $0 \le j \le m-1$，我们有

$$[\mathbf{f}]_{(j,d)} = [\mathbf{X}_j - \mathbf{x}_j]^{(d)} = 0, \;\; [\mathbf{f}]_{(j,d),(k,d')} = \delta_{(j,d),(k,d')}, \;\; [\mathbf{f}_p]_{(j,d)} = -\frac{\partial\,[\mathbf{x}_j]^{(d)}}{\partial p}. \tag{24}$$

我们通过使用以下方法计算训练所需的梯度

$$\nabla_p\tilde{g} = \tilde{\mathbf{g}}_p - \mathbf{r}^T\mathbf{f}_p. \tag{25}$$

为计算$\mathbf{r}$，我们针对随机轨迹中的每个样本$\mathbf{X}^{[i]}$，求解方程 20– 22 中给出的 m^{th}阶递推关系。 $\tilde{}\mathbf{g}_p = \partial\tilde{}\mathbf{g}/\partial p$ 和$\mathbf{f}_p$ 通过方程 23– 24 计算。最终梯度通过对随机轨迹的 M 个样本取平均来计算，这些样本由 $[i]$索引，并使用 $\nabla_p\bar{g} = \frac{1}{M}\sum_{1=1}^{M}\nabla_p\tilde{g}^{[i]}$。此外，我们还可以利用特定损失函数的结构进一步提高效率，详见第 4.3 节。我们使用方程 18– 25，为 m步生成模型开发高效的梯度计算方法。

参考文献

Gabriele Abbati、Philippe Wenk、Michael A. Osborne、Andreas Krause、Bernhard Sch¨olkopf 和 Stefan Bauer。《AReS 和 MaRS：针对随机微分方程的对抗性与MMD最小化回归》。载于 Kamalika Chaudhuri 和 Ruslan Salakhutdinov 编，《第36届国际机器学习大会论文集》，第 97 卷《机器学习研究论文集》，第 1–10 页。PMLR，2019年6月9日至15日。网址：https://proceedings. mlr.press/v97/abbati19a.html。

Yacine A¨ıt-Sahalia 和 Robert Kimmel。随机波动率模型的最大似然估计。《金融经济学杂志》，83(2):413–452，2007年。

Martin Arjovsky、Soumith Chintala 和 L´eon Bottou。Wasserstein 生成对抗网络。载于 Doina Precup 和 Yee Whye Teh 编，机器学习研究论文集，第 70 卷，机器学习研究论文集，第 214–223 页，澳大利亚悉尼国际会议中心，2017年8月6日至11日。PMLR。网址：http://proceedings.mlr.press/v70/arjovsky17a.html。

N. Aronszajn。再生核理论。《美国数学会会刊》，68(3):337–404，1950年。ISSN 00029947。网址：http://www.jstor.org/stable/ 1990404。

Paul J Atzberger、PeterRKramer 和 CharlesSPeskin。微观长度尺度下流固动力学的随机浸入边界方法。计算物理学杂志，224(2):1255–1292，2007年。

Alain Berlinet 和 Christine Thomas-Agnan。概率与统计学中的再生核希尔伯特空间。施普林格科学与商业媒体，2011年。

Miko-laj Binkowski、Dougal J.Sutherland、Michael Arbel 和 Arthur Gretton。揭开MMD生成对抗网络的神秘面纱。载于国际学习表征会议，2018年。网址：https://openreview.net/forum?id=r1lUOzWCW。

Christopher M. Bishop。《模式识别与机器学习》（信息科学与统计）。Springer-Verlag，柏林、海德堡，2006年。国际标准书号0387310738。

L´eon Bottou 等人。神经网络中的随机梯度学习。*Neuro-Nımes* 论文集，91(8):12，1991年。

Evelyn Buckwar 和 Renate Winkler。随机微分方程的多步法及其在小噪声问题中的应用。《*SIAM*数值分析杂志》，44(2):779–803，2006年。

