S27 / 23 · CC BY 4.0 · 全文中英逐段精译

底本：arXiv:2303.08797v4（2025-10-09），JMLR 26(209):1–80。作者：Michael S. Albergo、Nicholas M. Boffi、Eric Vanden-Eijnden。

许可核验以原 PDF 首页明示的 CC-BY 4.0 为准；arXiv 摘要页仍列非独占分发许可，两者并记于说明。全文正文、附录 A–C、数学推导、15 图、3 表、5 算法均保留。翻译、排版及标记为“译注”的校勘为本次改编，版权归原作者；CC BY 4.0。

作者单位：纽约大学宇宙学与粒子物理中心（Albergo）；纽约大学 Courant 数学科学研究所（Boffi、Vanden-Eijnden），New York, NY 10012, USA。编辑 Maxim Raginsky。原作记录：2023-12 投稿、2025-09 修订及发表；作者按字母排序，带星号的前两位作者同等贡献。

推荐阅读：定义 1 → 定理 6–8 → 推论 10、18 → 似然控制 → 算法 1–5。原始公式已离线展开；源锚点精确对应官方 HTML，页号按英文前缀匹配或相邻段定位。论文的数值结果为作者报告，本次未运行 GPU 实验。

S001 · Original / 原文 · PDF p.1 · p1.1

Michael S. Albergo, Nicholas M. Boffi, Eric Vanden-Eijnden

Michael S. Albergo、Nicholas M. Boffi、Eric Vanden-Eijnden。

Stochastic Interpolants: A Unifying Framework for Flows and Diffusions

随机插值：统一流与扩散的框架

Michael S. Albergo* albergo@nyu.edu Affiliation: Center for Cosmology and Particle Physics Affiliation: New York University Affiliation: New York, NY 10012, USA Nicholas M. Boffi boffi@cims.nyu.edu thanks: Author ordering alphabetical; authors contributed equally. Affiliation: Courant Institute of Mathematical Sciences Affiliation: New York University Affiliation: New York, NY 10012, USA Eric Vanden-Eijnden eve2@cims.nyu.edu Affiliation: Courant Institute of Mathematical Sciences Affiliation: New York University Affiliation: New York, NY 10012, USA

Abstract

摘要

S002 · Original / 原文 · PDF p.1 · abstract1.1

A class of generative models that unifies flow-based and diffusion-based methods is introduced. These models extend the framework proposed in *Albergo and Vanden-Eijnden (2023)*, enabling the use of a broad class of continuous-time stochastic processes called ‘stochastic interpolants’ to bridge any two probability density functions exactly in finite time. These interpolants are built by combining data from the two prescribed densities with an additional latent variable that shapes the bridge in a flexible way. The time-dependent probability density function of the stochastic interpolant is shown to satisfy a first-order transport equation as well as a family of forward and backward Fokker-Planck equations with tunable diffusion coefficient. Upon consideration of the time evolution of an individual sample, this viewpoint immediately leads to both deterministic and stochastic generative models based on probability flow equations or stochastic differential equations with an adjustable level of noise. The drift coefficients entering these models are time-dependent velocity fields characterized as the unique minimizers of simple quadratic objective functions, one of which is a new objective for the score of the interpolant density. We show that minimization of these quadratic objectives leads to control of the likelihood for generative models built

upon stochastic dynamics, while likelihood control for deterministic dynamics is more stringent. We also construct estimators for the likelihood and the cross-entropy of interpolant-based generative models, and we discuss connections with other methods such as score-based diffusion models, stochastic localization processes, probabilistic denoising techniques, and rectifying flows. In addition, we demonstrate that stochastic interpolants recover the Schrödinger bridge between the two target densities when explicitly optimizing over the interpolant. Finally, algorithmic aspects are discussed and the approach is illustrated on numerical examples.

本文提出一类统一流模型与扩散模型的生成模型，扩展了 Albergo 与 Vanden-Eijnden（2023）的框架。称为“随机插值”的广泛连续时间随机过程，可以在有限时间内精确连接任意两个概率密度。构造时，把来自两个指定密度的数据与一个额外潜变量结合，使桥接具有灵活形状。随机插值的时变密度满足一阶输运方程，以及具有可调扩散系数的一族前向、后向 Fokker–Planck 方程。考察单样本演化后，立即得到基于概率流方程的确定性生成模型，以及噪声水平可调的随机微分方程生成模型。这些模型中的漂移是时变速度场，可刻画为简单二次目标的唯一极小化解，其中包含一个新的插值密度得分目标。最小化这些目标能控制随机动力学生成模型的似然；确定性动力学的似然控制要求更强。本文还构造似然与交叉熵估计量，讨论与得分扩散、随机局部化、概率去噪、矫正流的联系，并证明显式优化插值可以恢复两目标密度之间的 Schrödinger 桥。最后讨论算法，并用数值例子演示。

heading: 26 2025 1- 12/23; Revised 9/25 9/25 23-1605shortheadings: Stochastic Interpolants / Albergo, Boffi, and Vanden-Eijndenfirstpage: 1editor: Maxim Raginsky

1 Introduction
1 引言

1.1 Background and motivation
1.1 背景与动机

S003 · Original / 原文 · PDF p.3 · S1.SS1.p1.1

Dynamical approaches for deterministic and stochastic transport have become a central theme in contemporary generative modeling research. At the heart of progress is the idea to use ordinary or stochastic differential equations (ODEs/SDEs) to continuously transform samples from a base probability density function (PDF) ρ_0 into samples from a target PDF ρ_1 (or vice-versa), and the realization that inference over the velocity field in these equations can be formulated as an empirical risk minimization problem over a parametric class of functions *(Grathwohl et al., 2019; Song and Ermon, 2019; Ho et al., 2020; Song et al., 2021b; Ben-Hamu et al., 2022; Albergo and Vanden-Eijnden, 2023; Liu et al., 2023b; Lipman et al., 2023)*.

确定性与随机输运的动力学方法已成为当代生成建模的核心方向。其关键思想是使用常微分方程或随机微分方程（ODE/SDE），把基分布 ρ_0 的样本连续变换为目标分布 ρ_1 的样本，或反向变换；同时认识到，对这些方程速度场的推断可写成参数化函数类上的经验风险最小化问题（Grathwohl 等，2019；Song 与 Ermon，2019；Ho 等，2020；Song 等，2021b；Ben-Hamu 等，2022；Albergo 与 Vanden-Eijnden，2023；Liu 等，2023b；Lipman 等，2023）。

S004 · Original / 原文 · PDF p.3 · S1.SS1.p2.1

A major milestone was the introduction of score-based diffusion methods (SBDM) *(Song et al., 2021b)*, which map an arbitrary density into a standard Gaussian by passing samples through an Ornstein-Uhlenbeck (OU) process. The key insight of SBDM is that this process can be reversed by introducing a backwards SDE whose drift coefficient depends on the score of the time-dependent density of the process. By learning this score – which can be done by minimization of a quadratic objective function known as the denoising loss *(Vincent, 2011)* – the backwards SDE can be used as a generative model that maps Gaussian noise into data from the target. Though theoretically exact, the mapping takes infinite time in both directions, and hence must be truncated in practice.

重要里程碑是基于得分的扩散方法 SBDM（Song 等，2021b）：把样本输入 Ornstein–Uhlenbeck（OU）过程，将任意密度映射到标准高斯。关键在于可引入一个后向 SDE 逆转该过程，其漂移依赖过程时变密度的得分。该得分可通过最小化称为去噪损失的二次目标来学习（Vincent，2011）；学得后，后向 SDE 便能把高斯噪声映射为目标数据。此映射在理论上精确，但双向都需要无限时间，所以实践中必须截断。

S005 · Original / 原文 · PDF p.3 · S1.SS1.p3.1

While diffusion-based methods have become state-of-the-art for tasks such as image generation, there remains considerable interest in developing methods that bridge two arbitrary densities (rather than requiring one to be Gaussian), that accomplish the transport exactly, and that do so on a finite time interval. Moreover, while the highest quality results from score-based diffusion were originally obtained using SDEs *(Song et al., 2021b)*, this has been challenged by recent works that find equivalent or better performance with ODE-based methods if the score is learned sufficiently well *(Karras et al., 2022)*. If made to match the performance of their stochastic counterparts, ODE-based methods exhibit a number of desirable characteristics, such as an exact, computationally tractable formula for the likelihood and the easy application of well-developed adaptive integration schemes for sampling. It is an open question of significant practical importance to understand if there exists a separation in sample quality between generative models based on deterministic dynamics and those based on stochastic dynamics.

扩散方法虽已在图像生成等任务达到先进水平，人们仍希望连接任意两个密度，而非要求其中一个为高斯；并且要求输运精确、在有限时间内完成。得分扩散最初用 SDE 得到最高质量结果（Song 等，2021b），但近期工作发现，只要得分学得足够准确，ODE 方法也能达到相同甚至更好性能（Karras 等，2022）。若质量与随机方法相当，ODE 有多项优点：似然有精确且计算可行的公式，采样可方便使用成熟的自适应积分器。确定性与随机动力学生成模型的样本质量是否存在本质差距，是一个具有重要实践意义的开放问题。

S006 · Original / 原文 · PDF p.4 · S1.SS1.p4.1

In order to satisfy the desirable characteristics outlined in the previous paragraph, we develop a framework for generative modeling based on the method proposed in *Albergo and Vanden-Eijnden (2023)*, which is built on the notion of a stochastic interpolant x_t used to bridge two arbitrary densities ρ_0 and ρ_1. We will consider more general designs below, but as one example the reader can keep in mind:

为满足上述要求，我们基于 Albergo 与 Vanden-Eijnden（2023）的方法建立生成框架，核心是连接任意密度 ρ_0、ρ_1 的随机插值 x_t。后文会考虑更一般设计，这里先给出一个可供读者把握的例子。

$$x_t = (1-t)x_0 + tx_1 + \sqrt{2t(1-t)}z, \quad t \in [0,1], \tag{1.1}$$

S007 · Original / 原文 · PDF p.4 · S1.SS1.p4.2

where x_0, x_1, and z are random variables drawn independently from ρ_0, ρ_1, and the standard Gaussian density $\mathsf{N}(0, Id)$, respectively. The stochastic interpolant x_t defined in (1.1) is a continuous-time stochastic process that, by construction, satisfies $x_{t=0} = x_0 \sim \rho_0$ and $x_{t=1} = x_1 \sim \rho_1$. Its paths therefore exactly bridge between samples from ρ_0 at $t = 0$ and from ρ_1 at $t = 1$. A key observation is that:

其中 x_0、x_1、z 分别独立取自 ρ_0、ρ_1 和标准高斯 $N(0,Id)$。式（1.1）的 x_t 是连续时间随机过程，按构造满足 $x_{t=0}=x_0\sim\rho_0$、$x_{t=1}=x_1\sim\rho_1$，因此其路径精确连接了两个端点分布的样本。关键观察如下。

> S008 · Original / 原文 · PDF p.4 · S1.SS1.p4.3.1
>
> The law of the interpolant x_t at any time $t \in [0,1]$ can be realized by many different processes, including an ODE and forward and backward SDEs whose drifts can be learned from data.
>
> 对任意 $t\in[0,1]$，插值 x_t 的分布可由许多不同过程实现，包括漂移能从数据学习的 ODE、前向 SDE 和后向 SDE。

S009 · Original / 原文 · PDF p.4 · S1.SS1.p4.4

To see why this is the case, one must consider the probability distribution of the interpolant x_t. As shown below, for a large class of densities ρ_0 and ρ_1 supported on $\mathbb{R}^d$, this distribution is absolutely continuous with respect to the Lebesgue measure. Moreover, its time-dependent density $\rho(t)$ satisfies a first-order transport equation and a family of forward and backward Fokker-Planck equations in which the diffusion coefficient can be varied at will. Out of these equations, we can readily derive generative models that satisfy ODEs and SDEs, respectively, and whose densities at time t are given by $\rho(t)$.

要理解这一点，需要考察 x_t 的概率分布。下文证明，对支撑在 R^d 上的一大类 ρ_0、ρ_1，该分布关于 Lebesgue 测度绝对连续，其时变密度 $\rho(t)$ 满足一阶输运方程，以及扩散系数可任意调节的一族前后向 Fokker–Planck 方程。由这些方程可直接导出分别满足 ODE、SDE 的生成模型，且 t 时刻密度都为 $\rho(t)$。

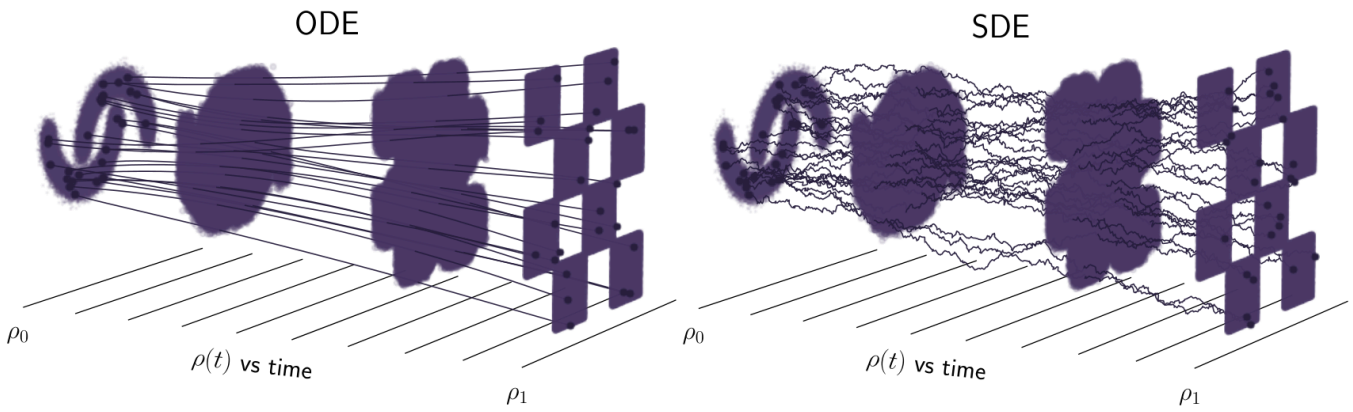

Figure 1: The stochastic interpolant paradigm. Example generative models based on the proposed framework, which connects two densities ρ_0 and ρ_1 using samples from both. The design of the time-dependent probability density $\rho(t)$ that bridges between ρ_0 and ρ_1 is separated from the choice of how to sample it, which can be accomplished with deterministic or stochastic generative models. Left panel: Sampling with a deterministic (ODE) generative model known as the probability flow equation. Right panel: Sampling with a stochastic generative model given by an SDE with a tunable diffusion coefficient. The probability flow equation and the SDE have different paths, but their time-dependent density $\rho(t)$ is the same. Moreover, the two equations rely on the same estimates for the velocity and the score.

图 1：随机插值范式。利用两个端点密度的样本构造生成模型；连接 ρ₀、ρ₁ 的时变密度 ρ(t) 的设计，与用什么过程采样相分离。左图是概率流 ODE，右图是扩散系数可调的 SDE。二者样本路径不同，但单时刻密度相同，并依赖相同的速度、得分估计。

S010 · Original / 原文 · PDF p.4 · S1.SS1.p5.1

Interestingly, the drift coefficients entering these ODEs/SDEs are the unique minimizers of quadratic objective functions that can be estimated empirically using data from ρ_0, ρ_1, and $\mathsf{N}(0, Id)$. The resulting least-squares regression problem allows us to estimate the drift coefficients of the ODE/SDEs, which can then be used to push samples from ρ_0 onto new samples from ρ_1 and vice-versa.

这些 ODE/SDE 的漂移系数是二次目标的唯一极小化解，而目标可用来自 ρ₀、ρ₁、N(0,Id) 的数据做经验估计。因此通过最小二乘回归可估计漂移，再将 ρ₀ 样本推至新的 ρ₁ 样本，反向亦然。

1.2 Main contributions and organization

1.2 主要贡献与组织

S011 · Original / 原文 · PDF p.4 · S1.SS2.p1.1

The approach introduced here is a versatile way to build generative models that unifies and extends many existing algorithms. In Sec. 2, we develop the framework in full generality, where we emphasize the following key contributions:

本文提供一种灵活的生成模型构造方式，统一并扩展许多已有算法。第 2 节给出一般框架，关键贡献如下。

- •

  S012 · Original / 原文 · PDF p.4 · S1.I1.i1.p1.1

  We prove that the stochastic interpolant defined in Section 2.1 has a distribution that is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$, and that its density $\rho(t)$ satisfies a first-order transport equation (TE) as well as a family of forward and backward Fokker-Planck equations (FPEs) with tunable diffusion coefficients.

  证明第 2.1 节定义的随机插值在 R^d 上的分布关于 Lebesgue 测度绝对连续，且其密度 ρ(t) 同时满足一阶输运方程 TE 和扩散系数可调的一族前后向 Fokker–Planck 方程 FPE。

- •

  S013 · Original / 原文 · PDF p.4 · S1.I1.i2.p1.1

We show how the stochastic interpolant can be used to learn the drift coefficients that enter the TE and the FPEs. We characterize these coefficients as the minimizers of simple quadratic objective functions given in Section 2.2. We introduce a new objective for the score $\nabla \log \rho(t)$ of the interpolant density, as well as an objective function for learning a denoiser η_z, which we relate to the score.

说明如何利用随机插值学习 TE、FPE 的漂移，并在第 2.2 节将其刻画为简单二次目标的极小化解。提出新的插值密度得分 ∇logρ(t) 目标，以及学习去噪器 ηz 的目标，并建立去噪器与得分的联系。

- •

S014 · Original / 原文 · PDF p.5 · S1.I1.i3.p1.1

In Section 2.3, we derive ordinary and stochastic differential equations associated with the TE and FPEs that lead to deterministic and stochastic generative models. In Section 2.4, we show that regressing the drift for SDE-based models controls the likelihood, but that regressing the drift alone is not sufficient for ODE-based models, which must also minimize a Fisher divergence. We show how to optimally tune the diffusion coefficient to maximize the likelihood for SDEs.

第 2.3 节从 TE、FPE 导出确定性和随机生成模型的 ODE、SDE。第 2.4 节说明，对 SDE 回归漂移能够控制似然；对 ODE，仅回归漂移不够，还须最小化 Fisher 散度。我们还说明如何最优调节扩散系数，以最大化 SDE 的似然。

- •

S015 · Original / 原文 · PDF p.5 · S1.I1.i4.p1.1

In Section 2.5, we develop a general formula to evaluate the likelihood of SDE-based generative models that serves as a natural counterpart to the continuous change-of-variables formula commonly used to compute the likelihood of ODE-based models. In addition, we give formulas to estimate the cross-entropy.

第 2.5 节给出计算 SDE 生成模型似然的一般公式，它自然对应于 ODE 模型常用的连续变量替换公式；同时给出交叉熵估计公式。

S016 · Original / 原文 · PDF p.5 · S1.SS2.p2.1

In Section 3, we discuss instantiations of the stochastic interpolant method. In Section 3.4 we first show that interpolants are equivalent to a class of stochastic bridges, but that they avoid the need for Doob's h-transform, which is generically unknown; we show that this simplifies the construction of a broad class of generative models. In Section 3.2, we define the one-sided interpolant, which corresponds to the conventional setting in which the base ρ_0 is taken to be a Gaussian. With a Gaussian base, several aspects of the interpolant simplify, and we detail the corresponding objective functions. In Section 3.3, we introduce a mirror interpolant in which the base ρ_0 and the target ρ_1 are identical. Finally, in Section 3.4, we show how the interpolant framework leads to a natural formulation of the Schrödinger bridge problem between two densities.