Richard L. Burden 和 Douglas Faires。《数值分析》。Brooks/Cole Cengage Learning，2010年。

Jean C´ea。最优设计或形状识别，成本函数的方向导数快速计算。*ESAIM*：数学建模与数值分析——模 ´型化数学 ´与数值分 ´析，20(3)：371–402，1986。

Tian Qi Chen、Yulia Rubanova、Jesse Bettencourt 和 David Duvenaud。神经常微分方程。载于 Samy Bengio、Hanna M. Wallach、Hugo Larochelle、Kristen Grauman、Nicol`o Cesa-Bianchi 和 Roman Garnett（编），神经信息处理系统进展 *31*：*2018* 年神经信息处理系统大会年会，*NeurIPS 2018*，*2018*年*12*月*3*日至*8*日，蒙特´利尔，加拿大，第 6572–6583 页，2018 年。网址：http://papers.nips.cc/paper/7892-neural-ordinary-differential-equations。

Thomas M. Cover 和 Joy A. Thomas。信息论基础（*Wiley*电信与信号处理丛书）。Wiley-Interscience，美国，2006年。ISBN 0471241954。

Jan Salomon Cramer。最大似然方法的计量经济学应用。CUP Archive，1989年。

Filipe de Avila Belbute-Peres、Kevin Smith、Kelsey Allen、Josh Tenenbaum 和 J Zico Kolter。用于学习与控制的端到端可微物理学。神经信息处理系统进展，31，2018年。

Ishan Deshpande、Yuan-Ting Hu、Ruoyu Sun、Ayis Pyrros、Nasir Siddiqui、Sanmi Koyejo、Zhizhen Zhao、David Forsyth 和 Alexander G. Schwing。最大切片Wasserstein距离及其在生成对抗网络中的应用，2019年。

Mucong Ding、Constantinos Daskalakis 和 Soheil Feizi。贝叶斯网络上概率散度的次可加性及其在时间序列生成对抗网络中的应用。预印本 *arxiv:2003.00652*，2020年。网址：https://arxiv.org/abs/2003.00652。

Mucong Ding、Constantinos Daskalakis 和 Soheil Feizi。带条件独立图的生成对抗网络：概率散度的次可加性。载于 Arindam Banerjee 和 Kenji Fukumizu 编，第*24*届国际人工智能与统计学会议论文集，第130卷，机器学习研究论文集，第3709–3717页。PMLR，2021年4月13日至15日。网址：https://proceedings.mlr.press/v130/ding21e.html。

OK Dudko、AE Filippov、J Klafter 和 M Urbakh。动态力谱：一种福克–普朗克方法。《化学物理快报》，352(5-6):499–504，2002年。

David Duvenaud。核函数指南：协方差函数建议。网址 *https://www. cs. toronto. edu/˜ duvenaud/cookbook*，2014年。

Gintare Karolina Dziugaite、Daniel M. Roy 和 Zoubin Ghahramani。通过最大均值差异优化训练生成式神经网络。载于《第三十一届人工智能不确定性会议论文集》，UAI' 15，第 258–267 页，阿灵顿，弗吉尼亚州，美国，2015年。AUAI出版社。国际标准书号 9780996643108。网址：https:// arxiv.org/abs/1505.03906。

Brian D Ewald 和 Roger T´emam。地球物理学中随机格式的数值分析。《*SIAM*数值分析杂志》，42(6):2257–2276，2005年。

R. A. Fisher 和 Edward John Russell。论理论统计学的数学基础。伦敦皇家学会哲学汇刊。*A*辑，包含数学或物理性质论文，222(594-604):309–368，1922年。doi: 10.1098/rsta.1922.0009。URL https://royalsocietypublishing.org/doi/abs/10.1098/rsta.1922.0009。Robert Fortet 和 Edith Mourier。经验分布向理论分布的收敛。高等师范学校科学年鉴´辑，第3辑，70 (3):267–285，1953。doi: 10.24033/asens.1013。URL http://www.numdam.org/articles/ 10.24033/asens.1013/。Daan Frenkel 和 Berend Smit。分子模拟：从算法到应用，第1卷。爱思唯尔，2001。C. W. Gardiner。随机方法手册。协同学系列。施普林格，1985。Andrew Gelman、John B Carlin、Hal S Stern 和 Donald B Rubin。贝叶斯数据分析。Chapman and Hall/CRC，1995年。Michael Giles 和 Paul Glasserman。Smoking adjoints：在蒙特卡罗计算中快速评估希腊值。*Risk*，2006年。URL http://citeseerx.ist.psu.edu/viewdoc/ summary?doi=10.1.1.132.2360。Michael B. Giles 和 Niles A. Pierce。伴随方法设计简介。流动、湍流与燃烧，65(3):393–415，2000年。ISSN 1573-1987。doi: 10.1023/A: 1011430410075。URL https://doi.org/10.1023/A:1011430410075。L. T. Giorgini、W. Moon 和 J. S. Wettlaufer。Ornstein-Uhlenbeck 过程的解析生存分析。统计物理学杂志，181(6):2404–2414，2020年。ISSN 1572-9613。doi: 10.1007/s10955-020-02669-y。URL https://doi.org/10.1007/ s10955-020-02669-y。R´emi Goerlich、Minghao Li、Samuel Albert、Giovanni Manfredi、Paul-Antoine Hervieux 和 Cyriaque Genet。光镊中布朗运动的噪声与遍历性质：观察 Ornstein-Uhlenbeck 过程中的状态转换。《物理评论 *E*》，103(3):032132，2021。J. R. Gomez-Solano、L. Bellon、A. Petrosyan 和 S. Ciliberto。受外部随机力驱动系统的稳态涨落关系。《欧洲物理快报》，89(6):60003，2010年3月。doi: 10.1209/0295-5075/89/60003。URL https://doi.org/10.1209/0295-5075/89/60003。Ian Goodfellow、Jean Pouget-Abadie、Mehdi Mirza、Bing Xu、David Warde-Farley、Sherjil Ozair、Aaron Courville 和 Yoshua Bengio。生成对抗网络。载于 Z. Ghahra- mani、M. Welling、C. Cortes、N. D. Lawrence 和 K. Q. Weinberger 编，神经信息处理系统进展第*27*卷，第2672–2680页。Curran Associates, Inc.，2014。URL http://papers.nips.cc/paper/5423-generative-adversarial-nets。pdf。

Ian Goodfellow、Yoshua Bengio 和 Aaron Courville。《深度学习》。麻省理工学院出版社，2016年。国际标准书号 0262035618。网址：https://www.deeplearningbook.org/。

Arthur Gretton、Karsten M. Borgwardt、Malte J. Rasch、Bernhard Sch¨olkopf 和 Alexander Smola。核双样本检验。《机器学习研究杂志》，13(25)：723–773，2012年。网址：http://jmlr.org/papers/v13/gretton12a.html。

Niels Gronbech-Jensen 和 Oded Farago。一种简单有效的 Verlet 型算法，用于模拟朗之万动力学。《style id='1'>《分子物理学》》，111(8)：983–991，2013年2月。国际标准刊号 1362-3028。doi：10.1080/00268976.2012.760055。网址：http://dx.doi.org/10.1080/ 00268976.2012.760055。

B.J. Gross、N. Trask、P. Kuberry 和 P.J. Atzberger。曲面上流体动力学流的流形无网格方法：一种广义移动最小二乘（gmls）方法。《计算物理学杂志》，409:109340，2020年。ISSN 0021-9991。doi: 10.1016/j.jcp.2020.109340。URL https://doi.org/10.1016/j.jcp.2020.109340。