第 3 节讨论随机插值的不同实例。原文先指第 3.4 节说明插值等价于一类随机桥，但不需要通常未知的 Doob h 变换，因此简化了一大类生成模型的构造。第 3.2 节定义单侧插值，即基分布 ρ_0 为高斯的常规设置；高斯基分布可简化多方面结构，我们给出相应目标。第 3.3 节引入基分布、目标相同的镜像插值。最后第 3.4 节说明，插值框架自然导出两密度间 Schrödinger 桥的表述。译注：关于随机桥的一般讨论亦见第 3.1 节，保留原文此处交叉引用。

S017 · Original / 原文 · PDF p.5 · S1.SS2.p3.1

In Section 4, we discuss a special case in which the interpolant is spatially linear in x_0 and x_1. In this case, the velocity field can be factorized, which we show in Section 4.1 leads to a simpler learning problem. We detail specific choices of linear interpolants in Section 4.2, and in Section 4.3 we illustrate how these choices influence the performance of the resulting generative model, with a particular focus on the role of the latent variable and the diffusion coefficient. For exposition, we focus on Gaussian mixture densities, for which the drift coefficients can be computed analytically. We provide the resulting formula in Appendix A. Finally, in Section 4.4, we discuss the case of spatially linear one-sided interpolants.

第 4 节讨论对 x_0、x_1 空间线性的特殊插值。第 4.1 节说明速度场可分解，从而简化学习问题。第 4.2 节给出具体线性插值，第 4.3 节分析这些选择如何影响生成模型，特别关注潜变量与扩散系数的作用。为便于解释，重点考察

可解析计算漂移的高斯混合密度，公式见附录 A。第 4.4 节讨论空间线性单侧插值。

S018 · Original / 原文 · PDF p.6 · S1.SS2.p4.1

In Section 5, we formalize the connection between stochastic interpolants and related classes of generative models. In Section 5.1, we show that score-based diffusion models can be re-written as one-sided interpolants after a reparameterization of time; we highlight how this approach eliminates singularities that appear when naively compressing score-based diffusion onto a finite-time interval. In Section 5.2, we show how interpolants can be used to derive the Bayes-optimal estimator for a denoiser, and we show how this approach can be iterated to create a generative model. In Section 5.3, we consider the possibility of rectifying the flow map of a learned generative model. We show that the rectification procedure does not change the underlying generative model, though it may change the time-dependent density of the interpolant.

第 5 节形式化随机插值与相关生成模型的联系。第 5.1 节在重参数化时间后，把得分扩散改写为单侧插值；强调这种方式消除了直接把得分扩散压缩到有限时间时出现的奇性。第 5.2 节利用插值导出去噪器的 Bayes 最优估计，并说明如何迭代形成生成模型。第 5.3 节考虑对学得生成模型的流映射做矫正，证明该过程不改变底层生成模型，但可能改变插值的时变密度。

S019 · Original / 原文 · PDF p.6 · S1.SS2.p5.1

In Section 6, we provide the details of practical algorithms associated with the mathematical results presented above. In Section 6.1, we describe how to numerically estimate the objectives given empirical datasets from the base and the target. In Section 6.2, we complement this discussion on learning with algorithms for sampling with the ODE or an SDE.

第 6 节给出对应数学结果的实践算法。第 6.1 节说明用基分布和目标分布经验数据估计目标；第 6.2 节补充 ODE、SDE 采样算法。

S020 · Original / 原文 · PDF p.6 · S1.SS2.p6.1

We provide numerical demonstrations in line with these recommendations in Section 7, and we conclude with some remarks in Section 8.

第 7 节给出遵循这些建议的数值演示，第 8 节作结。

1.3 Related work

1.3 相关工作

Deterministic Transport and Normalizing Flows.

确定性输运与归一化流

S021 · Original / 原文 · PDF p.6 · S1.SS3.SSS0.Px1.p1.1

Transport-based sampling and density estimation has its contemporary roots in Gaussianizing data via maximum entropy methods *(Friedman, 1987; Chen and Gopinath, 2000; Tabak and Vanden-Eijnden, 2010; Tabak and Turner, 2013)*. The change of measure under such transformation is the backbone of normalizing flow models. The first neural network realizations of these methods arose through imposing clever structure on the transformation to make the change of measure tractable in discrete, sequential steps *(Rezende and Mohamed, 2015; Dinh et al., 2017; Papamakarios et al., 2017; Huang et al., 2018; Durkan et al., 2019)*. A continuous time version of this procedure was made possible by viewing the map $T = X_t(x)$ as the solution of an ODE *(Chen et al., 2018; Grathwohl et al., 2019)*, whose parametric drift defining the transport is learned via maximum likelihood estimation. Training this way is intractable at scale, as computing the gradient of the objective via the adjoint method requires simulating an ODE. Various methods have introduced regularization on the path taken between the two densities to make the ODE solves more efficient *(Finlay et al., 2020; Onken et al., 2021; Tong et al., 2020)*, but the fundamental difficulty remains. We also work in continuous time; however, our approach allows us to learn the drift without simulation of the dynamics, and can be formulated at sample generation time through either deterministic or stochastic transport.

基于输运的采样与密度估计，其现代起源是通过最大熵方法将数据高斯化（Friedman，1987；Chen 与 Gopinath，2000；Tabak 与 Vanden-Eijnden，2010；Tabak 与 Turner，2013）。这种变换下的测度变换是归

一化流模型的基础。早期神经网络实现通过精心设计变换结构，使离散连续步骤中的测度变换可计算（Rezende 与 Mohamed，2015；Dinh 等，2017；Papamakarios 等，2017；Huang 等，2018；Durkan 等，2019）。将映射 $T=X_t(x)$ 看作 ODE 解即可得到连续时间版本（Chen 等，2018；Grathwohl 等，2019），参数化漂移用极大似然学习。但通过伴随法计算目标梯度需要模拟 ODE，难以大规模训练。路径正则可提高积分效率（Finlay 等，2020；Onken 等，2021；Tong 等，2020），却未消除根本困难。本文同样使用连续时间，但学习漂移无需模拟动力学，采样时则可使用确定性或随机输运。

Stochastic Transport and Score-Based Diffusion Models (SBDMs).

随机输运与基于得分的扩散模型

S022 · Original / 原文 · PDF p.7 · S1.SS3.SSS0.Px2.p1.1

Complementary to approaches based on deterministic maps, recent works have realized that connecting a data distribution to a Gaussian density can be viewed as the evolution of an Ornstein-Ulhenbeck (OU) process which gradually degrades samples from the distribution of interest to Gaussian noise *(Sohl-Dickstein et al., 2015; Ho et al., 2020; Song and Ermon, 2019; Song et al., 2021b)*. The OU process specifies a path in the space of probability densities; this path is simple to traverse in the forward direction by addition of noise, and can be reversed if access to the score of the time-dependent density $\nabla \log \rho(t)$ is available. This score can be approximated through solution of a least-squares regression problem *(Hyvärinen, 2005; Vincent, 2011)*, and the target can be sampled by reversing the path once the score has been learned. Interestingly, the resulting forward and backward stochastic processes have an equivalent formulation (at the distribution level) in terms of a deterministic probability flow equation, first noted by *Bakry and Émery (1985); Otto (2000); Kim and Milman (2010)* and then applied in *Maoutsa et al. (2020); Song et al. (2021a); Kingma et al. (2021); Boffi and Vanden-Eijnden (2023)*. The probability flow formulation is useful for density estimation and cross-entropy calculations, but it is worth noting that the probability flow and the reverse-time SDE will have densities that differ when using an approximate score. The SBDM framework, as it has been originally presented, has a number of features which are not a priori well motivated, including the dependence on mapping to a normal density, the complicated tuning of the time parameterization and noise scheduling *(Xiao et al., 2022; Hoogeboom et al., 2023)*, and the choice of the underlying stochastic dynamics *(Dockhorn et al., 2022; Karras et al., 2022)*.

与确定性映射互补，近期工作将数据分布连接到高斯视作 OU 过程：样本逐渐退化为高斯噪声（Sohl-Dickstein 等，2015；Ho 等，2020；Song 与 Ermon，2019；Song 等，2021b）。OU 指定密度空间中的路径，前向加噪容易遍历，若有时变得分 $\nabla \log \rho(t)$，便可逆向。得分可通过最小二乘回归近似（Hyvärinen，2005；Vincent，2011），学习后逆路径即可采样目标。前后向随机过程在分布层面还有确定性概率流的等价表述，Bakry 与 Émery（1985）、Otto（2000）、Kim 与 Milman（2010）首先注意，后由 Maoutsa 等（2020）、Song 等（2021a）、Kingma 等（2021）、Boffi 与 Vanden-Eijnden（2023）应用。概率流便于密度估计和交叉熵计算，但用近似得分时，它与逆时间 SDE 的密度会不同。原始 SBDM 的一些设计并无先验充分动机，包括依赖映射到正态密度、复杂时间参数化及噪声调度（Xiao 等，2022；Hoogeboom 等，2023），以及底层随机动力学选择（Dockhorn 等，2022；Karras 等，2022）。

Stochastic bridges.

随机桥

S023 · Original / 原文 · PDF p.7 · S1.SS3.SSS0.Px3.p1.1

Starting with *(Peluchetti, 2022)* there has been some recent effort *(Liu et al., 2022; Liu et al., 2023a; Somnath et al., 2023)* to remove the dependence of SBDMs on the OU process via stochastic bridges, which can be used to connect two arbitrary densities in finite time. As another step in this direction, we observe here that the key idea behind SBDMs – the bridging of two densities via a time-dependent density whose evolution equation is available – can be generalized to a much wider class of processes in a straightforward and computationally accessible manner. This viewpoint highlights the key property that the construction of the bridge between the two densities is decoupled from the process used to sample it.

自 Peluchetti（2022）起，一些研究（Liu 等，2022、2023a；Somnath 等，2023）用有限时间连接任意密度的随机桥，解除 SBDM 对 OU 的依赖。本文进一步指出，SBDM 的核心——借助演化方程已知的时变密度来连接两

个密度——可直接且计算可行地推广至更广过程族。此观点突出了“构造桥接”与“采样该桥接所用过程”相互解耦的性质。

Stochastic Interpolants, Rectified Flows, and Flow matching.

随机插值、矫正流与流匹配

S024 · Original / 原文 · PDF p.7 · S1.SS3.SSS0.Px4.p1.1

Variants of the stochastic interpolant method presented in *Albergo and Vanden-Eijnden (2023)* were also presented in *Liu et al. (2023b); Lipman et al. (2023)*. In *Liu et al. (2023b)*, a linear interpolant was proposed with a focus on straight paths. This was employed as a step toward rectifying the transport paths *(Liu, 2022)* through a procedure that improves sampling efficiency but introduces a bias. In *Lipman et al. (2023)*, the interpolant picture was assembled from the perspective of conditional probability paths connecting to a Gaussian, where a noise convolution was used to improve the learning at the cost of biasing the method. Extensions of *Lipman et al. (2023)* were presented in *Tong et al. (2023)* that generalize the method beyond the Gaussian base density. In the method proposed here, we introduce an unbiased means to incorporate noise into the process, both via the introduction of a latent variable into the stochastic interpolant and the inclusion of a tunable diffusion coefficient in the associated stochastic generative models. We provide theoretical and practical motivation for the presence of these noise terms.

Albergo 与 Vanden-Eijnden（2023）的随机插值方法也有 Liu 等（2023b）、Lipman 等（2023）的变体。前者提出强调直路径的线性插值，进一步通过矫正输运路径（Liu，2022）提高采样效率，但会引入偏差。后者从连接高斯的条件概率路径出发，通过噪声卷积改善学习，代价也是偏差。Tong 等（2023）把其推广至非高斯基分布。本文给出无偏的加噪方式：既在随机插值中加入潜变量，也在相应随机生成模型中加入可调扩散系数，并提供理论和实践动机。

Optimal Transport and Schrödinger Bridges.

最优输运与 Schrödinger 桥

S025 · Original / 原文 · PDF p.8 · S1.SS3.SSS0.Px5.p1.1

There is both theoretical and practical interest in minimizing the transport cost of connecting ρ_0 and ρ_1. In the case of deterministic maps, this is characterized by the optimal transport problem, and in the case of diffusive maps, by the Schrödinger Bridge problem *(Villani, 2009; Chen et al., 2021)*. Formally, these two problems can be related by viewing the Schrödinger Bridge as an entropy-regularized optimal transport. Optimal transport has primarily been employed as a means to regularize flow-based methods by imposing either a path length penalty *(Zhang et al., 2018; Onken et al., 2021; Finlay et al., 2020; Tong et al., 2020)* or structure on the parameterization itself *(Huang et al., 2021; Yang and Karniadakis, 2022)*. A variety of recent works have formulated the Schrödinger problem in the context of a learnable diffusion *(Bortoli et al., 2021; Su et al., 2023; Chen et al., 2022b)*. For the case of Gaussians, recent work has also identified an analytical solution *(Bunne et al., 2022)*. In the interpolant framework, *(Albergo and Vanden-Eijnden, 2023; Liu et al., 2023b; Lipman et al., 2023; Tong et al., 2023)* all propose optimal transport extensions to the learning procedure. The method proposed in *Liu et al. (2023b); Liu (2022)* allows one to sequentially lower the transport cost through rectification, at the cost of introducing a bias unless the velocity field is perfectly learned. The method proposed in *Albergo and Vanden-Eijnden (2023)* is an unbiased framework at the cost of solving an additional optimization problem over the interpolant function. The statement of optimal transport in *Lipman et al. (2023)* only applies to Gaussians, but is shown to be practically useful in experimental demonstrations.

理论与实践都关心连接 ρ_0、ρ_1 的输运成本最小化。确定性映射对应最优输运，扩散映射对应 Schrödinger 桥（Villani，2009；Chen 等，2021）；形式上后者可视作熵正则最优输运。最优输运主要用于正则化流模型：惩罚路径长度（Zhang 等，2018；Onken 等，2021；Finlay 等，2020；Tong 等，2020），或约束参数化结构（Huang 等，2021；Yang 与 Karniadakis，2022）。一些研究在可学习扩散中表述 Schrödinger 问题（Bortoli 等，2021；Su 等，2023；Chen 等，2022b）；高斯情形还有解析解（Bunne 等，2022）。插值框架中的多项工作均提出最优输运扩展。Liu 等（2023b）与 Liu（2022）通过矫正逐步降低输运成本，但除非速度学得完美，否则引入偏差；Albergo 与 Vanden-Eijnden（2023）无偏，但须额外优化插值函数。Lipman 等（2023）的最优输运论述只适用于高斯，不过实验显示有实践价值。

S026 · Original / 原文 · PDF p.8 · S1.SS3.SSS0.Px5.p2.1

In the method proposed below, we provide two approaches for optimizing the transport under a stochastic dynamics. Our primary approach, based on the scheme introduced in *Albergo and Vanden-Eijnden (2023)*, is presented in Section 3.4. It offers an alternative route to solve the Schrödinger bridge problem under the Benamou-Brenier hydrodynamic formulation of transport by maximizing over the interpolant *(Benamou and Brenier, 2000)*. However, we stress that this additional optimization step is not necessary in practice, as our approach leads to bias-free generative models for any fixed interpolant. In addition, Section 5.3 discusses an unbiased variant of the rectification scheme proposed in *Liu et al. (2023b)*.

下文给出两种优化随机动力学输运的方式。主要方法见第 3.4 节，基于 Albergo 与 Vanden-Eijnden（2023）：在 Benamou–Brenier 流体动力学表述（Benamou 与 Brenier，2000）下，对插值做最大化，为求 Schrödinger 桥提供另一条路线。但实践中并不需要这项额外优化，因为任意固定插值都可导出无偏生成模型。第 5.3 节还讨论 Liu 等（2023b）矫正方案的无偏变体。

Convergence bounds.

收敛界

S027 · Original / 原文 · PDF p.8 · S1.SS3.SSS0.Px6.p1.1

Inspired by the successes of score-based diffusion, significant recent research effort has been expended to understand the control that can be obtained on suitable distances between the distribution of the generative model and the target data distribution, such as KL, W_2, or TV. Perhaps the first line of work in this direction is *Song et al. (2021a)*, which showed that standard score-based diffusion training techniques bound the likelihood of the resulting SDE model. Importantly, as we show here, the likelihood of the corresponding probability flow is not bounded in general by this technique, as first highlighted in the context of SBDM by *Lu et al. (2022)*. Control for SBDM-based techniques was later quantified more rigorously under the assumption of functional inequalities in a discretized setting by *Lee et al. (2022)*, which were removed by *Lee et al. (2023)* and *Chen et al. (2023)* via Girsanov-based techniques. Most relevant to the PDE-based methods considered here is *Chen et al. (2022a)*, which applies similar techniques to our own in the SBDM context to obtain sharp guarantees with minimal assumptions.

得分扩散的成功激发了大量研究，旨在控制生成分布与目标之间的 KL、W_2、TV 等距离。较早的 Song 等（2021a）证明，标准得分扩散训练可以界定相应 SDE 模型的似然。重要的是，此技术通常不能界定对应概率流的似然；Lu 等（2022）在 SBDM 中首先指出，本文也予以说明。Lee 等（2022）在离散化设置和函数不等式假设下作了严格量化，Lee 等（2023）、Chen 等（2023）用 Girsanov 技术移除这些假设。与本文 PDE 方法最相关的是 Chen 等（2022a），他们在 SBDM 中以相似技术、较少假设得到紧保证。

1.4 Notation

1.4 记号

S028 · Original / 原文 · PDF p.9 · S1.SS4.p1.1

Throughout, we denote probability density functions as $\rho_0(x)$, $\rho_1(x)$, and $\rho(t,x)$, with $t \in [0,1]$ and $x \in \mathbb{R}^d$, omitting the function arguments when clear from the context. We proceed similarly for other functions of time and space, such as $b(t,x)$ or $I(t,x_0,x_1)$. We use the subscript t to denote the time-dependency of stochastic processes, such as the stochastic interpolant x_t or the Wiener process W_t. To specify that the random variable x_0 is drawn from the probability distribution with density ρ_0, say, with a slight abuse of notations we use $x_0 \sim \rho_0$. Similarly, we use $\mathsf{N}(0, Id)$ to denote both the density and the distribution of the Gaussian random variable with mean zero and covariance identity. We denote expectation by $\mathbb{E}$, and usually specify the random variables this expectation is taken over. With a slight abuse of terminology, we say that the law of the process x_t is $\rho(t)$ if $\rho(t)$ is the density of the probability distribution of x_t at time t.

全文用 $\rho_0(x)$、$\rho_1(x)$、$\rho(t,x)$ 表示概率密度，$t\in[0,1]$、$x\in\mathsf{R}^d$；语境明确时省略自变量。$b(t,x)$、$I(t,x_0,x_1)$ 等函数也同样处理。随机过程的时间依赖用下标 t 表示，如 x_t、W_t。略微滥用记号，以 $x_0\sim\rho_0$ 表示随机变量取自密度为 ρ_0 的分布；$\mathsf{N}(0,\mathrm{Id})$ 同时表示零均值、单位协方差高斯的密度与分布。E 表示期望，通常注明所平均的随机变量。若 $\rho(t)$ 是 x_t 在时刻 t 的密度，文中也略宽泛地说过程 x_t 的“分布”为 $\rho(t)$。

S029 · Original / 原文 · PDF p.9 · S1.SS4.p2.1

We use standard notation for function spaces: for example, $C^1([0,1])$ is the space of continuously differentiable functions from $[0,1]$ to $\mathbb{R}$, $(C^2(\mathbb{R}^d))^d$ is the space of twice continuously differentiable functions from $\mathbb{R}^d$ to $\mathbb{R}^d$, and $C_0^p(\mathbb{R}^d)$ is the space of compactly supported functions from $\mathbb{R}^d$ to $\mathbb{R}$ that are continuously differentiable p times. Given a function $b : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ with value $b(t,x)$ at (t,x), we use $b \in C^1([0,1]; (C^2(\mathbb{R}^d))^d)$ to indicate that b is continuously differentiable in t for all $(t,x) \in [0,1] \times \mathbb{R}^d$ and that $b(t,\cdot)$ is an element of $(C^2(\mathbb{R}^d))^d$ for all $t \in [0,1]$.

函数空间使用标准记号。例如 $C^1([0,1])$ 是从 [0,1] 到 R 的连续可微函数空间；$(C^2(R^d))^d$ 是从 R^d 到 R^d 的二阶连续可微函数空间；$C_0^p(R^d)$ 是具有紧支撑、p 阶连续可微的实值函数空间。对 $b:[0,1]\times R^d\to R^d$，记 $b\in C^1([0,1]; (C^2(R^d))^d)$，表示 b 关于 t 连续可微，且对每个 t，b(t,·) 关于 x 二阶连续可微。

2 Stochastic interpolant framework

2 随机插值框架

2.1 Definitions and assumptions

2.1 定义与假设

S030 · Original / 原文 · PDF p.9 · S2.SS1.p1.1

We begin by defining the stochastic processes that are central to our approach:

先定义本方法的核心随机过程。

Definition 1 (Stochastic interpolant).定义 1（随机插值）

S031 · Original / 原文 · PDF p.9 · Thmtheorem1.p1.1

Given two probability density functions $\rho_0, \rho_1 : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$, a stochastic interpolant between ρ_0 and ρ_1 is a stochastic process x_t defined as

给定概率密度 ρ_0、ρ_1:$\mathbb{R}^d \to \mathbb{R}_{\geq 0}$，二者之间的随机插值，是按下式定义的随机过程 x_t。

$$x_t = I(t, x_0, x_1) + \gamma(t)z, \qquad t \in [0, 1], \tag{2.1}$$

S032 · Original / 原文 · PDF p.9 · Thmtheorem1.p1.2

where:

其中各项满足下列条件。

1. 1.

   S033 · Original / 原文 · PDF p.9 · S2.I1.i1.p1.2

   $I \in C^2([0, 1], C^2(\mathbb{R}^d \times \mathbb{R}^d)^d)$ satisfies the boundary conditions $I(0, x_0, x_1) = x_0$ and $I(1, x_0, x_1) = x_1$, as well as

   $I \in C^2([0,1], C^2(\mathbb{R}^d \times \mathbb{R}^d)^d)$ 满足端点条件 $I(0,x_0,x_1)=x_0$、$I(1,x_0,x_1)=x_1$，以及下列增长约束。

   $$\exists C_1 < \infty : \qquad |\partial_t I(t, x_0, x_1)| \leq C_1 |x_0 - x_1| \qquad \forall (t, x_0, x_1) \in [0, 1] \times \mathbb{R}^d \times \mathbb{R}^d. \tag{2.2}$$

2. 2.

   S034 · Original / 原文 · PDF p.9 · S2.I1.i2.p1.1

   $\gamma : [0, 1] \to \mathbb{R}$ satisfies $\gamma(0) = \gamma(1) = 0$, $\gamma(t) > 0$ for all $t \in (0, 1)$, and $\gamma^2 \in C^2([0, 1])$.

   $\gamma:[0,1]\to\mathbb{R}$ 满足 $\gamma(0)=\gamma(1)=0$；对全部内点 $t\in(0,1)$，$\gamma(t)>0$；并且 $\gamma^2\in C^2([0,1])$。

3. 3.

   S035 · Original / 原文 · PDF p.9 · S2.I1.i3.p1.1

   The pair (x_0, x_1) is drawn from a probability measure ν that marginalizes on ρ_0 and ρ_1, i.e.

   端点对 (x_0,x_1) 从概率测度 ν 采样，其两个边缘分别为 ρ_0 与 ρ_1，即下式。

   $$\nu(dx_0, \mathbb{R}^d) = \rho_0(x_0)dx_0, \qquad \nu(\mathbb{R}^d, dx_1) = \rho_1(x_1)dx_1. \tag{2.3}$$

4. 4.

S036 · Original / 原文 · PDF p.9 · S2.I1.i4.p1.1

z is a Gaussian random variable independent of (x_0, x_1), i.e. $z \sim \mathsf{N}(0, Id)$ and $z \perp (x_0, x_1)$.

z 是独立于 (x_0,x_1) 的高斯变量，即 $z\sim N(0,Id)$、$z\perp(x_0,x_1)$。

S037 · Original / 原文 · PDF p.10 · S2.SS1.p2.1

Eq. (2.2) states that $I(t, x_0, x_1)$ does not move too fast along the way from x_0 at $t = 0$ to x_1 at $t = 1$, and as a result does not wander too far from either endpoint – this assumption is made for convenience but is not necessary for most arguments below. Later, we will find it useful to consider choices for I that are spatially nonlinear, which we show can recover the solution to the Schrödinger bridge problem. Nevertheless, a simple example that serves as a valid I in the sense of Definition 1 is given in (1.1). The measure ν allows for a coupling between the two densities ρ_0 and ρ_1, which affects the properties of the stochastic interpolant, but a simple choice is to take the product measure $\nu(dx_0, dx_1) = \rho_0(x_0)\rho_1(x_1)dx_0dx_1$, in which case x_0 and x_1 are independent. In Section 6 we discuss how to design the stochastic interpolant in (2.1) and state some properties of the corresponding process x_t. Examples of stochastic interpolants are also shown in Figure 2 for various choices of I and γ.

式（2.2）说明 I 从 x_0 走向 x_1 时不会移动过快，从而不会离两端太远；这个假设为方便而设，下文大部分论证并非必需。后文会使用空间非线性的 I，并证明它可以恢复 Schrödinger 桥解。简单的合法实例已见式（1.1）。测度 ν 容许 ρ_0、ρ_1 之间耦合，并影响插值性质；最简单是乘积测度 $\nu(dx_0,dx_1)=\rho_0(x_0)\rho_1(x_1)dx_0dx_1$，此时端点独立。第 6 节讨论设计式（2.1）的方式与性质；图 2 展示不同 I、γ 的例子。

Remark 2 (Comparison with *Albergo and Vanden-Eijnden (2023)*).说明 2（与先前框架比较）

S038 · Original / 原文 · PDF p.10 · Thmtheorem2.p1.1

The main difference between the stochastic interpolant defined in (2.1) and the one originally introduced in *Albergo and Vanden-Eijnden (2023)* is the inclusion of the latent variable $\gamma(t)z$. Many of the results below also hold when we set $\gamma(t)z = 0$, but the objective of the present paper is to elucidate the advantages that this additional term provides when neither of the endpoints are Gaussian. We note that we could generalize the construction by making $\gamma(t)$ a tensor; here we focus on the scalar case for simplicity. Another difference is the possibility to couple ρ_0 and ρ_1 via ν. While the latent variable can be drawn from any noise distribution, as we will see, it will be convenient to choose it to be a Gaussian.

与 Albergo 与 Vanden-Eijnden（2023）相比，主要不同是加入潜变量 γ(t)z。令它为零时，许多下述结果仍成立，但本文着重阐明两端都非高斯时额外噪声项的好处。γ 可推广为张量，此处为简单取标量。另一差别是可通过 ν 耦合两端。潜变量可来自任意噪声分布，不过高斯选择会使分析更方便。

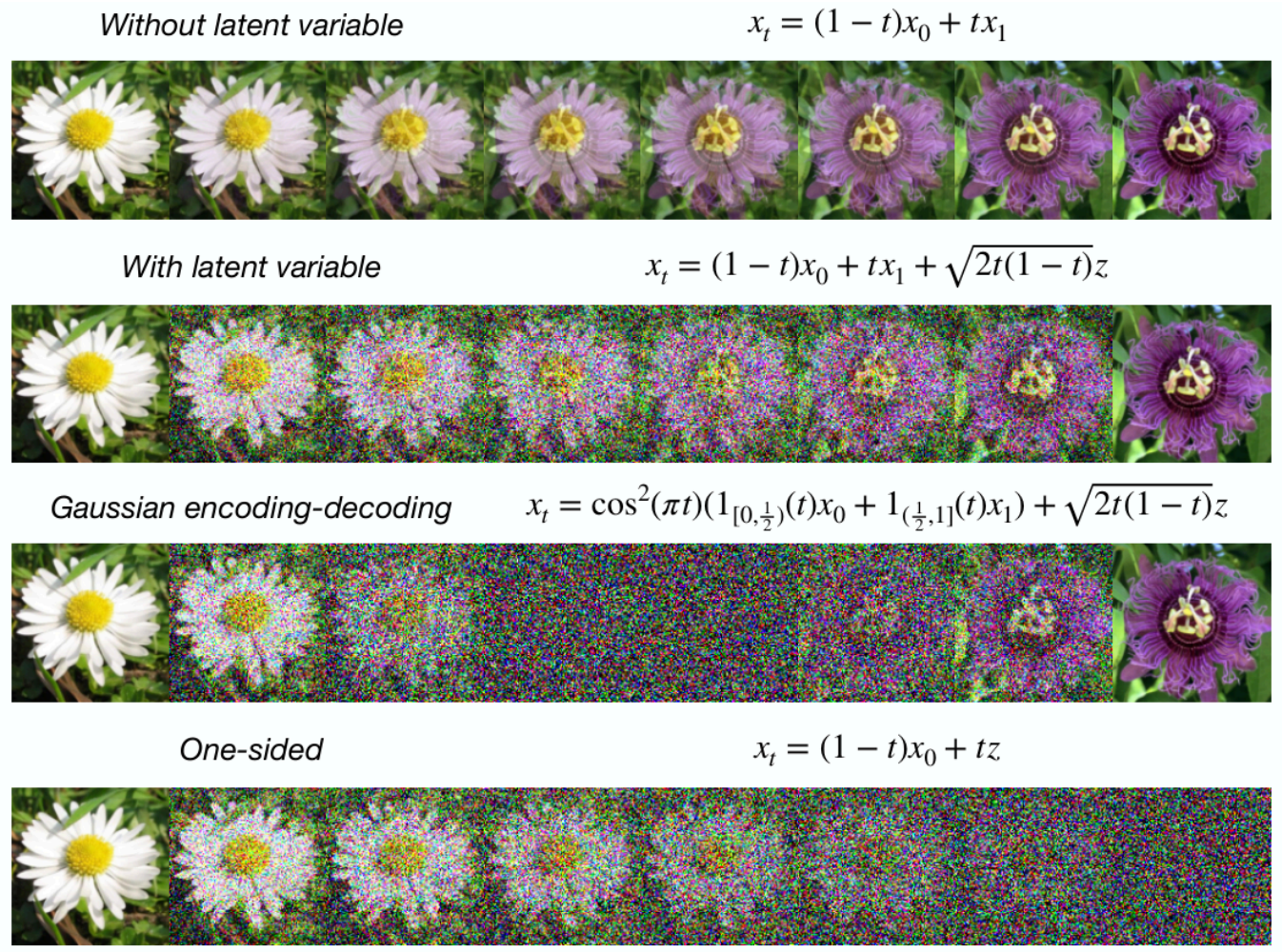

Figure 2: Design flexibility. An illustration of how stochastic interpolants can be tailored to specific aims. All examples show one realization of x_t with one $x_0 \sim \rho_0$, one $x_1 \sim \rho_1$ (the flowers at the left and right of the figures), and one $z \sim \mathsf{N}(0, Id)$. Top, Upper middle, and lower middle: various interpolants, ranging from direct interpolation with no latent variable (as in *Albergo and Vanden-Eijnden (2023)*) to Gaussian encoding-decoding in which the data transitions to pure noise at the mid-point. Bottom: one-sided interpolant, which connects with score-based diffusion methods.

图 2：设计灵活性。各例固定一个 $x_0 \sim \rho_0$、一个 $x_1 \sim \rho_1$（两端花朵）及一个 $z \sim \mathsf{N}(0, Id)$，展示一次 x_t 实现。上面三行从不含潜变量的直接插值，到在中点变成纯噪声的高斯编码—解码。最下一行为与得分扩散相联系的单侧插值。

S039 · Original / 原文 · PDF p.10 · S2.SS1.p3.1

The stochastic interpolant x_t in (2.1) is a continuous-time stochastic process whose realizations are samples from ρ_0 at time $t = 0$ and from ρ_1 at time $t = 1$ by construction. As a result, it offers a way to bridge ρ_0 and ρ_1 – we are interested in characterizing the law of x_t over the full interval $[0, 1]$, as it will allow us to design generative models. Mathematically, we want to characterize the properties of the time-dependent probability distribution $\mu(t, dx)$ such that

按构造，式（2.1）的连续时间过程在 t=0、t=1 分别取样于 ρ_0、ρ_1，因而实现桥接。为构建生成模型，我们关注完整区间内 x_t 的分布。数学上，希望刻画满足下式的时变概率测度 μ(t,dx)。

$$\forall t \in [0,1] \quad : \quad \int_{\mathbb{R}^d} \phi(x)\mu(t,dx) = \mathbb{E}[\phi(x_t)] \quad \text{for any test function} \quad \phi \in C_0^\infty(\mathbb{R}^d), \tag{2.4}$$

S040 · Original / 原文 · PDF p.10 · S2.SS1.p3.2

where x_t is defined in (2.1) and the expectation is taken independently over $(x_0, x_1) \sim \nu$, and $z \sim \mathsf{N}(0, Id)$. To this end, we will need to use conditional expectations over x_t 1 Formally, in terms of the Dirac delta distribution, we can write

$$\mathbb{E}\left[f(t,x_0,x_1,z)|x_t=x\right]=\frac{\mathbb{E}\left[f(t,x_0,x_1,z)\delta(x-x_t)\right]}{\mathbb{E}[\delta(x-x_t)]}$$

and in this notation we also have $\mu(t,x)=\mathbb{E}[\delta(x-x_t)]$., as described in the following definition.

其中 x_t 由式（2.1）定义，期望分别对 $(x_0,x_1)\sim\nu$ 与独立 $z\sim N(0,Id)$ 计算。为此需定义给定 x_t 的条件期望。注 1：形式上可用 Dirac delta 写成 $\mathbb{E}[f(t,x_0,x_1,z)|x_t=x]=\mathbb{E}[f(t,x_0,x_1,z)\delta(x-x_t)]/\mathbb{E}[\delta(x-x_t)]$，且 $\mu(t,x)=E[\delta(x-x_t)]$。

Definition 3.定义 3

S041 · Original / 原文 · PDF p.10 · Thmtheorem3.p1.1

Given any $f:[0,1]\times\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}$, its conditional expectation $\mathbb{E}\left[f(t,x_0,x_1,z)|x_t=x\right]$ is the function of x such that, for any test function $\phi\in C_0^\infty(\mathbb{R}^d)$, we have

给定任意 $f:[0,1]\times R^d\times R^d\times R^d\to R$，其条件期望 $E[f(t,x_0,x_1,z)|x_t=x]$ 是关于 x 的函数，对任意测试函数 $\varphi\in C_0^\infty(R^d)$ 都满足下式。

$$\forall t\in[0,1]\quad:\quad\int_{\mathbb{R}^d}\phi(x)\mathbb{E}\left[f(t,x_0,x_1,z)|x_t=x\right]\mu(t,dx)=\mathbb{E}[\phi(x_t)f(t,x_0,x_1,z)],\tag{2.5}$$

S042 · Original / 原文 · PDF p.11 · Thmtheorem3.p1.2

where $\mu(t,dx)$ is the time-dependent distribution of x_t defined by (3), and the expectation on the right-hand side is taken independently over $(x_0,x_1)\sim\nu$ and $z\sim N(0,Id)$ with x_t given by (2.1).

其中 $\mu(t,dx)$ 是定义 3 所用 x_t 的时变分布；右侧期望分别对 $(x_0,x_1)\sim\nu$ 与独立 $z\sim N(0,Id)$ 计算，x_t 由式（2.1）给定。

S043 · Original / 原文 · PDF p.11 · S2.SS1.p4.1

Vector-valued functions have conditional expectations that are defined analogously. Note that, with our definition, $\mathbb{E}\left[f(t,x_0,x_1,z)|x_t=x\right]$ is a deterministic function of $(t,x)\in[0,1]\times\mathbb{R}^d$, not to be confused with the random variable $\mathbb{E}\left[f(t,x_0,x_1,z)|x_t\right]$ that can be defined analogously.

向量值函数的条件期望类似定义。注意，这里的 $E[f|x_t=x]$ 是 (t,x) 的确定性函数，不应与可类似定义的随机变量 $E[f|x_t]$ 混淆。

Remark 4.说明 4

S044 · Original / 原文 · PDF p.11 · Thmtheorem4.p1.1

Another seemingly more general way to define the stochastic interpolant is via

另一种看似更一般的随机插值定义方式如下。

$$x_t^{\mathsf{d}}=I(t,x_0,x_1)+N_t\tag{2.6}$$

S045 · Original / 原文 · PDF p.11 · Thmtheorem4.p1.2

where $N:[0,1]\to\mathbb{R}^d$ is a zero-mean Gaussian stochastic process constrained to satisfy $N_{t=0}=N_{t=1}=0$. As we will show below, our construction only depends on the single-time properties of N_t, which are completely specified by $\mathbb{E}[N_tN_t^{\mathsf{T}}]$. That is, if we take $\gamma(t)$ in (2.1) such that $\mathbb{E}[N_tN_t^{\mathsf{T}}]=\gamma^2(t)Id$, then the probability distribution of x_t will coincide with that of x_t' defined in (2.6), $x_t\overset{\mathrm{d}}{=}x_t^{\mathsf{d}}$. For example, taking $\gamma(t)=\sqrt{t(1-t)}$ in (2.1) – a choice we will consider below in Sec. 3.4 – is equivalent to choosing N_t to be a Brownian bridge in (2.6), i.e. the stochastic process realizable in terms of the Wiener process W_t as $N_t=W_t-tW_1$. This observation will also help us draw an analogy between our approach and the

construction used in score-based diffusion models as well as methods based on stochastic bridges. As we will show in Sec. 3.4, it is simpler for both analysis and practical implementation to work with the definition (2.1) for x_t.

其中 N 是零均值高斯过程，受约束 $N_0=N_1=0$。下文构造只依赖 N_t 的单时刻性质，而这些性质完全由 $E[N_tN_t^T]$ 指定。因此若选择 γ 使该协方差为 $\gamma^2(t)Id$，则式（2.1）的 x_t 与式（2.6）的过程具有相同单时刻分布。例如 $\gamma(t)=\sqrt{t(1-t)}$ 等价于在式（2.6）中采用 Brownian 桥 $N_t=W_t-tW_1$。这也帮助建立与得分扩散、随机桥的类比。第 3.4 节将说明，分析及实现时用式（2.1）更简单。

S046 · Original / 原文 · PDF p.12 · S2.SS1.p5.1

To proceed, we will make the following assumption on the densities ρ_0, ρ_1, and the interplay between the measure ν to the function I:

接下来对密度 ρ_0、ρ_1，以及测度 ν 与函数 I 的配合作如下假设。

Assumption 5.假设 5

S047 · Original / 原文 · PDF p.12 · Thmtheorem5.p1.1

The densities ρ_0 and ρ_1 are strictly positive elements of $C^2(\mathbb{R}^d)$ and are such that

ρ_0、ρ_1 是 $C^2(R^d)$ 中严格为正的密度，并满足下列 Fisher 信息有限条件。

$$\int_{\mathbb{R}^d} |\nabla \log \rho_0(x)|^2 \rho_0(x) dx < \infty \quad \text{and} \quad \int_{\mathbb{R}^d} |\nabla \log \rho_1(x)|^2 \rho_1(x) dx < \infty. \tag{2.7}$$

S048 · Original / 原文 · PDF p.12 · Thmtheorem5.p1.2

The measure ν and the function I are such that

测度 ν、函数 I 满足下列统一矩约束。

$$\exists M_1, M_2 < \infty \ : \ \mathbb{E}\left[|\partial_t I(t, x_0, x_1)|^4\right] \leq M_1; \quad \mathbb{E}\left[|\partial_t^2 I(t, x_0, x_1)|^2\right] \leq M_2, \quad \forall t \in [0, 1], \tag{2.8}$$

S049 · Original / 原文 · PDF p.12 · Thmtheorem5.p1.3

where the expectation is taken over $(x_0, x_1) \sim \nu$.