Ishaan Gulrajani、Faruk Ahmed、Martin Arjovsky、Vincent Dumoulin 和 Aaron C Courville。改进 Wasserstein 生成对抗网络的训练。载于 I. Guyon、U. V. Luxburg、S. Bengio、H. Wallach、R. Fergus、S. Vishwanathan 和 R. Gar- nett 编，《神经信息处理系统进展 30》，第 5767–5777 页。Curran Associates, Inc.，2017年。网址：http://papers.nips.cc/paper/ 7159-improved-training-of-wasserstein-gans.pdf。Anders Hald。论最大似然与逆概率和最小二乘法相关的历史。《统计科学》，14(2):214–222，1999年。ISSN 08834237。网址：http: // www.jstor.org/stable/2676741。Lars Peter Hansen。广义矩估计量的大样本性质。《计量经济学》，第 1029–1054 页，1982。A Stan Hurn、JI Jeisman 和 Kenneth A Lindsay。见树木而不见森林：对随机微分方程参数估计方法的批判性评估。《金融计量经济学杂志》，5(3):390–455，2007年。Arieh Iserles。《微分方程数值分析入门》。剑桥大学出版社，2009年。Dev Jasuja 和 P. J. Atzberger。用于刚性时变随机系统的 Magnus 指数积分器。《预印本 arxiv 2212.08978》，2022年。doi: 10.48550/ARXIV.2212.08978。网址：https://arxiv.org/abs/2212.08978。Bjarke Jensen 和 Rolf Poulsen。扩散过程的转移密度。《衍生品杂志》，9(4):18–32，2002年。ISSN 1074-1240。doi: 10.3905/jod.2002.319183。网址：https://jod.pm-research.com/content/9/4/18。K. Jittorntrum。隐函数定理。《优化理论与应用杂志》，25(4):575–577，1978年。ISSN 1573-2878。doi: 10.1007/BF00933522。网址：https://doi.org/10.1007/BF00933522。

C. Kaebe、J. H. Maruhn 和 E. W. Sachs。基于伴随方法的金融市场模型蒙特卡罗校准。《Finance and Stochastics》，13(3):351–379，2009。ISSN 1432-1122。doi: 10.1007/s00780-009-0097-9。网址：https://doi.org/10.1007/s00780-009-0097-9。R. E. Kalman。线性滤波与预测问题的新方法。《Journal of Basic Engineering》，82(1):35–45，1960年03月。ISSN 0021-9223。doi: 10.1115/1.3662552。网址：https://doi.org/10.1115/1.3662552。Patrick Kidger、James Foster、Xuechen Li 和 Terry J Lyons。神经随机微分方程作为无限维生成对抗网络。载于《国际机器学习会议》，第 5453–5463 页。PMLR，2021年。Sangjoon Kim、Neil Shephard 和 Siddhartha Chib。随机波动率：似然推断及与 ARCH 模型的比较。《The Review of Economic Studies》，65(3):361–393，1998年。网址：https://doi.org/10.1111/1467-937X.00050。Gary King。《统一政治方法论：统计推断的似然理论》。密歇根大学出版社，1998年。Diederik P Kingma 和 Jimmy Ba。Adam：一种随机优化方法。《预印本 arXiv:1412.6980》，2014年。网址：https://arxiv.org/abs/1412.6980。P.E. Kloeden 和 E. Platen。《随机微分方程的数值解》。Springer-Verlag，1992 年。Naveen Kodali 、Jacob Abernethy、James Hays 和 Zsolt Kira。论生成对抗网络的收敛性与稳定性。《arXiv 预印本 arXiv:1705.07215》，2017年。S. Kullback 和 R. A. Leibler。论信息与充分性。《数学统计年刊》，22 (1):79–86，1951年03月。doi: 10.1214/aoms/1177729694。网址：https://doi.org/10.1214/ aoms/1177729694。Steven J Large。微观非平衡系统中的随机控制。载于《Dissipation and Control in Microscopic Nonequilibrium Systems》，第 91–111 页。施普林格，2021年。Yann LeCun、Yoshua Bengio 和 Geoffrey Hinton。深度学习。《Nature》，521(7553):436–444，2015年5月。ISSN 1476-4687。网址：https://doi.org/10.1038/nature14539。Yujia Li、Kevin Swersky 和 Richard S. Zemel。生成矩匹配网络。载于 Francis R. Bach 和 David M. Blei 编，《第 32 届国际机器学习会议论文集（ICML 2015），法国里尔，2015年7月6日至11日》，第 37 卷《JMLR Workshop and Conference Proceedings》，第 1718–1727 页。JMLR.org，2015年。网址：http://proceedings.mlr.press/v37/li15.html。J. Lin。基于香农熵的散度度量。《IEEE Transactions on Information Theory》，37(1):145–151，1991年。doi: 10.1109/18.61115。Jacques Louis Lions。《偏微分方程控制系统的最优控制》。Springer-Verlag，柏林，1971年。