期望对 $(x_0,x_1)\sim\nu$ 计算。

S050 · Original / 原文 · PDF p.12 · S2.SS1.p6.1

Note that for the interpolant (1.1), Assumption 5 holds if ρ_0 and ρ_1 both have finite fourth moments.

对式（1.1）的插值，只要 ρ_0、ρ_1 都有有限四阶矩，假设 5 就成立。

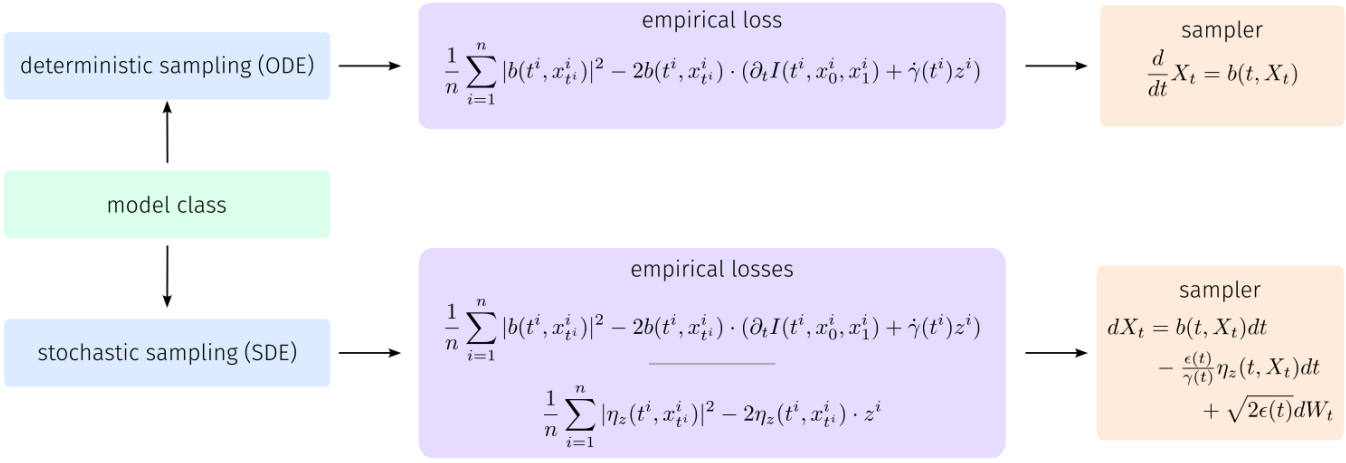

Figure 3: Algorithmic implementation. A simple overview of suggested implementation strategies. For deterministic sampling, a single velocity field b can be learned by minimizing the empirical loss in the top row. For stochastic sampling, the velocity field b, along with the denoiser η_z, can be learned by minimizing the two empirical losses specified in the bottom row. To sample deterministically, off-the-shelf ODE integrators can be used to integrate the probability flow equation. To sample stochastically, the listed SDE can be integrated using standard techniques such as the Euler-Maruyama method or the Heun sampler introduced in *Karras et al. (2022)*. The time-dependent diffusion coefficient $\epsilon(t)$ can be specified after learning to maximize sample quality.

图 3：算法实现概览。确定性采样只需最小化上行经验损失学习速度 b；随机采样则最小化下行两个损失，学习 b 和去噪器 ηz。前者用现成 ODE 积分器；后者用 Euler–Maruyama 或 Karras 等（2022）的 Heun 采样器。学习之后仍可指定时变扩散系数 ε(t)，优化样本质量。

2.2 Transport equations, score, and quadratic objectives

2.2 输运方程、得分与二次目标

S051 · Original / 原文 · PDF p.13 · S2.SS2.p1.1

We now state a result that specifies some important properties of the probability distribution of the stochastic interpolant x_t:

现在给出随机插值 x_t 的概率分布的一些重要性质。

Theorem 6 (Stochastic interpolant properties).定理 6（随机插值的性质）

S052 · Original / 原文 · PDF p.13 · Thmtheorem6.p1.1

The probability distribution of the stochastic interpolant x_t defined in (2.1) is absolutely continuous with respect to the Lebesgue measure at all times $t \in [0,1]$ and its time-dependent density $\rho(t)$ satisfies $\rho(0)=\rho_0$, $\rho(1)=\rho_1$, $\rho \in C^1([0,1];C^p(\mathbb{R}^d))$ for any $p \in \mathbb{N}$, and $\rho(t,x)>0$ for all $(t,x)\in[0,1]\times\mathbb{R}^d$. In addition, ρ solves the transport equation

式（2.1）定义的随机插值在全部 t∈[0,1] 时关于 Lebesgue 测度绝对连续。其时变密度满足 ρ(0)=ρ₀、ρ(1)=ρ₁；对任意自然数 p，ρ∈C¹([0,1];Cᵖ(Rᵈ))；且对全部 (t,x)∈[0,1]×Rᵈ，ρ(t,x)>0。此外，ρ 满足下列输运方程。

$$\partial_t \rho + \nabla\cdot(b\rho)=0, \tag{2.9}$$

S053 · Original / 原文 · PDF p.13 · Thmtheorem6.p1.2

where we defined the velocity

其中速度定义如下。

$$b(t,x)=\mathbb{E}[\dot x_t | x_t=x]=\mathbb{E}[\partial_t I(t,x_0,x_1)+\dot\gamma(t)z|x_t=x]. \tag{2.10}$$

S054 · Original / 原文 · PDF p.13 · Thmtheorem6.p1.3

This velocity is in $C^0([0,1];(C^p(\mathbb{R}^d))^d)$ for any $p \in \mathbb{N}$, and such that

对任意自然数 p，此速度属于 $C^0([0,1];(C^p(\mathbb{R}^d))^d)$，并满足下列二阶可积条件。

$$\forall t \in [0,1] \quad : \quad \int_{\mathbb{R}^d} |b(t,x)|^2 \rho(t,x) dx < \infty. \tag{2.11}$$

S055 · Original / 原文 · PDF p.13 · S2.SS2.p2.1

Note that this theorem means that we can write (2.4) as

此定理意味着，可以将式（2.4）写成下式。

$$\forall t \in [0,1] \quad : \quad \int_{\mathbb{R}^d} \phi(x)\rho(t,x)dx = \mathbb{E}\phi(x_t) \quad \text{for any test function} \quad \phi \in C_0^\infty(\mathbb{R}^d). \tag{2.12}$$

S056 · Original / 原文 · PDF p.13 · S2.SS2.p2.2

The transport equation (2.9) can be solved either forward in time from the initial condition $\rho(0) = \rho_0$, in which case $\rho(1) = \rho_1$, or backward in time from the final condition $\rho(1) = \rho_1$, in which case $\rho(0) = \rho_0$.

输运方程（2.9）既可从 $\rho(0)=\rho_0$ 向前求解、此时得到 $\rho(1)=\rho_1$，也可从终值 $\rho(1)=\rho_1$ 向后求解、此时得到 $\rho(0)=\rho_0$。

S057 · Original / 原文 · PDF p.13 · S2.SS2.p3.1

The proof of Theorem 6 is given in Appendix B.1; it mostly relies on manipulations involving the characteristic function of the stochastic interpolant x_t. The transport equation (2.9) for ρ lead to methods for generative modeling and density estimation, as explained in Secs. 2.3 and 2.5, provided that we can estimate the velocity b. This velocity is explicitly available only in special cases, for example when ρ_0 and ρ_1 are both Gaussian mixture densities: this case is treated in Appendix A. In general b must be calculated numerically, which can be performed via empirical risk minimization of a quadratic objective function, as characterized by our next result:

定理 6 的证明见附录 B.1，主要依赖对随机插值特征函数的操作。只要能估计速度 b，ρ 的输运方程便能导出生成建模、密度估计方法，见第 2.3、2.5 节。b 只有在特殊情形能显式获得，例如两端均为高斯混合密度，见附录 A。一般须数值计算，可通过二次目标的经验风险最小化实现，下一结果对此作出刻画。

Theorem 7 (Objective).定理 7（目标）

S058 · Original / 原文 · PDF p.13 · Thmtheorem7.p1.1

The velocity b defined in (2.10) is the unique minimizer in $C^0([0,1];(C^1(\mathbb{R}^d))^d)$ of the quadratic objective

式（2.10）的速度 b，是函数空间 $C^0([0,1];(C^1(\mathbb{R}^d))^d)$ 中下列二次目标的唯一极小化解。

$$\mathcal{L}_b[\hat{b}] = \int_0^1 \mathbb{E}\left(\tfrac{1}{2}|\hat{b}(t,x_t)|^2 - (\partial_t I(t,x_0,x_1) + \dot{\gamma}(t)z)\cdot\hat{b}(t,x_t)\right)dt \tag{2.13}$$

S059 · Original / 原文 · PDF p.13 · Thmtheorem7.p1.2

where x_t is defined in (2.1) and the expectation is taken independently over $(x_0,x_1) \sim \nu$ and $z \sim \mathsf{N}(0,Id)$.

其中 x_t 由式（2.1）定义，期望分别对 $(x_0,x_1)\sim\nu$ 与独立 $z\sim\mathsf{N}(0,Id)$ 计算。

S060 · Original / 原文 · PDF p.13 · S2.SS2.p4.1

The proof of Theorem 7 is given in Appendix B.1: it relies on the definitions of b in (2.10), as well as the definition of ρ in (2.12) and some elementary properties of the conditional expectation. We discuss how to estimate the objective function (2.13) in practice in Section 6. Interestingly, we also have access to the score of the probability density, as shown by our next result:

定理 7 的证明见附录 B.1，使用式（2.10）对 b 的定义、式（2.12）对 ρ 的定义及条件期望的基本性质。第 6 节讨论实际估计目标（2.13）的方式。我们还可获得概率密度的得分，如下所示。

Theorem 8 (Score).定理 8（得分）

S061 · Original / 原文 · PDF p.14 · Thmtheorem8.p1.1

The score of the probability density ρ specified in Theorem 6 is in $C^1([0,1];(C^p(\mathbb{R}^d))^d)$ for any $p \in \mathbb{N}$ and given by

定理 6 的密度得分对任意自然数 p 属于 $C^1([0,1];(C^p(\mathbb{R}^d))^d)$，并由下式给出。

$$s(t,x) = \nabla \log \rho(t,x) = -\gamma^{-1}(t)\mathbb{E}(z|x_t = x) \quad \forall (t,x) \in (0,1) \times \mathbb{R}^d \tag{2.14}$$

S062 · Original / 原文 · PDF p.14 · Thmtheorem8.p1.2

In addition it satisfies

此外，它满足下列二阶可积条件。

$$\forall t \in [0,1] \quad : \quad \int_{\mathbb{R}^d} |s(t,x)|^2 \rho(t,x) dx < \infty, \tag{2.15}$$

S063 · Original / 原文 · PDF p.14 · Thmtheorem8.p1.3

and is the unique minimizer in $C^1([0,1];(C^1(\mathbb{R}^d))^d)$ of the quadratic objective

并且是在 $C^1([0,1];(C^1(\mathbb{R}^d))^d)$ 中下列二次目标的唯一极小化解。

$$\mathcal{L}_s[\hat{s}] = \int_0^1 \mathbb{E}\left(\tfrac{1}{2}|\hat{s}(t,x_t)|^2 + \gamma^{-1}(t) z \cdot \hat{s}(t,x_t)\right) dt \tag{2.16}$$

S064 · Original / 原文 · PDF p.14 · Thmtheorem8.p1.4

where x_t is defined in (2.1) and the expectation is taken independently over $(x_0, x_1) \sim \nu$ and $z \sim \mathsf{N}(0, Id)$

其中 x_t 由式（2.1）定义，期望分别对 $(x_0,x_1)\sim\nu$ 与独立 $z\sim\mathsf{N}(0,\mathrm{Id})$ 计算。

S065 · Original / 原文 · PDF p.14 · S2.SS2.p5.1

The proof of Theorem 8 is given in Appendix B.1. We stress that the objective function is well defined despite the fact that $\gamma(0) = \gamma(1) = 0$: see Section 6 for more details about how to evaluate this objective in practice.

定理 8 证明见附录 B.1。尽管 $\gamma(0)=\gamma(1)=0$，该目标仍定义良好；第 6 节给出实际计算细节。

Remark 9 (Denoiser).说明 9（去噪器）

S066 · Original / 原文 · PDF p.14 · Thmtheorem9.p1.1

The quantity

以下定义的量：

$$\eta_z(t,x) = \mathbb{E}(z|x_t = x), \tag{2.17}$$

S067 · Original / 原文 · PDF p.14 · Thmtheorem9.p1.2

will be referred as the denoiser, for reasons that will be made clear in Section 5.2. By (2.14), this quantity gives access to the score on $t \in (0,1)$ (where $\gamma(t) > 0$) since, from (2.14),

称为去噪器，理由见第 5.2 节。由式（2.14），在 γ(t)>0 的内点 t∈(0,1)，可通过它获得得分，具体如下。

$$s(t,x) = -\gamma^{-1}(t)\eta_z(t,x). \tag{2.18}$$

S068 · Original / 原文 · PDF p.14 · Thmtheorem9.p1.3

This denoiser is the minimizer of an equivalent expression to (2.16),

这个去噪器极小化下列与式（2.16）等价的目标。

$$\mathcal{L}_{\eta_z}[\hat{\eta}_z] = \int_0^1 \mathbb{E}\left(\tfrac{1}{2}|\hat{\eta}_z(t,x_t)|^2 - z\cdot\hat{\eta}_z(t,x_t)\right) dt. \tag{2.19}$$

S069 · Original / 原文 · PDF p.14 · Thmtheorem9.p1.4

The denoiser η_z is useful for numerical realizations. In particular, the objective in (2.19) is easier to use than the one in (2.16) because it does not contain the factor $\gamma^{-1}(t)$, which needs careful handling as t approaches 0 and 1.

ηz 便于数值实现。尤其是，式（2.19）不含端点附近须谨慎处理的 $\gamma^{-1}(t)$ 因子，因此比式（2.16）更易使用。

S070 · Original / 原文 · PDF p.14 · S2.SS2.p6.1

Having access to the score immediately allows us to rewrite the TE (2.9) as forward and backward Fokker-Planck equations, which we state as:

获得得分后，可立即将输运方程（2.9）改写为前向、后向 Fokker–Planck 方程，表述如下。

Corollary 10 (Fokker-Planck equations).推论 10（Fokker–Planck 方程）

S071 · Original / 原文 · PDF p.14 · Thmtheorem10.p1.1

For any $\epsilon \in C^0([0,1])$ with $\epsilon(t) \geq 0$ for all $t \in [0,1]$, the probability density ρ specified in Theorem 6 satisfies:

对任意 ε∈C⁰([0,1]) 且 ε(t)≥0，定理 6 给定的密度 ρ 满足下列方程。

1. 1.

   S072 · Original / 原文 · PDF p.14 · S2.I2.i1.p1.1

   The forward Fokker-Planck equation

   前向 Fokker–Planck 方程如下。

$$\partial_t\rho + \nabla\cdot(b_F\rho) = \epsilon(t)\Delta\rho, \qquad \rho(0) = \rho_0, \tag{2.20}$$

   S073 · Original / 原文 · PDF p.14 · S2.I2.i1.p1.2

   where we defined the forward drift

   其中前向漂移定义如下。

$$b_F(t,x) = b(t,x) + \epsilon(t)s(t,x). \tag{2.21}$$

S074 · Original / 原文 · PDF p.14 · S2.I2.i1.p1.3

Equation (2.20) is well-posed when solved forward in time from $t = 0$ to $t = 1$, and its solution for the initial condition $\rho(t = 0) = \rho_0$ satisfies $\rho(t = 1) = \rho_1$.

从 t=0 向 t=1 求解式（2.20）是适定的；从初值 ρ(0)=ρ₀ 出发的解满足 ρ(1)=ρ₁。

2. 2.

S075 · Original / 原文 · PDF p.14 · S2.I2.i2.p1.1

The backward Fokker-Planck equation

后向 Fokker–Planck 方程如下。

$$\partial_t \rho + \nabla \cdot (b_{\mathrm{B}} \rho) = -\epsilon(t) \Delta \rho, \qquad \rho(1) = \rho_1, \tag{2.22}$$

S076 · Original / 原文 · PDF p.14 · S2.I2.i2.p1.2

where we defined the backward drift

其中后向漂移定义如下。

$$b_{\mathrm{B}}(t, x) = b(t, x) - \epsilon(t) s(t, x). \tag{2.23}$$

S077 · Original / 原文 · PDF p.15 · S2.I2.i2.p1.3

Equation (2.22) is well-posed when solved backward in time from $t = 1$ to $t = 0$, and its solution for the final condition $\rho(1) = \rho_1$ satisfies $\rho(0) = \rho_0$.

从 t=1 向 t=0 求解式（2.22）是适定的；满足终值 ρ(1)=ρ₁ 的解会得到 ρ(0)=ρ₀。

S078 · Original / 原文 · PDF p.15 · S2.SS2.p7.1

In Section 2.3 we will use the results of this theorem to design generative models based on forward and backward stochastic differential equations. Note that we can replace the diffusion coefficient $\epsilon(t)$ by a positive semi-definite tensor; also note that if we define $\rho_{\mathrm{B}}(t_{\mathrm{B}}, x) = \rho(1 - t_{\mathrm{B}}, x)$, the reversed FPE (2.22) can be written as

第 2.3 节将利用该推论设计前后向 SDE 生成模型。扩散系数 ε(t) 还可换成半正定张量。另若定义 ρB(tB,x)=ρ(1−tB,x)，则逆向 FPE（2.22）可改写如下。

$$\partial_{t_{\mathrm{B}}} \rho_{\mathrm{B}}(t_{\mathrm{B}}, x) - \nabla \cdot (b_{\mathrm{B}}(1 - t_{\mathrm{B}}, x) \rho_{\mathrm{B}}(t_{\mathrm{B}}, x)) = \epsilon(1 - t_{\mathrm{B}}) \Delta \rho_{\mathrm{B}}(t_{\mathrm{B}}, x), \qquad \rho_{\mathrm{B}}(t_{\mathrm{B}} = 0) = \rho_1, \tag{2.24}$$

S079 · Original / 原文 · PDF p.15 · S2.SS2.p7.2

which is now well-posed forward in (reversed) time t_{B}. So as to have only one definition of time t, it is more convenient to work with (2.22).

这样，方程在反转时间 tB 下向前求解便是适定的。为统一时间 t 的定义，下文仍用式（2.22）更方便。

S080 · Original / 原文 · PDF p.15 · S2.SS2.p8.1

Let us make a few remarks about the statements made so far:

对以上结论补充若干说明。

Remark 11.说明 11

S081 · Original / 原文 · PDF p.15 · Thmtheorem11.p1.1

If we set $\gamma(t) = 0$ in x_t (i.e, if we remove the latent variable), the stochastic interpolant (2.1) reduces to the one originally considered in *Albergo and Vanden-Eijnden (2023)*. In this setup, the results above formally stand except that we cannot guarantee the spatial regularity of $b(t, x)$ and $s(t, x)$, since it relies on the presence of the latent variable (as shown in the proof of Theorem 6). Hence, we expect the introduction of the latent variable $\gamma(t)z$ to help for generative modeling, where the solution to the corresponding ODEs/SDEs will be better behaved, and for statistical approximation, since the targets b and s will be more regular. We will see in Section 6 that it also gives us much greater flexibility in the way we can bridge ρ_0 and ρ_1, which will enable us to design generative models with appealing properties.

若令 γ(t)=0、即移除潜变量，式（2.1）退化为 Albergo 与 Vanden-Eijnden（2023）的原始插值。此时上述结果形式上仍成立，但无法保证 b(t,x)、s(t,x) 的空间正则性，因为正则性依赖潜变量，见定理 6 证明。因此预计加入 γ(t)z 有助生成建模：ODE/SDE 解更良性，统计近似的目标 b、s 也更光滑。第 6 节还将说明，额外噪声让连接两密度的设计更灵活，从而可构造具有优良性质的生成模型。

Remark 12.说明 12

S082 · Original / 原文 · PDF p.15 · Thmtheorem12.p1.1

We will see in Section 2.4 that the forward and backward FPE in (2.20) and (2.22) are more robust than the TE in (2.9) against approximation errors in the velocity b and the score s, which has practical implications for generative models based on these equations.

第 2.4 节将说明，相比输运方程（2.9），前后向 FPE（2.20）、（2.22）对速度 b、得分 s 的近似误差更稳健，这影响相应生成模型的实践表现。

Remark 13.说明 13

S083 · Original / 原文 · PDF p.15 · Thmtheorem13.p1.1

We could also obtain $b(t, \cdot)$ at any $t \in [0, 1]$ by minimizing

也可以对每个固定 t∈[0,1]，通过最小化下式求 b(t,·)。

$$\mathbb{E}\left(\tfrac{1}{2}|\hat{b}(t, x_t)|^2 - (\partial_t I(t, x_0, x_1) + \dot{\gamma}(t)z) \cdot \hat{b}(t, x_t)\right) \qquad t \in [0, 1] \tag{2.25}$$

S084 · Original / 原文 · PDF p.15 · Thmtheorem13.p1.2

and $s(t, \cdot)$ at any $t \in (0, 1)$ by minimizing

并对每个固定 t∈(0,1)，通过最小化下式求 s(t,·)。

$$\mathbb{E}\left(\tfrac{1}{2}|\hat{s}(t, x_t)|^2 + \gamma^{-1}(t)z \cdot \hat{s}(t, x_t)\right) \qquad t \in (0, 1) \tag{2.26}$$

S085 · Original / 原文 · PDF p.15 · Thmtheorem13.p1.3

Using the time-integrated versions of these objectives given in (2.13) and (2.16) is more convenient numerically as it allows one to parameterize $\hat{b}$ and $\hat{s}$ globally for $(t, x) \in [0, 1] \times \mathbb{R}^d$.

数值上，使用式（2.13）、（2.16）的时间积分目标更方便，因为可在整个 (t,x)∈[0,1]×R^d 上统一参数化 $\hat{b}$、$\hat{s}$。

Remark 14.说明 14

S086 · Original / 原文 · PDF p.15 · Thmtheorem14.p1.1

From (2.10) we can write

由式（2.10）可将 b 写成下式。

$$b(t,x) = v(t,x) - \dot{\gamma}(t)\gamma(t)s(t,x), \tag{2.27}$$

S087 · Original / 原文 · PDF p.16 · Thmtheorem14.p1.2

where s is the score given in (2.14) and we defined the velocity field

其中 s 是式（2.14）的得分，速度场 v 定义如下。

$$v(t,x) = \mathbb{E}(\partial_t I(t,x_0,x_1)|x_t = x). \tag{2.28}$$

S088 · Original / 原文 · PDF p.16 · Thmtheorem14.p1.3

The velocity field $v \in C^0([0,1];(C^p(\mathbb{R}^d))^d)$ for any $p \in \mathbb{N}$ and can be characterized as the unique minimizer of

对任意自然数 p，$v \in C^0([0,1];(C^p(\mathbb{R}^d))^d)$，并可刻画为下列目标的唯一极小化解。

$$\mathcal{L}_v[\hat{v}] = \int_0^1 \mathbb{E}\left(\tfrac{1}{2}|\hat{v}(t,x_t)|^2 - \partial_t I(t,x_0,x_1)\cdot\hat{v}(t,x_t)\right) dt \tag{2.29}$$

S089 · Original / 原文 · PDF p.16 · Thmtheorem14.p1.4

Learning this velocity and the score separately may be useful in practice.

分别学习这个速度与得分，在实践中可能有益。

Remark 15.说明 15

S090 · Original / 原文 · PDF p.16 · Thmtheorem15.p1.1

The objectives in (2.13) and (2.16) (as well as the ones in(2.19) and (2.29)) are amenable to empirical estimation if we have samples $(x_0,x_1) \sim \nu$, since in that case we can generate samples of $x_t = I(t,x_0,x_1) + \gamma(t)z$ at any time $t \in [0,1]$. We will use this feature in the numerical experiments presented below.

若能取得 $(x_0,x_1)\sim\nu$ 样本，就可经验估计式（2.13）、（2.16）、（2.19）、（2.29）的目标：对任意 $t\in[0,1]$ 都能生成 $x_t=I(t,x_0,x_1)+\gamma(t)z$。后文数值实验将利用此性质。

Remark 16.说明 16

S091 · Original / 原文 · PDF p.16 · Thmtheorem16.p1.1

Since s is the score of ρ, an alternative objective to estimate it is *(Hyvärinen, 2005)*

由于 s 是 ρ 的得分，还可用 Hyvärinen（2005）的下列目标估计。

$$\int_0^1 \mathbb{E}\left(|\hat{s}(t,x_t)|^2 + 2\nabla\cdot\hat{s}(t,x_t)\right) dt. \tag{2.30}$$

S092 · Original / 原文 · PDF p.16 · Thmtheorem16.p1.2

The derivation of (2.30) is standard: for the reader's convenience we recall it at the end of Appendix B.1. The advantage of using (2.16) over (2.30) is that it does not require us to take the divergence of $\hat{s}$.

式（2.30）的推导是标准结论，为方便读者，附录 B.1 末尾予以回顾。式（2.16）的优势是不需计算 ŝ 的散度。

Remark 17 (Energy-based models).说明 17（能量模型）

S093 · Original / 原文 · PDF p.16 · Thmtheorem17.p1.1

By definition, the score $s(t,x) = \nabla \log \rho(t,x)$ is a gradient field. As a result, if we model $\hat{s}(t,x) = -\nabla \hat{E}(t,x)$, we can turn (2.16) into an objective function for $\hat{E}(t,x)$

按定义，s(t,x)=∇logρ(t,x) 是梯度场。因此若设 ŝ(t,x)=−∇Ê(t,x)，便可将式（2.16）转为下列关于 Ê 的目标。

$$\mathcal{L}_E[\hat{E}] = \int_0^1 \mathbb{E}\left(\tfrac{1}{2}|\nabla \hat{E}(t,x_t)|^2 + \gamma^{-1}(t) z \cdot \nabla \hat{E}(t,x_t)\right) dt \tag{2.31}$$

S094 · Original / 原文 · PDF p.16 · Thmtheorem17.p1.2

This objective is invariant to constant shifts in $\hat{E}$ and should therefore be minimized under some constraint, such as $\min_x \hat{E}(t,x) = 0$ for all $t \in [0,1]$. The minimizer of (2.31) provides us with an energy-based model (EBM) *(LeCun et al., 2007; Song and Kingma, 2021)* that can in principle be used to sample the PDF of the stochastic interpolant, $\rho(t,x)$, at any fixed $t \in [0,1]$ using e.g. Langevin dynamics. We will not exploit this possibility here, and instead rely on generative models to sample $\rho(t,x)$, as discussed next in Sec. 2.3.

目标对 Ê 加常数不变，因此应在某种约束下极小化，例如对任意 t∈[0,1] 令 $\min_x$Ê(t,x)=0。式（2.31）的极小化解给出能量模型 EBM（LeCun 等，2007；Song 与 Kingma，2021），原则上可在固定 t 用 Langevin 动力学等方法采样插值密度 ρ(t,x)。本文不继续利用此途径，而采用第 2.3 节的生成模型采样。

2.3 Generative models

2.3 生成模型

S095 · Original / 原文 · PDF p.16 · S2.SS3.p1.1

Our next result is a direct consequence of Theorem 6, and it shows how to design generative models using the stochastic processes associated with the TE (2.9), the forward FPE (2.20), and the backward FPE (2.22):

下一结论直接来自定理 6，说明如何使用与 TE（2.9）、前向 FPE（2.20）、后向 FPE（2.22）相联系的随机过程来设计生成模型。

Corollary 18 (Generative models).推论 18（生成模型）

S096 · Original / 原文 · PDF p.17 · Thmtheorem18.p1.1

At any time $t \in [0,1]$, the law of the stochastic interpolant x_t coincides with the law of the three processes X_t, X_t^{F}, and X_t^{B}, respectively defined as:

在任意 t∈[0,1]，随机插值 x_t 的分布分别与如下三个过程 X_t、X_t^F、X_t^B 的分布相同。

1. 1.

S097 · Original / 原文 · PDF p.17 · S2.I3.i1.p1.1

The solutions of the probability flow associated with the transport equation (2.9)

与输运方程（2.9）相联系的概率流解如下。

$$\frac{d}{dt}X_t = b(t, X_t), \tag{2.32}$$

S098 · Original / 原文 · PDF p.17 · S2.I3.i1.p1.2

solved either forward in time from the initial data $X_{t=0} \sim \rho_0$ or backward in time from the final data $X_{t=1} = x_1 \sim \rho_1$.

它既可从初值 $X_0 \sim \rho_0$ 向前求解，也可从终值 $X_1 = x_1 \sim \rho_1$ 向后求解。

2. 2.

S099 · Original / 原文 · PDF p.17 · S2.I3.i2.p1.1

The solutions of the forward SDE associated with the FPE (2.20)

与 FPE（2.20）相联系的前向 SDE 解如下。

$$dX_t^{\mathsf{F}} = b_{\mathsf{F}}(t, X_t^{\mathsf{F}})dt + \sqrt{2\epsilon(t)}\, dW_t, \tag{2.33}$$

S100 · Original / 原文 · PDF p.17 · S2.I3.i2.p1.2

solved forward in time from the initial data $X_{t=0}^{\mathsf{F}} \sim \rho_0$ independent of W.

从独立于 W 的初值 $X_0^{\mathsf{F}} \sim \rho_0$ 向前求解。

3. 3.

S101 · Original / 原文 · PDF p.17 · S2.I3.i3.p1.1

The solutions of the backward SDE associated with the backward FPE (2.22)

与后向 FPE（2.22）相联系的后向 SDE 解如下。

$$dX_t^{\mathsf{B}} = b_{\mathsf{B}}(t, X_t^{\mathsf{B}})dt + \sqrt{2\epsilon(t)}\, dW_t^{\mathsf{B}}, \quad W_t^{\mathsf{B}} = -W_{1-t}, \tag{2.34}$$

S102 · Original / 原文 · PDF p.17 · S2.I3.i3.p1.2

solved backward in time from the final data $X_{t=1}^{\mathsf{B}} \sim \rho_1$ independent of W^{B}; the solution of (2.34) is by definition $X_t^{\mathsf{B}} = Z_{1-t}^{\mathsf{F}}$ where Z_t^{F} satisfies

它从独立于 WB 的终值 $X_1^{\mathsf{B}} \sim \rho_1$ 向后求解。按定义，式（2.34）的解为 $X_t^{\mathsf{B}} = Z_{1-t}^{\mathsf{F}}$，其中 Z_t^{F} 满足下式。

$$dZ_t^{\mathsf{F}} = -b_{\mathsf{B}}(1-t, Z_t^{\mathsf{F}})dt + \sqrt{2\epsilon(t)}\, dW_t, \tag{2.35}$$

S103 · Original / 原文 · PDF p.17 · S2.I3.i3.p1.3

solved forward in time from the initial data $Z_{t=0}^{\mathsf{F}} \sim \rho_1$ independent of W.

该方程从独立于 W 的初值 $Z_0^{\mathsf{F}} \sim \rho_1$ 向前求解。

S104 · Original / 原文 · PDF p.17 · S2.SS3.p2.1

To avoid repeated applications of the transformation $t \mapsto 1-t$, it is convenient to work with (2.34) directly using the reversed Itô calculus rules stated in the following lemma, which follows from the results in *Anderson (1982)* and is proven in

Appendix B.2:

为避免反复使用 t↦1−t 的变换，可以直接用下列引理的逆向 Itô 规则处理式（2.34）。引理来自 Anderson（1982），证明见附录 B.2。

Lemma 19 (Reverse Itô Calculus).引理 19（逆向 Itô 演算）

S105 · Original / 原文 · PDF p.17 · Thmtheorem19.p1.1

If X_t^{B} solves the backward SDE (2.34):

若 X_t^{B} 满足后向 SDE（2.34），则有下列性质。

1. 1.

   S106 · Original / 原文 · PDF p.17 · S2.I4.i1.p1.1

   For any $f \in C^1([0,1]; C_0^2(\mathbb{R}_d))$ and $t \in [0,1]$, the backward Itô formula holds

   对任意 $f \in C^1([0,1]; C_0^2(\mathbb{R}^d))$、$t \in [0,1]$，后向 Itô 公式如下。

$$df(t, X_t^{\mathsf{B}}) = \partial_t f(t, X_t^{\mathsf{B}})dt + \nabla f(X_t^{\mathsf{B}}) \cdot dX_t^{\mathsf{B}} - \epsilon(t)\Delta f(t, X_t^{\mathsf{B}})dt. \tag{2.36}$$

2. 2.

   S107 · Original / 原文 · PDF p.17 · S2.I4.i2.p1.2

   For any $g \in C^0([0,1]; (C_0(\mathbb{R}_d))^d)$ and $t \in [0,1]$, the backward Itô isometries hold:

   对任意 $g \in C^0([0,1]; (C_0(\mathbb{R}^d))^d)$、$t \in [0,1]$，后向 Itô 等距性质如下。

$$\mathbb{E}_{\mathsf{B}}^x \int_t^1 g(t, X_t^{\mathsf{B}}) \cdot dW_t^{\mathsf{B}} = 0; \qquad \mathbb{E}_{\mathsf{B}}^x \left| \int_t^1 g(t, X_t^{\mathsf{B}}) \cdot dW_t^{\mathsf{B}} \right|^2 = \int_t^1 \mathbb{E}_{\mathsf{B}}^x \left| g(t, X_t^{\mathsf{B}}) \right|^2 dt, \tag{2.37}$$

   S108 · Original / 原文 · PDF p.17 · S2.I4.i2.p1.3

   where $\mathbb{E}_{\mathsf{B}}^x$ denotes expectation conditioned on the event $X_{t=1}^{\mathsf{B}} = x$.

   其中 $\mathbb{E}_{\mathsf{B}}^x$ 表示在 $X_1^{\mathsf{B}} = x$ 的条件下计算的期望。

S109 · Original / 原文 · PDF p.17 · S2.SS3.p3.1

The relevance of Corollary 18 for generative modeling is clear. Assuming, for example, that ρ_0 is a simple density that can be sampled easily (e.g. a Gaussian or a Gaussian mixture density), we can use the ODE (2.32) or the SDE (2.33) to push these samples forward in time and generate samples from a complex target density ρ_1. In Section 2.5, we will show how to use the ODE (2.32) or the reverse SDE (2.34) to estimate ρ_1 at any $x \in \mathbb{R}^d$ assuming that we can evaluate ρ_0 at any $x \in \mathbb{R}^d$. We will also show how similar ideas can be used to estimate the cross entropy between ρ_0 and ρ_1.

推论 18 在生成建模中的意义直接可见。例如，若 ρ_0 是高斯或高斯混合等容易采样的简单密度，就可用 ODE（2.32）或 SDE（2.33）向前推动这些样本，生成复杂目标 ρ_1 的样本。第 2.5 节还将说明：只要能在任意 $x \in \mathbb{R}^d$ 计算 ρ_0，便可用 ODE（2.32）或逆向 SDE（2.34）估计任意位置的 ρ_1，并可用类似思想估计二者的交叉熵。

Remark 20.说明 20

S110 · Original / 原文 · PDF p.18 · Thmtheorem20.p1.1

We stress that the stochastic interpolant x_t, the solution X_t to the ODE (2.32), and the solutions X_t^{F} and X_t^{B} of the forward and backward SDEs (2.33) and (2.34) are different stochastic processes, but their laws all coincide with $\rho(t)$ at any time $t \in [0,1]$. This is all that matters when applying these processes as generative models. However, the fact that

these processes are different has implications for the accuracy of the numerical integration used to sample from them at any t as well as for the propagation of statistical errors (see also the next remark).

强调：插值 x_t、ODE 解 X_t、前后向 SDE 解 X_t^F、X_t^B 是不同的随机过程，但在每个时刻都具有相同密度 $\rho(t)$。作为生成模型时，只需这一点。但过程不同会影响任意时刻采样所用数值积分的精度，也影响统计误差传播，见下一说明。

S111 · Original / 原文 · PDF p.18 · S2.SS3.p4.1

Generative models based on solutions X_t to the ODE (2.32), solutions X_t^F to the forward SDE (2.33), and solutions X_t^B to the backward SDE (2.34) will typically involve drifts b, b_F, and b_B that are, in practice, imperfectly estimated via minimization of (2.13) and (2.16) over finite datasets. It is important to estimate how this statistical estimation error propagates to errors in sample quality, and how the propagation of error depends on the generative model used, which is the object of our next section.

实践中，ODE（2.32）、前向 SDE（2.33）、后向 SDE（2.34）的漂移 b、bF、bB，都通过有限数据上最小化式（2.13）、（2.16）估计，难免不完美。必须研究统计估计误差如何传播为样本质量误差，以及这种传播如何依赖所选生成模型，下一节对此分析。

2.4 Likelihood control

2.4 似然控制

S112 · Original / 原文 · PDF p.18 · S2.SS4.p1.1

In this section, we demonstrate that jointly minimizing the objective functions (2.29) and (2.16) (or the losses (2.13) and (2.16)) controls the KL-divergence from the target density ρ_1 to the model density $\hat{\rho}_1$. We focus on bounds involving the score, but we note that analogous results hold for learning the denoiser $\eta_z(t,x)$ defined in (2.17) by the relation $\eta_z(t,x) = -s(t,x)/\gamma(t)$. The derivation is based on a simple and exact characterization of the KL-divergence between two transport equations or two Fokker-Planck equations with different drifts. Remarkably, we find that the presence of a diffusive term determines whether or not it is sufficient to learn the drift to control KL. This can be seen as a generalization of the result for score-based diffusion models described in *Song et al. (2021a)* to arbitrary generative models described by ODEs or SDEs. The proofs of the statements in this section are provided in Appendix B.3. These proofs rely on manipulation of the time derivative of the KL divergence, which is a practice that has proven useful elsewhere in the literature *(Vempala and Wibisono, 2019)*.