Lennart Ljung。《系统辨识》。Prentice Hall，1987年。ISBN ISBN 10: 0136566952 ISBN 13: 9780136566953。URL https://www.amazon.com/ System-Identification-Theory-User-2nd/dp/0136566952。Ryan Lopez 和 Paul J. Atzberger。Gd-vaes：用于学习非线性动力学和降维的几何动态变分自编码器。预印本 *arxiv:2206.05183*，2022年。URL https://arxiv.org/abs/2206.05183。Agnes Lydia 和 Sagayaraj Francis。Adagrad——用于随机梯度下降的优化器。*Int. J. Inf. Comput. Sci*，6(5):566–568，2019年。Anton Mallasto、Guido Mont´ufar 和 Augusto Gerolin。WGAN 如何估计 Wasserstein度量？预印本 *arxiv:1910.03875*，2019年。doi: 10.48550/ARXIV.1910.03875。URL https://arxiv.org/abs/1910.03875。Gisiro Maruyama。连续马尔可夫过程与随机方程。巴勒莫数学学会会刊，4:48–90，1955年。URL https://link.springer.com/ article/10.1007/BF02846028。G. N. Milstein。随机微分方程的近似积分。概率论及其应用，19(3):557–562，1975年。doi: 10.1137/1119062。URL https: //doi.org/10.1137/1119062。Karri Muinonen 和 Edward Bowell。利用贝叶斯概率确定小行星轨道。*Icarus*，104(2):255–279，1993年。ISSN 0019-1035。doi: https://doi.org/10. 1006/icar.1993.1100。URL https://www.sciencedirect.com/science/article/pii/ S0019103583711000。Alfred Muller。积分概率度量及其生成函数类。应用概率进展，29(2):429–443，1997年9月。ISSN 00018678。URL http://www.jstor.org/stable/1428011。Kevin P. Murphy。《机器学习：概率视角》。麻省理工学院出版社，2012年。ISBN 0262018020。John Nash。非合作博弈。数学年刊，54(2):286–295，1951年。ISSN 0003486X。URL http://www.jstor.org/stable/1969529。Jan Nygaard Nielsen、Henrik Madsen 和 Peter C Young。随机微分方程中的参数估计：综述。控制年度评论，24:83–94，2000年。Noam Nisan、Tim Roughgarden、Eva Tardos 和 Vijay V. Vazirani。《算法博弈论》。剑桥大学出版社，2007年。B. Oksendal。《随机微分方程导论》。施普林格，2000年。Martin M. Olsen、Jan Swevers 和 Walter Verdonck。机器人动力学模型的最大似然辨识：实现问题。国际机器人研究杂志，21(2):89–96，2002年。doi: 10.1177/027836402760475379。URL https://doi. org/10.1177/027836402760475379。