本节证明，联合最小化式（2.29）与（2.16），或式（2.13）与（2.16），可控制从目标密度 ρ_1 到模型密度 $\hat{\rho}_1$ 的 KL 散度。重点给出涉及得分的界；学习式（2.17）的去噪器 η_z 也有类似结果。原文此处写 $\eta_z=-s/\gamma$；译注：按式（2.18）$s=-\eta_z/\gamma$，应为 $\eta_z=-\gamma s$。推导基于对不同漂移的两个输运方程或 FPE 之间 KL 的简单精确刻画。扩散项的存在，决定了仅学习漂移是否足以控制 KL，这将 Song 等（2021a）关于得分扩散的结果推广到一般 ODE/SDE 生成模型。证明见附录 B.3，使用 KL 时间导数的操作；此做法也见 Vempala 与 Wibisono（2019）。

S113 · Original / 原文 · PDF p.18 · S2.SS4.p2.1

We first characterize the KL divergence between two densities transported by two different continuity equations but initialized from the same initial condition:

先刻画从相同初值出发、由两个不同连续性方程输运的密度之间的 KL 散度。

Lemma 21.引理 21

S114 · Original / 原文 · PDF p.18 · Thmtheorem21.p1.2

Let $\rho_0 : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ denote a fixed base probability density function. Given two velocity fields $b, \hat{b} \in C^0([0,1], (C^1(\mathbb{R}^d))^d)$, let the time-dependent densities $\rho : [0,1] \times \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ and $\hat{\rho} : [0,1] \times \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ denote the solutions to the transport equations

固定基密度 $\rho_0:\mathbb{R}^d\to\mathbb{R}_{\geq 0}$。给定 b、$\hat{b}\in C^0([0,1],(C^1(\mathbb{R}^d))^d)$，设时变密度 ρ、$\hat{\rho}:[0,1]\times\mathbb{R}^d\to\mathbb{R}_{\geq 0}$ 分别解下列输运方程。

$$\begin{aligned} \partial_t \rho + \nabla \cdot (b\rho) &= 0, & \rho(0) &= \rho_0, \\ \partial_t \hat{\rho} + \nabla \cdot (\hat{b}\hat{\rho}) &= 0, & \hat{\rho}(0) &= \rho_0. \end{aligned} \tag{2.38}$$

S115 · Original / 原文 · PDF p.18 · Thmtheorem21.p1.3

Then, the Kullback-Leibler divergence of $\rho(1)$ from $\hat{\rho}(1)$ is given by

则从 $\rho(1)$ 到 $\hat{\rho}(1)$ 的 Kullback–Leibler 散度如下。

$$\mathsf{KL}(\rho(1) \,\|\, \hat{\rho}(1)) = \int_0^1 \int_{\mathbb{R}^d} (\nabla \log \hat{\rho}(t,x) - \nabla \log \rho(t,x)) \cdot \big(\hat{b}(t,x) - b(t,x)\big) \rho(t,x) dx dt. \tag{2.39}$$

S116 · Original / 原文 · PDF p.18 · S2.SS4.p3.1

Lemma 21 shows that it is insufficient in general to match $\hat{b}$ with b to obtain control on the KL divergence. The essence of the problem is that a small error in $\hat{b} - b$ does not ensure control on the Fisher divergence $\mathsf{FI}(\rho(t) \,\|\, \hat{\rho}(t)) = \int_{\mathbb{R}^d} |\nabla \log \rho(t,x) - \nabla \log \hat{\rho}(t,x)|^2 \, \rho(t,x) dx$, which is necessary due to the presence of $(\nabla \log \hat{\rho} - \nabla \log \rho)$ in (2.39).

引理 21 表明，一般仅使 $\hat{b}$ 接近 b 还不足以控制 KL。根本原因是 $\hat{b}-b$ 小，不保证 Fisher 散度 $\mathsf{FI}(\rho(t)\|\hat{\rho}(t))$ 小；该散度为 $\int |\nabla \log \rho - \nabla \log \hat{\rho}|^2 \rho \, dx$。式（2.39）包含两密度对数梯度之差，因此需要这种控制。

S117 · Original / 原文 · PDF p.19 · S2.SS4.p4.1

In the next lemma, we study the case for two Fokker-Planck equations, and highlight that the situation becomes quite different.

下一个引理研究两个 FPE，情形将明显不同。

Lemma 22.引理 22

S118 · Original / 原文 · PDF p.19 · Thmtheorem22.p1.2

Let $\rho_0 : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ denote a fixed base probability density function. Given two velocity fields $b_{\mathsf{F}}, \hat{b}_{\mathsf{F}} \in C^0([0,1], (C^1(\mathbb{R}^d))^d)$, let the time-dependent densities $\rho : [0,1] \times \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ and $\hat{\rho} : [0,1] \times \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ denote the solutions to the Fokker-Planck equations

固定基密度 $\rho_0:\mathbb{R}^d\to\mathbb{R}_{\geq 0}$。给定 b_{F}、$\hat{b}_{\mathsf{F}}\in C^0([0,1],(C^1(\mathbb{R}^d))^d)$，设 ρ、$\hat{\rho}$ 是下列 Fokker–Planck 方程的时变密度解。

$$\begin{aligned} \partial_t \rho + \nabla \cdot (b_{\mathsf{F}}\rho) &= \epsilon \Delta \rho, & \rho(0) &= \rho_0, \\ \partial_t \hat{\rho} + \nabla \cdot (\hat{b}_{\mathsf{F}}\hat{\rho}) &= \epsilon \Delta \hat{\rho}, & \hat{\rho}(0) &= \rho_0. \end{aligned} \tag{2.40}$$

S119 · Original / 原文 · PDF p.19 · Thmtheorem22.p1.3

where $\epsilon > 0$. Then, the Kullback-Leibler divergence from $\rho(1)$ to $\hat{\rho}(1)$ is given by

其中 $\varepsilon>0$，则从 $\rho(1)$ 到 $\hat{\rho}(1)$ 的 KL 散度如下。

$$\begin{aligned}\mathsf{KL}(\rho(1)\,\|\,\hat{\rho}(1)) = &\int_0^1 \int_{\mathbb{R}^d} (\nabla \log \hat{\rho}(t,x) - \nabla \log \rho(t,x)) \cdot \left(\hat{b}_{\mathsf{F}}(t,x) - b_{\mathsf{F}}(t,x)\right) \rho(t,x) dx dt \\ &-\epsilon \int_0^1 \int_{\mathbb{R}^d} |\nabla \log \rho(t,x) - \nabla \log \hat{\rho}(t,x)|^2 \rho(t,x) dx dt,\end{aligned} \tag{2.41}$$

S120 · Original / 原文 · PDF p.19 · Thmtheorem22.p1.4

and as a result

因此得到下列上界。

$$\mathsf{KL}(\rho(1)\,\|\,\hat{\rho}(1)) \leq \frac{1}{4\epsilon} \int_0^1 \int_{\mathbb{R}^d} \left|\hat{b}_{\mathsf{F}}(t,x) - b_{\mathsf{F}}(t,x)\right|^2 \rho(t,x) dx dt. \tag{2.42}$$

S121 · Original / 原文 · PDF p.19 · S2.SS4.p5.1

Lemma 22 shows that, unlike for transport equations, the KL-divergence between the solutions of two Fokker-Planck equations is controlled by the error in their drifts. The diffusive term in each Fokker-Planck equation provides an additional negative term in the KL-divergence, which eliminates the need for explicit control on the Fisher divergence.

引理 22 表明，与输运方程不同，两个 FPE 解之间的 KL 受漂移误差控制。每个 FPE 的扩散项在 KL 导数中提供一个额外负项，从而无需显式控制 Fisher 散度。

S122 · Original / 原文 · PDF p.19 · S2.SS4.p6.1

Putting the above results together, we can state the following result, which demonstrates that the losses (2.13) and (2.16) control the likelihood for learned approximations to the FPE (2.20).

综合可得下列结果：损失（2.13）、（2.16）控制 FPE（2.20）学习近似的似然。

Theorem 23.定理 23

S123 · Original / 原文 · PDF p.19 · Thmtheorem23.p1.1

Let ρ denote the solution of the Fokker-Planck equation (2.20) with $\epsilon(t) = \epsilon > 0$. Given two velocity fields $\hat{b}, \hat{s} \in C^0([0,1], (C^1(\mathbb{R}^d))^d)$, define

设 ρ 是 ε(t)=ε>0 时 FPE（2.20）的解。给定 $\hat{b}$、$\hat{s}\in C^0([0,1],(C^1(\mathbb{R}^d))^d)$，定义下式。

$$\hat{b}_{\mathsf{F}}(t,x) = \hat{b}(t,x) + \epsilon \hat{s}(t,x), \qquad \hat{v}(t,x) = \hat{b}(t,x) + \gamma(t)\dot{\gamma}(t)\hat{s}(t,x) \tag{2.43}$$

S124 · Original / 原文 · PDF p.19 · Thmtheorem23.p1.2

where the function γ satisfies the properties listed in Definition 1. Let $\hat{\rho}$ denote the solution to the Fokker-Planck equation

其中 γ 满足定义 1。令 $\hat{\rho}$ 解下列 FPE。

$$\partial_t \hat{\rho} + \nabla \cdot (\hat{b}_{\mathsf{F}} \hat{\rho}) = \epsilon \Delta \hat{\rho}, \qquad \hat{\rho}(0) = \rho_0. \tag{2.44}$$

S125 · Original / 原文 · PDF p.19 · Thmtheorem23.p1.3

Then,

则有下列结论。

$$\mathsf{KL}(\rho_1\,\|\,\hat{\rho}(1)) \leq \frac{1}{2\epsilon}\left(\mathcal{L}_b[\hat{b}] - \min_{\hat{b}} \mathcal{L}_b[\hat{b}]\right) + \frac{\epsilon}{2}\left(\mathcal{L}_s[\hat{s}] - \min_{\hat{s}} \mathcal{L}_s[\hat{s}]\right), \tag{2.45}$$

S126 · Original / 原文 · PDF p.19 · Thmtheorem23.p1.4

where $\mathcal{L}_b[\hat{b}]$ and $\mathcal{L}_s[\hat{s}]$ are the objective functions defined in (2.13) and (2.16), and

其中 Lb[b̂]、Ls[ŝ] 分别为式（2.13）、（2.16）的目标，并有下列另一上界。

$$\mathsf{KL}(\rho_1 \,\|\, \hat{\rho}(1)) \leq \frac{1}{2\epsilon}\left(\mathcal{L}_v[\hat{v}] - \min_{\hat{v}} \mathcal{L}_v[\hat{v}]\right) + \frac{\sup_{t\in[0,1]}(\gamma(t)\dot{\gamma}(t)-\epsilon)^2}{2\epsilon}\left(\mathcal{L}_s[\hat{s}] - \min_{\hat{v}} \mathcal{L}_s[\hat{s}]\right). \tag{2.46}$$

S127 · Original / 原文 · PDF p.19 · Thmtheorem23.p1.5

where $\mathcal{L}_v[\hat{v}]$ is the objective function defined in (2.29).

其中 Lv[v̂] 为式（2.29）的目标。

Remark 24 (Generative modeling).说明 24（生成建模）

S128 · Original / 原文 · PDF p.20 · Thmtheorem24.p1.1

The above results have practical ramifications for generative modeling. In particular, they show that minimizing either the losses (2.13) and (2.16) or (2.29) and (2.16) maximize the likelihood of the stochastic generative model

上述结果对生成建模有直接意义：最小化式（2.13）与（2.16），或式（2.29）与（2.16），能最大化下列随机生成模型的似然。

$$d\hat{X}_t^{\mathsf{F}} = \left(\hat{b}(t, \hat{X}_t^{\mathsf{F}}) + \epsilon \hat{s}(t, \hat{X}_t^{\mathsf{F}})\right) dt + \sqrt{2\epsilon} dW_t, \tag{2.47}$$

S129 · Original / 原文 · PDF p.20 · Thmtheorem24.p1.2

but that minimizing the objective (2.13) is insufficient in general to maximize the likelihood of the deterministic generative model

但仅最小化式（2.13），一般不足以最大化下列确定性生成模型的似然。

$$\dot{\hat{X}}_t = \hat{b}(t, \hat{X}_t). \tag{2.48}$$

S130 · Original / 原文 · PDF p.20 · Thmtheorem24.p1.3

Moreover, they show that, when learning $\hat{b}$ and $\hat{s}$, the choice of ϵ that minimizes the upper bound is given by

此外，学习 b̂、ŝ 时，使似然上界最小的 ε 由下式给出。

$$\epsilon^* = \left(\frac{\mathcal{L}_b[\hat{b}] - \min_{\hat{b}} \mathcal{L}_b[\hat{b}]}{\mathcal{L}_s[\hat{s}] - \min_{\hat{s}} \mathcal{L}_s[\hat{s}]}\right)^{1/2}, \tag{2.49}$$

S131 · Original / 原文 · PDF p.20 · Thmtheorem24.p1.4

so that $\epsilon^* > 1$ if the score is learned to higher accuracy than $\hat{b}$ and $\epsilon^* < 1$ in the opposite situation. Note that (2.49) suggests to take $\epsilon = 0$ if $\hat{b}$ is learned perfectly but $\hat{s}$ is not, and send $\epsilon \to \infty$ in the opposite situation. While taking $\epsilon = 0$ is achievable in practice and leads to the ODE (2.32), taking $\epsilon \to \infty$ is not, as increasing ϵ increases the expense of the numerical integration in (2.33) and (2.34).

若得分比 b̂ 学得更准，则 ε*>1；反之 ε*<1。式（2.49）意味着，若 b̂ 完美但 ŝ 不完美，应取 ε=0；反过来则应令 ε→∞。取零可在实践中实现，对应 ODE（2.32）；趋向无限却不实际，因为增大 ε 会增加式（2.33）、（2.34）的数值积分成本。

2.5 Density estimation and cross-entropy calculation

2.5 密度估计与交叉熵计算

S132 · Original / 原文 · PDF p.20 · S2.SS5.p1.1

It is well-known that the solution of the TE (2.9) can be expressed in terms of the solution to the probability flow ODE (2.32); for completeness, we now recall this fact:

众所周知，TE（2.9）的解可用概率流 ODE（2.32）的解表示。为完整起见，回顾如下。

Lemma 25.引理 25

S133 · Original / 原文 · PDF p.20 · Thmtheorem25.p1.1

Given the velocity field $\hat{b} \in C^0([0,1],(C^1(\mathbb{R}^d))^d)$, let $\hat{\rho}$ satisfy the transport equation

给定 $\hat{b}\in C^0([0,1],(C^1(\mathbb{R}^d))^d)$，令 $\hat{\rho}$ 满足下列输运方程。

$$\partial_t \hat{\rho} + \nabla \cdot (\hat{b}\hat{\rho}) = 0, \tag{2.50}$$

S134 · Original / 原文 · PDF p.20 · Thmtheorem25.p1.2

and let $X_{s,t}(x)$ solve the ODE

令 $X_{s,t}(x)$ 满足下列 ODE。

$$\frac{d}{dt}X_{s,t}(x) = b(t, X_{s,t}(x)), \qquad X_{s,s}(x) = x, \qquad t,s \in [0,1] \tag{2.51}$$

S135 · Original / 原文 · PDF p.20 · Thmtheorem25.p1.3

Then, given the PDFs ρ_0 and ρ_1:

则对给定概率密度 ρ_0、ρ_1，有下列结果。

1. 1.

   S136 · Original / 原文 · PDF p.20 · S2.I5.i1.p1.1

   The solution to (2.50) for the initial condition $\hat{\rho}(0) = \rho_0$ is given at any time $t \in [0,1]$ by

   对初值 $\hat{\rho}(0)=\rho_0$，式（2.50）在任意 $t\in[0,1]$ 的解如下。

$$\hat{\rho}(t,x) = \exp\left(-\int_0^t \nabla \cdot b(\tau, X_{t,\tau}(x))d\tau\right)\rho_0(X_{t,0}(x)) \tag{2.52}$$

2. 2.

   S137 · Original / 原文 · PDF p.20 · S2.I5.i2.p1.1

   The solution to (2.50) for the final condition $\hat{\rho}(1) = \rho_1$ is given at any time $t \in [0,1]$ by

   对终值 $\hat{\rho}(1)=\rho_1$，式（2.50）在任意 $t\in[0,1]$ 的解如下。

$$\hat{\rho}(t,x) = \exp\left(\int_t^1 \nabla \cdot b(\tau, X_{t,\tau}(x))d\tau\right)\rho_1(X_{t,1}(x)) \tag{2.53}$$

S138 · Original / 原文 · PDF p.21 · S2.SS5.p2.1

The proof of Lemma 25 can be found in Appendix B.4. Interestingly, we can obtain a similar result for the solution of the forward and backward FPEs in (2.20) and (2.22). These results make use of auxiliary forward and backward SDEs in which the roles of the forward and backward drifts are switched:

引理 25 证明见附录 B.4。有趣的是，对前后向 FPE（2.20）、（2.22）也能获得类似结果。这里要用辅助前后向 SDE，并交换前后向漂移的角色。

Theorem 26.定理 26

S139 · Original / 原文 · PDF p.21 · Thmtheorem26.p1.1

Given $\epsilon > 0$ and two velocity fields $\hat{b}, \hat{s} \in C^0([0,1], (C^1(\mathbb{R}^d))^d)$, define

给定 ε>0 以及 $\hat{b}$、$\hat{s}\in C^0([0,1],(C^1(\mathbb{R}^d))^d)$，定义下式。

$$\hat{b}_{\mathsf{F}}(t,x) = \hat{b}(t,x) + \epsilon\hat{s}(t,x), \qquad \hat{b}_{\mathsf{B}}(t,x) = \hat{b}(t,x) - \epsilon\hat{s}(t,x), \tag{2.54}$$

S140 · Original / 原文 · PDF p.21 · Thmtheorem26.p1.2

and let Y_t^{F} and Y_t^{B} denote solutions of the following forward and backward SDEs:

令 Y_t^{F}、Y_t^{B} 分别为下列前后向 SDE 的解。

$$dY_t^{\mathsf{F}} = b_{\mathsf{B}}(t, Y_t^{\mathsf{F}})dt + \sqrt{2\epsilon}dW_t, \tag{2.55}$$

S141 · Original / 原文 · PDF p.21 · Thmtheorem26.p1.3

to be solved forward in time from the initial condition $Y_{t=0}^{\mathsf{F}} = x$ independent of W; and

前向方程从独立于 W 的初值 $Y_0^{\mathsf{F}}=x$ 向前求解；后向方程如下。

$$dY_t^{\mathsf{B}} = b_{\mathsf{F}}(t, Y_t^{\mathsf{B}})dt + \sqrt{2\epsilon}dW_t^{\mathsf{B}}, \quad W_t^{\mathsf{B}} = -W_{1-t}, \tag{2.56}$$

S142 · Original / 原文 · PDF p.21 · Thmtheorem26.p1.4

to be solved backwards in time from the final condition $Y_{t=1}^{\mathsf{B}} = x$ independent of W^{B}. Then, given the densities ρ_0 and ρ_1:

后向方程从独立于 WB 的终值 $Y_1^{\mathsf{B}}=x$ 向后求解。于是，对给定密度 ρ_0、ρ_1，有下列结果。

1. 1.

   S143 · Original / 原文 · PDF p.21 · S2.I6.i1.p1.1

   The solution to the forward FPE

   下列前向 FPE 的解：

   $$\partial_t\hat{\rho}_{\mathsf{F}} + \nabla\cdot(\hat{b}_{\mathsf{F}}\hat{\rho}_{\mathsf{F}}) = \epsilon\Delta\hat{\rho}_{\mathsf{F}}, \qquad \hat{\rho}_{\mathsf{F}}(0) = \rho_0, \tag{2.57}$$

   S144 · Original / 原文 · PDF p.21 · S2.I6.i1.p1.2

   can be expressed at $t = 1$ as

   在 t=1 可表示如下。

   $$\hat{\rho}_{\mathsf{F}}(1,x) = \mathbb{E}_{\mathsf{B}}^x\left(\exp\left(-\int_0^1 \nabla\cdot\hat{b}_{\mathsf{F}}(t,Y_t^{\mathsf{B}})dt\right)\rho_0(Y_{t=0}^{\mathsf{B}})\right), \tag{2.58}$$

   S145 · Original / 原文 · PDF p.21 · S2.I6.i1.p1.3

where $\mathbb{E}_{\mathsf{B}}^{x}$ denotes expectation on the path of Y_t^{B} conditional on the event $Y_{t=1}^{\mathsf{B}} = x$.

其中 $\mathbb{E}_{\mathsf{B}}^{x}$ 表示给定 $Y_1^{\mathsf{B}}=x$ 后，对 YB 路径的期望。

2. 2.

S146 · Original / 原文 · PDF p.21 · S2.I6.i2.p1.1

The solution to the backward FPE

下列后向 FPE 的解：

$$\partial_t \hat{\rho}_{\mathsf{B}} + \nabla \cdot (\hat{b}_{\mathsf{B}} \hat{\rho}_{\mathsf{B}}) = -\epsilon \Delta \hat{\rho}_{\mathsf{B}}, \qquad \hat{\rho}_{\mathsf{B}}(1) = \rho_1, \tag{2.59}$$

S147 · Original / 原文 · PDF p.21 · S2.I6.i2.p1.2

can be expressed at any $t = 0$ as

在 t=0 可表示如下。

$$\hat{\rho}_{\mathsf{B}}(0, x) = \mathbb{E}_{\mathsf{F}}^{x} \left(\exp\left(\int_0^1 \nabla \cdot \hat{b}_{\mathsf{B}}(t, Y_t^{\mathsf{F}}) dt \right) \rho_1(Y_{t=1}^{\mathsf{F}}) \right), \tag{2.60}$$

S148 · Original / 原文 · PDF p.21 · S2.I6.i2.p1.3

where $\mathbb{E}_{\mathsf{F}}^{x}$ denotes expectation on the path of Y_t^{F} conditional on $Y_{t=0}^{\mathsf{F}} = x$.

其中 $\mathbb{E}_{\mathsf{F}}^{x}$ 表示给定 $Y_0^{\mathsf{F}}=x$ 后，对 YF 路径的期望。

S149 · Original / 原文 · PDF p.21 · S2.SS5.p3.1

The proof of Theorem 26 can be found in Appendix B.4. Note that to generate data from either $\hat{\rho}_{\mathsf{F}}(1)$ or $\hat{\rho}_{\mathsf{B}}(0)$ assuming that we can sample exactly the PDF at the other end, i.e. ρ_0 and ρ_1 respectively, we would still rely on the equivalent of the forward and backward SDE in (2.33) and (2.34), now used with the approximate drifts in (2.54), i.e.

定理 26 证明见附录 B.4。若能精确采样另一端的密度，即分别为 ρ_0、ρ_1，则从 $\rho\mathsf{F}(1)$ 或 $\rho\mathsf{B}(0)$ 生成数据时，仍使用与前后向 SDE（2.33）、（2.34）对应的方程，但换成式（2.54）的近似漂移，如下。

$$d\hat{X}_t^{\mathsf{F}} = \hat{b}_{\mathsf{F}}(t, \hat{X}_t^{\mathsf{F}}) dt + \sqrt{2\epsilon} dW_t, \tag{2.61}$$

S150 · Original / 原文 · PDF p.21 · S2.SS5.p4.1

and

以及下式。

$$d\hat{X}_t^{\mathsf{B}} = \hat{b}_{\mathsf{B}}(t, \hat{X}_t^{\mathsf{B}}) dt + \sqrt{2\epsilon} dW_t^{\mathsf{B}}, \quad W_t^{\mathsf{B}} = -W_{1-t}, \tag{2.62}$$

S151 · Original / 原文 · PDF p.22 · S2.SS5.p4.2

If we solve (2.61) forward in time from initial data $\hat{X}_{t=0}^{\mathsf{F}} \sim \rho_0$, we then have $\hat{X}_{t=1}^{\mathsf{F}} \sim \hat{\rho}_{\mathsf{F}}(1)$ where $\hat{\rho}_{\mathsf{F}}$ is the solution to the forward FPE (2.57). Similarly If we solve (2.62) backward in time from final data $\hat{X}_{t=1}^{\mathsf{B}} \sim \rho_1$, we then have $\hat{X}_{t=0}^{\mathsf{B}} \sim \hat{\rho}_{\mathsf{B}}(0)$ where $\hat{\rho}_{\mathsf{B}}$ is the solution to the backward FPE (2.59).

若从初始分布 $\hat{X}_{t=0}^{\mathsf{F}} \sim \rho_0$ 出发向前求解式（2.61），则终点满足 $\hat{X}_{t=1}^{\mathsf{F}} \sim \hat{\rho}_{\mathsf{F}}(1)$，其中 $\hat{\rho}_{\mathsf{F}}$ 是前向 Fokker–Planck 方程（2.57）的解。类似地，若从终端分布 $\hat{X}_{t=1}^{\mathsf{B}} \sim \rho_1$ 出发向后求解式（2.62），则 $\hat{X}_{t=0}^{\mathsf{B}} \sim \hat{\rho}_{\mathsf{B}}(0)$；$\hat{\rho}_{\mathsf{B}}$ 是后向 Fokker–Planck 方程（2.59）的解。

S152 · Original / 原文 · PDF p.22 · S2.SS5.p5.1

The results of Lemma 25 and Theorem 26 can be used to test the quality of samples generated by either the ODE (2.32) or the forward and backward SDEs (2.33) and (2.34). In particular, the following two results are direct consequences of Lemma 25 and Theorem 26, respectively:

引理 25 与定理 26 可用于检验由常微分方程（2.32）或前、后向随机微分方程（2.33）、（2.34）生成的样本质量。具体而言，下面两个结果分别直接来自引理 25 和定理 26：

Corollary 27.推论 27

S153 · Original / 原文 · PDF p.22 · Thmtheorem27.p1.3

Under the same conditions as Lemma 25, if $\hat{\rho}(0) = \rho_0$, the cross-entropy of $\hat{\rho}(1)$ relative to ρ_1 is given by

在引理 25 的条件下，若 $\hat{\rho}(0) = \rho_0$，则 $\hat{\rho}(1)$ 相对于 ρ_1 的交叉熵为

$$\begin{aligned} \mathsf{H}(\rho_1 \,\|\, \hat{\rho}(1)) &= -\int_{\mathbb{R}^d} \log \hat{\rho}(1,x)\rho_1(x)dx \\ &= \mathbb{E}_1 \int_0^1 \nabla \cdot b(\tau, X_{1,\tau}(x_1))d\tau - \mathbb{E}_1 \log \rho_0(X_{1,0}(x_1)) \end{aligned} \tag{2.63}$$

S154 · Original / 原文 · PDF p.22 · Thmtheorem27.p1.4

where $\mathbb{E}_1$ denotes an expectation over $x_1 \sim \rho_1$. Similarly, if $\hat{\rho}(1) = \rho_1$, the cross-entropy of $\hat{\rho}(0)$ relative to ρ_0 is given by

其中 $\mathbb{E}_1$ 表示对 $x_1 \sim \rho_1$ 取期望。同理，若 $\hat{\rho}(1) = \rho_1$，则 $\hat{\rho}(0)$ 相对于 ρ_0 的交叉熵为

$$\begin{aligned} \mathsf{H}(\rho_0 \,\|\, \hat{\rho}(0)) &= -\int_{\mathbb{R}^d} \log \hat{\rho}(0,x)\rho_0(x)dx \\ &= -\mathbb{E}_0 \int_0^1 \nabla \cdot b(\tau, X_{0,\tau}(x_0))d\tau - \mathbb{E}_0 \log \rho_1(X_{0,1}(x_0)) \end{aligned} \tag{2.64}$$

S155 · Original / 原文 · PDF p.22 · Thmtheorem27.p1.5

where $\mathbb{E}_0$ denotes an expectation over $x_0 \sim \rho_0$.

其中 $\mathbb{E}_0$ 表示对 $x_0 \sim \rho_0$ 取期望。

Corollary 28.推论 28

S156 · Original / 原文 · PDF p.22 · Thmtheorem28.p1.3

Under the same conditions as Theorem 26, the cross-entropy of $\hat{\rho}_{\mathsf{F}}(1)$ relative to ρ_1 is given by

在定理 26 的条件下，$\hat{\rho}_{\mathsf{F}}(1)$ 相对于 ρ_1 的交叉熵为

$$\begin{aligned} \mathsf{H}(\rho_1 \,\|\, \hat{\rho}_{\mathsf{F}}(1)) &= -\int_{\mathbb{R}^d} \log \hat{\rho}_{\mathsf{F}}(1,x)\rho_1(x)dx \\ &= -\mathbb{E}_1 \log \mathbb{E}_{\mathsf{B}}^{x_1} \left(\exp\left(-\int_0^1 \nabla \cdot b_{\mathsf{F}}(t, Y_t^{\mathsf{B}})dt \right) \rho_0(Y_{t=0}^{\mathsf{B}}) \right), \end{aligned} \tag{2.65}$$

S157 · Original / 原文 · PDF p.22 · Thmtheorem28.p1.4

where $\mathbb{E}_{\mathsf{B}}^{x_1}$ denotes an expectation over Y_t^{B} conditioned on the event $Y_{t=1}^{\mathsf{B}} = x_1$, and $\mathbb{E}_1$ denotes an expectation over $x_1 \sim \rho_1$. Similarly, the cross-entropy of $\hat{\rho}_{\mathsf{B}}(0)$ relative to ρ_0 is given by

其中 $\mathbb{E}_{\mathsf{B}}^{x_1}$ 表示在 $Y_{t=1}^{\mathsf{B}} = x_1$ 的条件下对 Y_t^{B} 取期望，$\mathbb{E}_1$ 表示对 $x_1 \sim \rho_1$ 取期望。同理，$\hat{\rho}_{\mathsf{B}}(0)$ 相对于 ρ_0 的交叉熵为

$$\begin{aligned}\mathsf{H}(\rho_0 \,\|\, \hat{\rho}_{\mathsf{B}}(0)) &= -\int_{\mathbb{R}^d} \log \hat{\rho}_{\mathsf{B}}(0,x)\rho_0(x)dx \\ &= -\mathbb{E}_0 \log \mathbb{E}_{\mathsf{F}}^{x_0}\left(\exp\left(\int_0^1 \nabla\cdot b_{\mathsf{B}}(t,Y_t^{\mathsf{F}})dt\right)\rho_1(Y_{t=1}^{\mathsf{F}})\right),\end{aligned} \tag{2.66}$$

S158 · Original / 原文 · PDF p.22 · Thmtheorem28.p1.5

where $\mathbb{E}_{\mathsf{B}}^{x_0}$ denotes an expectation over Y_t^{F} conditioned on the event $Y_{t=0}^{\mathsf{F}} = x_0$, and $\mathbb{E}_0$ denotes an expectation over $x_0 \sim \rho_0$.

其中，原文此处写作 $\mathbb{E}_{\mathsf{B}}^{x_0}$，指在 $Y_{t=0}^{\mathsf{F}} = x_0$ 的条件下对 Y_t^{F} 取期望；$\mathbb{E}_0$ 表示对 $x_0 \sim \rho_0$ 取期望。

S159 · Original / 原文 · PDF p.23 · S2.SS5.p6.1

If in (2.63), (2.64), (2.65), and (2.66) we approximate the expectations $\mathbb{E}_0$ and $\mathbb{E}_1$ over ρ_0 and ρ_1 by empirical expectations over the available data, these equations allow us to cross-validate different approximations of $\hat{b}$ and $\hat{s}$, as well as to compare the cross-entropies of densities evolved by the TE (2.50) with those of the forward and backward FPEs (2.57) and (2.59).

在式（2.63）至（2.66）中，若以已有数据的经验期望近似分布 ρ_0、ρ_1 下的 $\mathbb{E}_0$、$\mathbb{E}_1$，便能对 $\hat{b}$ 和 $\hat{s}$ 的不同近似进行交叉验证，也能将输运方程（2.50）演化所得密度的交叉熵，与前、后向 Fokker–Planck 方程（2.57）、（2.59）的结果比较。

Remark 29.说明 29

S160 · Original / 原文 · PDF p.23 · Thmtheorem29.p1.5

When using (2.65) and (2.66) in practice, taking the log of the expectations $\mathbb{E}_{\mathsf{B}}^{x_1}$ and $\mathbb{E}_{\mathsf{F}}^{x_0}$ may create difficulties, such as when using Hutchinson's trace estimator to compute the divergence of b_{F} or b_{B}, which will introduce a bias. One way to remove this bias is to use Jensen's inequality, which leads to the upper bounds

实际使用式（2.65）、（2.66）时，对期望 $\mathbb{E}_{\mathsf{B}}^{x_1}$ 与 $\mathbb{E}_{\mathsf{F}}^{x_0}$ 取对数可能带来困难。例如，用 Hutchinson 迹估计量计算 b_{F} 或 b_{B} 的散度时，这一步会引入偏差。避免这一偏差的一种办法是应用 Jensen 不等式，得到如下上界：

$$\mathsf{H}(\rho_1 \,\|\, \hat{\rho}_{\mathsf{F}}(1)) \le \int_0^1 \mathbb{E}_1\mathbb{E}_{\mathsf{B}}^{x_1}\nabla\cdot b_{\mathsf{F}}(t,Y_t^{\mathsf{B}})dt - \mathbb{E}_1\mathbb{E}_{\mathsf{B}}^{x_1}\log\rho_0(Y_{t=0}^{\mathsf{B}}), \tag{2.67}$$

S161 · Original / 原文 · PDF p.23 · Thmtheorem29.p1.6

and

以及

$$\mathsf{H}(\rho_0 \,\|\, \hat{\rho}_{\mathsf{B}}(0)) \le -\mathbb{E}_0\mathbb{E}_{\mathsf{F}}^{x_0}\int_0^1 \nabla\cdot b_{\mathsf{B}}(t,Y_t^{\mathsf{F}})dt - \mathbb{E}_0\mathbb{E}_{\mathsf{F}}^{x_0}\log\rho_1(Y_{t=1}^{\mathsf{F}}). \tag{2.68}$$

S162 · Original / 原文 · PDF p.23 · Thmtheorem29.p1.7

However, these bounds are not sharp in general – in fact, using calculations similar to the one presented in the proof of Theorem 28, we can derive exact expressions that capture precisely what is lost when applying Jensen's inequality:

不过，这些上界一般并不紧。事实上，依照定理 28 证明中的类似计算，可以推得以下精确表达式，恰好说明应用 Jensen 不等式损失了什么：

$$\mathsf{H}(\rho_1 \,\|\, \hat{\rho}_{\mathsf{F}}(1)) = \int_0^1 \mathbb{E}_1\mathbb{E}_{\mathsf{B}}^{x_1}\left(\nabla\cdot b_{\mathsf{F}}(t,Y_t^{\mathsf{B}}) - \epsilon|\nabla\log\hat{\rho}_{\mathsf{F}}(t,Y_t^{\mathsf{B}})|^2\right)dt - \mathbb{E}_1\mathbb{E}_{\mathsf{B}}^{x_1}\log\rho_0(Y_{t=0}^{\mathsf{B}}), \tag{2.69}$$

S163 · Original / 原文 · PDF p.23 · Thmtheorem29.p1.8

and

以及

$$\mathsf{H}(\rho_0 \,\|\, \hat{\rho}_{\mathsf{B}}(0)) = -\mathbb{E}_0\mathbb{E}_{\mathsf{F}}^{x_0}\int_0^1 \left(\nabla\cdot b_{\mathsf{B}}(t,Y_t^{\mathsf{F}}) + \epsilon|\nabla\log\hat{\rho}_{\mathsf{B}}(t,Y_t^{\mathsf{F}})|^2\right)dt - \mathbb{E}_0\mathbb{E}_{\mathsf{F}}^{x_0}\log\rho_1(Y_{t=1}^{\mathsf{F}}). \quad (2.70)$$

S164 · Original / 原文 · PDF p.23 · Thmtheorem29.p1.9

Unfortunately, since $\nabla\log\hat{\rho}_{\mathsf{F}} \neq \hat{s}$ and $\nabla\log\hat{\rho}_{\mathsf{B}} \neq \hat{s}$ in general due to approximation errors, we do not know how to estimate the extra terms on the right-hand side of (2.69) and (2.70). One possibility is to use $\hat{s}$ as a proxy for $\nabla\log\hat{\rho}_{\mathsf{F}}$ and $\nabla\log\hat{\rho}_{\mathsf{B}}$, which may be useful in practice, but this approximation is uncontrolled in general.

遗憾的是，受近似误差影响，通常有 $\nabla\log\hat{\rho}_{\mathsf{F}} \neq \hat{s}$、$\nabla\log\hat{\rho}_{\mathsf{B}} \neq \hat{s}$，因此还不知道怎样估计式（2.69）、（2.70）右侧的附加项。一种可能是用 $\hat{s}$ 代替这两个对数密度梯度。实践中这或许有用，但一般无法控制其近似误差。

3 Instantiations and extensions

3 实例与扩展

S165 · Original / 原文 · PDF p.23 · S3.p1.1

In this section, we instantiate the stochastic interpolant framework discussed in Section 2.

本节给出第 2 节随机插值框架的具体实例。

3.1 Diffusive interpolants

3.1 扩散插值

S166 · Original / 原文 · PDF p.23 · S3.SS1.p1.1

Recently, there has been a surge of interest in the construction of generative models through diffusive bridge processes *(Peluchetti, 2022; Liu et al., 2022; Somnath et al., 2023)*. In this section, we connect these approaches with our own, highlighting that stochastic interpolants allow us to manipulate certain bridge processes in a simpler and more direct manner. We also show that this perspective leads to a generative process that samples any target density ρ_1 by pushing a point mass at any $x_0 \in \mathbb{R}^d$ through an SDE. We begin by introducing a new kind of interpolant:

近来，基于扩散桥过程构造生成模型引起了广泛兴趣（Peluchetti，2022；Liu 等，2022；Somnath 等，2023）。本节将这些方法与我们的框架联系起来，说明怎样借助随机插值更直接、更简便地处理某些桥过程。我们还将说明，这一视角给出一种生成过程：从任意 $x_0 \in \mathbb{R}^d$ 处的点质量出发，按随机微分方程推进即可从任意目标密度 ρ_1 采样。首先引入一种新的插值过程：

Definition 30 (Diffusive interpolant).定义 30（扩散插值）

S167 · Original / 原文 · PDF p.24 · Thmtheorem30.p1.1

Given two probability density functions $\rho_0, \rho_1 : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$, a diffusive interpolant between ρ_0 and ρ_1 is a stochastic process x_t^{d} defined as

给定两个概率密度函数 $\rho_0, \rho_1 : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$，称按下式定义的随机过程 x_t^{d} 为 ρ_0 与 ρ_1 之间的扩散插值：

$$x_t^{\mathsf{d}} = I(t, x_0, x_1) + \sqrt{2a(t)} B_t, \qquad t \in [0, 1], \tag{3.1}$$

S168 · Original / 原文 · PDF p.24 · Thmtheorem30.p1.2

where:

其中：

1. 1.

   S169 · Original / 原文 · PDF p.24 · S3.I1.i1.p1.1

   $I(t, x_0, x_1)$ is as in Definition 1;

   $I(t, x_0, x_1)$ 与定义 1 相同；

2. 2.