Guillermo Owen。《博弈论》。Emerald Group Publishing，2013年。Eulogio Pardo-Ig´uzquiza。空间协方差参数的最大似然估计。数学地质学，30(1):95–108，1998年。Bart Peeters 和 Guido De Roeck。用于运行模态分析的随机系统辨识：综述。《动态系统、测量与控制杂志》，123(4):659–667，2001年2月。ISSN 0022-0434。doi: 10.1115/1.1410370。URL https://doi.org/10. 1115/1.1410370。CE. Rasmussen 和 CKI. Williams。《机器学习中的高斯过程》。自适应计算与机器学习。MIT Press，剑桥，马萨诸塞州，美国，2006年1月。L. E. Reichl。《现代统计物理学教程》。John Wiley & Sons, Inc.，1997年。URL https://onlinelibrary.wiley.com/doi/book/10.1002/9783527690497。Luigi M Ricciardi 和 Laura Sacerdote。Ornstein-Uhlenbeck 过程作为神经元活动模型。生物控制论，35(1):1–9，1979年。Herbert Robbins 和 Sutton Monro。一种随机逼近方法。《数学统计年刊》，22(3):400–407，1951年9月。doi: 10.1214/aoms/1177729586。URL https://doi. org/10.1214/aoms/1177729586。Tim Roughgarden。算法博弈论。《美国计算机协会通讯》，53(7):78–86，2010年。Saburou Saitoh 和 Yoshihiro Sawano。RKHS 的基本性质，2016年。ISSN 1389- 2177。URL https://doi.org/10.1007/978-981-10-0530-5_2。Rainer Sch¨obel 和 Jianwei Zhu。以 Ornstein–Uhlenbeck 过程为模型的随机波动率：扩展。《金融评论》，3(1):23–46，1999年。Adrian FM Smith 和 Gareth O Roberts。通过吉布斯抽样器及相关马尔可夫链蒙特卡罗方法进行贝叶斯计算。《皇家统计学会会刊：*B*辑（方法论）》，55(1):3–23，1993年。Kennan T Smith。《现代分析学入门》。施普林格科学与商业媒体，2012年。Alexander J Smola、A Gretton 和 K Borgwardt。最大均值差异。载于第*13*届国际会议，*ICONIP 2006*，中国香港，*2006*年*10*月*3*日至*6*日：会议论文集，2006年。Helle Sørensen。对离散时间点观测的扩散过程进行参数推断：综述。《国际统计评论》，72(3):337–354，2004年。Bharath K. Sriperumbudur、Kenji Fukumizu、Arthur Gretton、Bernhard Sch¨olkopf 和 Gert R. G. Lanckriet。关于积分概率度量、 ϕ-散度和二元分类。*arxiv*，2009年。URL https://arxiv.org/abs/0901.2698。

Bharath K. Sriperumbudur、Arthur Gretton、Kenji Fukumizu、Bernhard Sch¨olkopf 和 Gert R.G. Lanckriet。概率测度的希尔伯特空间嵌入与度量。机器学习研究杂志，11：1517–1561，2010年8月。ISSN 1532-4435。

J Michael Steele。《随机微积分与金融应用》，第 1 卷。施普林格，2001 年。Panos Stinis、Tobias Hagge、Alexandre M. Tartakovsky 和 Enoch Yeung。生成对抗网络中插值与外推的约束实施。《J. Com- put. Phys.》，397:108844，2019 年。ISSN 0021-9991。doi：10.1016/j.jcp.2019.07.042。URL https://www.sciencedirect.com/science/article/pii/S0021999119305285。Gilbert Strang。《计算科学与工程》，第 791 卷。Wellesley-Cambridge Press Wellesley，2007 年。Luis Enrique Sucar。概率图模型。《计算机视觉与模式识别进展。伦敦：施普林格伦敦。doi》，10(978):1，2015 年。G´abor J. Sz´ekely 和 Maria L. Rizzo。能量统计：一种基于距离的统计学。《统计规划与推断杂志》，143(8):1249 – 1272，2013 年。ISSN 0378-3758。doi：https://doi.org/10.1016/j.jspi.2013.03.018。URL http://www.sciencedirect.com/science/article/pii/S0378375813000633。Gil Tabak 和 Paul J Atzberger。受热涨落影响的惯性流体-结构相互作用的随机约化。《应用数学学会杂志》，75(4):1884–1914，2015 年。URL https://doi.org/10.1137/15M1019088。Daniel Soukup Thomas Pinetz 和 Thomas Pock。Wasserstein 生成对抗网络优化的是什么？见 Zuzana Kukelov ´ a 和 J ´ ulia ´ Skovierov ˇ a（编），《第 23 届计算机视觉冬季研讨会》，2018 年。URL http://cmp.felk.cvut.cz/cvww2018/papers/16.pdf。T. Tieleman 和 G. Hinton。第 6.5 讲——RMSProp：用梯度近期幅值的运行平均值除以梯度。《幻灯片》，2012 年。URL http://www.cs.toronto.edu/~hinton/coursera/lecture6/lec6.pdf。Nathaniel Trask、Ravi Patel、Paul Atzberger 和 Ben Gross。GMLS-Nets：面向非结构化数据的机器学习框架。见《AAAI-MLPS 论文集》，2020 年。URL http://ceur-ws.org/Vol-2587/article_9.pdf。Belinda Tzen 和 Maxim Raginsky。神经随机微分方程：扩散极限下的深度潜在高斯模型。《arXiv 预印本 arXiv:1905.09883》，2019 年。URL https://arxiv.org/abs/1905.09883。George E Uhlenbeck 和 Leonard S Ornstein。论布朗运动理论。《物理评论》，36(5):823，1930 年。Oldrich Vasicek。期限结构的均衡刻画。《金融经济学杂志》，5(2):177–188，1977 年。URL https://www.sciencedirect.com/science/article/abs/pii/0304405X77900162。

Loup Verlet。关于经典流体的计算机“实验”。I. Lennard-Jones 分子的热力学性质。《物理评论》，159:98–103，1967 年 7 月。doi：10.1103/PhysRev.159.98。URL https://link.aps.org/doi/10.1103/PhysRev.159.98。C. Villani。《最优传输：旧与新》。数学科学基础。Springer Berlin Heidelberg，2008 年。国际标准书号 9783540710509。URL https: //books.google.com/books?id=hV8o5R7_5tkC。Martin J. Wainwright 和 Michael I. Jordan。图模型、指数族与变分推断。《机器学习基础与趋势》，1(1-2):1–305，2008 年。doi：10.1561/ 2200000001。Larry Wasserman。《统计学全书：统计推断简明教程》。Springer Publishing Company, Incorporated，2010 年。国际标准书号 1441923225。URL http://www.stat. cmu.edu/~larry/all-of-statistics/。M Wilkinson 和 H R Kennard。微观杆状物与涡量之间排列的模型。《物理学杂志 A：数学与理论》，45(45):455502，2012 年 10 月。doi：10.1088/1751-8113/45/45/455502。URL https://doi.org/10.1088/1751-8113/ 45/45/455502。Christopher K. I. Williams 和 Carl Edward Rasmussen。用于回归的高斯过程。见 D. S. Touretzky、M. C. Mozer 和 M. E. Hasselmo（编），《神经信息处理系统进展 8》，第 514–520 页。MIT Press，1996 年。URL http://papers.nips.cc/paper/1048-gaussian-processes-for-regression.pdf。Bing Xu、Naiyan Wang、Tianqi Chen 和 Mu Li。卷积网络中整流激活函数的经验评估。《arXiv 预印本 arXiv:1505.00853》，2015 年。Kailai Xu、Weiqiang Zhu 和 Eric Darve。利用物理知识学习生成式神经网络。《数学科学研究》，9(2):1–19，2022 年。Liu Yang、Dongkun Zhang 和 George Em Karniadakis。用于随机微分方程的物理信息生成对抗网络。《SIAM 科学计算杂志》，42(1):A292–A317，2020 年。doi：10.1137/18M1225409。Liu Yang、Constantinos Daskalakis 和 George E Karniadakis。生成式集成回归：利用物理信息深度生成模型，从集成观测数据中学习粒子动力学。《SIAM 科学计算杂志》，44(1):B80–B99，2022 年。OV Yushchenko 和 A Yu Badalyan。纳米粒子集体运动的统计描述。《物理评论 E》，85(5):051127，2012 年。