   S170 · Original / 原文 · PDF p.24 · S3.I1.i2.p1.1

   $(x_0, x_1) \sim \nu$ with ν satisfying (2.3) in Definition 1;

   $(x_0, x_1) \sim \nu$，其中 ν 满足定义 1 中的条件（2.3）；

3. 3.

S171 · Original / 原文 · PDF p.24 · S3.I1.i3.p1.1

$a(t) \in C^2([0,1])$ with $a(0) > 0$ and $a(t) \geq 0$ for all $t \in (0,1]$, and;

$a(t) \in C^2([0,1])$，且 $a(0) > 0$，对一切 $t \in (0,1]$ 均有 $a(t) \geq 0$；

4. 4.

S172 · Original / 原文 · PDF p.24 · S3.I1.i4.p1.1

B_t is a standard Brownian bridge process, independent of x_0 and x_1.

B_t 为标准布朗桥过程，独立于 x_0 和 x_1。

S173 · Original / 原文 · PDF p.24 · S3.SS1.p2.1

Pathwise, (3.1) is different from the stochastic interpolant introduced in Definition 1: in particular, x_t^{d} is continuous but not differentiable in time. At the same time, since B_t is a Gaussian process with mean zero and variance $\mathbb{E}B_t^2 = t(1-t)$, (3.1) has the same single-time statistics and time-dependent density $\rho(t,x)$ as the stochastic interpolant (2.1) if we set $\gamma(t) = \sqrt{2a(t)t(1-t)}$, i.e.

从样本路径看，式（3.1）与定义 1 的随机插值不同：特别地，x_t^{d} 关于时间连续却不可微。另一方面，B_t 是均值为零、方差为 $\mathbb{E}B_t^2 = t(1-t)$ 的高斯过程。因此只要取 $\gamma(t) = \sqrt{2a(t)t(1-t)}$，式（3.1）便与随机插值（2.1）具有相同的单时刻统计量和含时密度 $\rho(t,x)$，即

$$x_t = I(t, x_0, x_1) + \sqrt{2a(t)t(1-t)}z \quad \text{with} \quad (x_0, x_1) \sim \nu,\ z \sim \mathsf{N}(0, Id),\ (x_0, x_1) \perp z. \tag{3.2}$$

S174 · Original / 原文 · PDF p.24 · S3.SS1.p2.2

As a result, (3.1) and (3.2) lead to the same generative models. Technically, it is easier to work with (3.2) than with (3.1), because it avoids the use of Itô calculus, and enables direct sampling of x_t using samples from ρ_0, ρ_1, and $\mathsf{N}(0, Id)$. However, (3.1) sheds light on some interesting properties of the generative models based on (3.2), i.e. stochastic interpolants with $\gamma(t) = \sqrt{2a(t)t(1-t)}$. To see why, we now re-derive the transport equation for the density $\rho(t,x)$ shared by (3.1) and (3.2) using the relation (3.1) using Fourier analysis. For simplicity, we focus on the case where $a(t)$ is constant in time, i.e. we set $a(t) = a > 0$ in (3.1).

因此式（3.1）和（3.2）导出相同的生成模型。技术上，处理式（3.2）更容易：无需使用 Itô 微积分，直接从 ρ_0、ρ_1 和标准正态分布 $\mathsf{N}(0, \mathrm{Id})$ 中抽取样本即可构造 x_t。但式（3.1）揭示了基于式（3.2）的模型——即取 $\gamma(t) = \sqrt{2a(t)t(1-t)}$ 的随机插值——的一些有趣性质。为说明这一点，下面从关系式（3.1）出发，用傅里叶分析重新推导这两个过程共有的密度 $\rho(t,x)$ 的输运方程。为简便起见，只考虑 $a(t)$ 为常数的情形，即令式（3.1）中的 $a(t) = a > 0$。

S175 · Original / 原文 · PDF p.24 · S3.SS1.p3.2

To begin, recall that the Brownian Bridge B_t can be expressed in terms of the Wiener process W_t as $B_t = W_t - tW_{t=1}$. Moreover, it satisfies the SDE obtained by conditioning on $B_{t=1} = 0$ via Doob's h-transform *(Doob, 1984)*:

首先回顾，布朗桥 B_t 可由维纳过程 W_t 表为 $B_t = W_t - tW_{t=1}$。此外，通过 Doob 的 h 变换对 $B_{t=1} = 0$ 施加条件，可以得到它所满足的随机微分方程（Doob，1984）：

$$dB_t = -\frac{B_t}{1-t}dt + dW_t, \qquad B_{t=0} = 0. \tag{3.3}$$

S176 · Original / 原文 · PDF p.24 · S3.SS1.p3.3

A direct application of Itô's formula implies that

直接应用 Itô 公式可得

$$de^{ik\cdot x_t^{\mathrm{d}}} = ik\cdot\Big(\partial_t I(t,x_0,x_1) - \frac{\sqrt{2a}B_t}{1-t}\Big)e^{ik\cdot x_t^{\mathrm{d}}}dt - a|k|^2 e^{ik\cdot x_t^{\mathrm{d}}}dt + \sqrt{2a}ik\cdot dW_t e^{ik\cdot x_t^{\mathrm{d}}}. \tag{3.4}$$

S177 · Original / 原文 · PDF p.24 · S3.SS1.p3.4

Taking the expectation of this expression and using the independence between (x_0,x_1) and B_t, we deduce that

对此式取期望，并利用 (x_0,x_1) 与 B_t 相互独立，可推出

$$\partial_t \mathbb{E}e^{ik\cdot x_t^{\mathrm{d}}} = ik\cdot\mathbb{E}\Big(\Big(\partial_t I(t,x_0,x_1) - \frac{\sqrt{2a}B_t}{1-t}\Big)e^{ik\cdot x_t^{\mathrm{d}}}\Big) - a|k|^2\mathbb{E}e^{ik\cdot x_t^{\mathrm{d}}}. \tag{3.5}$$

S178 · Original / 原文 · PDF p.24 · S3.SS1.p3.5

Since for all fixed $t\in[0,1]$ we have $B_t \overset{d}{=} \sqrt{t(1-t)}z$ and $x_t^{\mathrm{d}} \overset{d}{=} x_t$ with x_t defined in (3.2), the time derivative (3.5) can also be written as

对任意固定的 $t\in[0,1]$，B_t 与 $\sqrt{t(1-t)}z$ 同分布，x_t^{d} 与式（3.2）定义的 x_t 同分布。因此式（3.5）的时间导数也可写为

$$\partial_t \mathbb{E}e^{ik\cdot x_t} = ik\cdot\mathbb{E}\Big(\Big(\partial_t I(t,x_0,x_1) - \frac{\sqrt{2at}\,z}{\sqrt{1-t}}\Big)e^{ik\cdot x_t}\Big) - a|k|^2\mathbb{E}e^{ik\cdot x_t}. \tag{3.6}$$

S179 · Original / 原文 · PDF p.24 · S3.SS1.p3.6

Moreover, since by definition of their probability density we have $\mathbb{E}e^{ik\cdot x_t^{\mathrm{d}}} = \mathbb{E}e^{ik\cdot x_t} = \int_{\mathbb{R}^d} e^{ik\cdot x}\rho(t,x)dx$, we can deduce from (3.6) that $\rho(t)$ satisfies

此外，由概率密度的定义，$\mathbb{E}e^{ik\cdot x_t^{\mathrm{d}}} = \mathbb{E}e^{ik\cdot x_t} = \int_{\mathbb{R}^d} e^{ik\cdot x}\rho(t,x)\,dx$。故由式（3.6）可推出 $\rho(t)$ 满足

$$\partial_t\rho + \nabla\cdot(u\rho) = a\Delta\rho, \tag{3.7}$$

S180 · Original / 原文 · PDF p.24 · S3.SS1.p3.7

where we defined

这里定义了

$$u(t,x) = \mathbb{E}\Big(\partial_t I(t,x_0,x_1) - \frac{\sqrt{2at}\,z}{\sqrt{1-t}}\Big|x_t = x\Big). \tag{3.8}$$

S181 · Original / 原文 · PDF p.25 · S3.SS1.p3.8

For the interpolant x_t in (3.2), we have from the definitions of b and s in (2.10) and (2.14) that

对于式（3.2）中的插值 x_t，由式（2.10）和（2.14）中 b 与 s 的定义可得

$$\begin{aligned} b(t,x) &= \mathbb{E}\Big(\partial_t I(t,x_0,x_1) + \frac{a(1-2t)z}{\sqrt{2t(1-t)}}\Big|x_t = x\Big),\\ s(t,x) &= \nabla\log\rho(t,x) = -\frac{1}{\sqrt{2at(1-t)}}\mathbb{E}(z|x_t = x), \end{aligned} \tag{3.9}$$

S182 · Original / 原文 · PDF p.25 · S3.SS1.p3.9

As a result, $u - s = b$ and (3.7) can also be written as the TE (2.9) using $\Delta\rho = \nabla \cdot (s\rho)$.

因此，原文写作 $u - s = b$；利用 $\Delta\rho = \nabla \cdot (s\rho)$，式（3.7）也可改写为输运方程（2.9）。【译注：式（3.7）的扩散项为 $a\Delta\rho$；一般 a 下移项所得输运速度应为 $u - as$，原文简式对应 $a = 1$。】

Conditional sampling.

条件采样

S183 · Original / 原文 · PDF p.25 · S3.SS1.SSS0.Px1.p1.1

Remarkably, the drift u defined in (3.8) remains non-singular for all $t \in [0, 1]$ (including $t = 0$) even if ρ_0 is replaced by a point mass at x_0; by contrast, both b and s are singular at $t = 0$ in this case. Hence, the SDE associated with the FPE (3.7) provides us with a generative model that samples ρ_1 from a base measure concentrated at a single x_0 (i.e. such that the density ρ_0 is replaced by a point mass measure at $x = x_0$). We formalize this result in the following theorem:

值得注意的是，即便将 ρ_0 换成 x_0 处的点质量，式（3.8）定义的漂移 u 在整个 $t \in [0, 1]$（包括 $t = 0$）上仍无奇性；相比之下，此时 b 和 s 在 $t = 0$ 都是奇异的。因此，与 Fokker–Planck 方程（3.7）对应的随机微分方程构成了一个生成模型：它能从集中于单点 x_0 的基测度（即用 $x = x_0$ 处的点质量测度代替密度 ρ_0）出发，对 ρ_1 采样。下述定理给出正式表述：

Theorem 31.定理 31

S184 · Original / 原文 · PDF p.25 · Thmtheorem31.p1.1

Assume that $I(t, x_0, x_1) = x_0$ for $t \in [0, \delta]$ with some $\delta \in (0, 1]$. Given any $a > 0$, let

假设存在 $\delta \in (0, 1]$，使得对 $t \in [0, \delta]$ 均有 $I(t, x_0, x_1) = x_0$。给定任意 $a > 0$，令

$$u^{\mathsf{d}}(t, x, x_0) = \mathbb{E}_{x_1, z}\Big(\partial_t I(t, x_0, x_1) - \frac{\sqrt{2at}\, z}{\sqrt{1 - t}}\Big| x_t = x\Big), \tag{3.10}$$

S185 · Original / 原文 · PDF p.25 · Thmtheorem31.p1.2

where x_t is given by (3.2) and where $\mathbb{E}_{x_1, z}(\cdot | x_t = x)$ denotes an expectation over $x_1 \sim \rho_1 \perp z \sim \mathsf{N}(0, Id)$ conditioned on $x_t = x$ with $x_0 \in \mathbb{R}^d$ fixed. Then $u^{\mathsf{d}}(\cdot, \cdot, x_0) \in C^0([0, 1]; (C^p(\mathbb{R}^d))^d)$ for any $p \in \mathbb{N}$ and $x_0 \in \mathbb{R}^d$. Moreover, the solutions to the forward SDE

其中 x_t 由式（3.2）给出；固定 $x_0 \in \mathbb{R}^d$，$\mathbb{E}_{x_1, z}(\cdot \mid x_t = x)$ 表示从 ρ_1 与标准正态分布 $\mathsf{N}(0, \mathrm{Id})$ 独立抽取 x_1、z 后，在 $x_t = x$ 下取的条件期望。于是，对任意 $p \in \mathbb{N}$、$x_0 \in \mathbb{R}^d$，有 $u^{\mathsf{d}}(\cdot, \cdot, x_0) \in C^0([0, 1]; (C^p(\mathbb{R}^d))^d)$。此外，下列前向随机微分方程的解

$$dX_t^{\mathsf{d}} = u^{\mathsf{d}}(t, X_t^{\mathsf{d}}, x_0)dt + \sqrt{2a}\, dW_t, \qquad X_{t=0}^{\mathsf{d}} = x_0, \tag{3.11}$$

S186 · Original / 原文 · PDF p.25 · Thmtheorem31.p1.3

are such that $X_{t=1}^{\mathsf{d}} \sim \rho_1$.

满足 $X_{t=1}^{\mathsf{d}} \sim \rho_1$。

S187 · Original / 原文 · PDF p.25 · S3.SS1.SSS0.Px1.p2.1

Note that the additional assumption we make on $I(t, x_0, x_1)$ is consistent with the requirements in Definition 1 and Assumption 5: this additional assumption is made for simplicity and can probably be relaxed to $\partial_t I(t = 0, x_0, x_1) = 0$.

我们对 $I(t, x_0, x_1)$ 施加的额外假设与定义 1 和假设 5 相容。这个假设只是为了简化推导，很可能可以放宽为 $\partial_t I(t = 0, x_0, x_1) = 0$。

S188 · Original / 原文 · PDF p.25 · S3.SS1.SSS0.Px1.p3.1

The proof of Theorem 31 is given in Appendix B.5. It relies on the calculations that led to (3.8), along with the observation that at $t = 0$ and $x = x_0$,

定理 31 的证明见附录 B.5。它依赖导出式（3.8）的计算，并用到下列观察：在 $t = 0$、$x = x_0$ 时，

$$u^{\mathsf{d}}(t = 0, x_0, x_0) = \mathbb{E}_{x_1} \left(\partial_t I(t = 0, x_0, x_1) \right), \tag{3.12}$$

S189 · Original / 原文 · PDF p.25 · S3.SS1.SSS0.Px1.p3.2

whereas at $t = 1$ and any $x \in \mathbb{R}^d$, we have

而当 $t = 1$ 时，对任意 $x \in \mathbb{R}^d$ 均有

$$u^{\mathsf{d}}(t = 1, x, x_0) = \partial_t I(t = 1, x_0, x) + 2a \nabla \log \rho_1(x), \tag{3.13}$$

S190 · Original / 原文 · PDF p.26 · S3.SS1.SSS0.Px1.p3.3

which are both well-defined. To put the result in Theorem (31) in perspective, observe that no probability flow ODE with $b \in C^0([0, 1]; (C^p(\mathbb{R}^d))^d)$ can achieve the same feat as the diffusion in (3.11). This is because the solutions of such an ODE are unique, and therefore can only map x_0 onto a single point at time $t = 1$. Moreover, $u^{\mathsf{d}}(t, x, x_0)$ is the unique minimizer of the objective function

这两个端点的表达式都有良好定义。要理解定理 31 的意义，可注意到：任何满足 $b \in C^0([0, 1]; (C^p(\mathbb{R}^d))^d)$ 的概率流常微分方程，都无法像式（3.11）的扩散那样从单点生成一个分布。因为这样的常微分方程解唯一，从 x_0 出发，在 $t = 1$ 只能到达一个点。此外，$u^{\mathsf{d}}(t, x, x_0)$ 是下列目标函数的唯一极小化解：

$$\mathcal{L}_{u^{\mathsf{d}}}[\hat{u}^{\mathsf{d}}] = \int_0^1 \mathbb{E}_{x_1, z} \left(|\hat{u}^d(t, x_t, x_0)|^2 - 2 \Big(\partial_t I(t, x_0, x_1) - \frac{\sqrt{2at} z}{\sqrt{(1 - t)}} \Big) \cdot \hat{u}^d(t, x_t, x_0) \right) dt.$$

3.2 One-sided interpolants for Gaussian ρ_0

3.2 高斯基分布的单侧插值

S191 · Original / 原文 · PDF p.26 · S3.SS2.p1.1

A common choice of base density for generative modeling in the absence of prior information is to choose $\rho_0 = \mathsf{N}(0, Id)$. In this setting, we can group the effect of the latent variable z with x_0. This leads to a simpler type of stochastic interpolant that, in particular, will enables us to instantiate score-based diffusion within our general framework (see Section 3.4).

没有先验信息时，生成建模常取 $\rho_0 = \mathsf{N}(0, \mathrm{Id})$ 作为基密度。此时可以将潜变量 z 的作用与 x_0 合并，得到一种形式更简单的随机插值；特别地，可以由此在统一框架内导出基于得分的扩散模型（见第 3.4 节）。

Definition 32 (One-sided stochastic interpolant).定义 32（单侧随机插值）

S192 · Original / 原文 · PDF p.26 · Thmtheorem32.p1.1

Given a probability density function $\rho_1 : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$, a one-sided stochastic interpolant between $\mathsf{N}(0, Id)$ and ρ_1 is a stochastic process x_t^{os}

给定概率密度函数 $\rho_1 : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$，标准正态分布 $\mathsf{N}(0, \mathrm{Id})$ 与 ρ_1 之间的单侧随机插值为如下随机过程 x_t^{os}：

$$x_t^{\mathsf{os}} = \alpha(t) z + J(t, x_1), \qquad t \in [0, 1] \tag{3.14}$$

S193 · Original / 原文 · PDF p.26 · Thmtheorem32.p1.2

that fulfills the requirements:

它满足下列条件：

1. 1.

   S194 · Original / 原文 · PDF p.26 · S3.I2.i1.p1.1

   $J \in C^2([0,1], C^2(\mathbb{R}^d)^d)$ satisfies the boundary conditions $J(0,x_1)=0$ and $J(1,x_1)=x_1$.

   $J \in C^2([0,1], C^2(\mathbb{R}^d)^d)$，且满足边界条件 $J(0,x_1)=0$、$J(1,x_1)=x_1$；

2. 2.

   S195 · Original / 原文 · PDF p.26 · S3.I2.i2.p1.1

   x_1 and z are independent random variables drawn from ρ_1 and $\mathsf{N}(0, Id)$, respectively.

   x_1 与 z 相互独立，分别取自 ρ_1 与 $\mathsf{N}(0, \mathrm{Id})$；

3. 3.

   S196 · Original / 原文 · PDF p.26 · S3.I2.i3.p1.1

   $\alpha : [0,1] \to \mathbb{R}$ satisfies $\alpha(0)=1$, $\alpha(1)=0$, $\alpha(t)>0$ for all $t \in [0,1)$, and $\alpha^2 \in C^2([0,1])$.

   $\alpha : [0,1] \to \mathbb{R}$ 满足 $\alpha(0)=1$、$\alpha(1)=0$，对一切 $t \in [0,1)$ 均有 $\alpha(t)>0$，并且 $\alpha^2 \in C^2([0,1])$。

S197 · Original / 原文 · PDF p.26 · S3.SS2.p2.1

By construction, $x^{\mathsf{os}}_{t=0} = z \sim \mathsf{N}(0, Id)$ and $x^{\mathsf{os}}_{t=1} = x_1 \sim \rho_1$, so that the distribution of the stochastic process x^{os}_t bridges $\mathsf{N}(0, Id)$ and ρ_1. It is easy to see that the one-sided stochastic interpolant defined in (3.14) will have the same density as the stochastic interpolant defined in (2.1) if we set $I(t, x_0, x_1) = J_t(x_1) + \delta(t)x_0$ and take $\delta^2(t) + \gamma^2(t) = \alpha^2(t)$. Restricting to this case, our earlier theoretical results apply where the velocity field b defined in (2.10) becomes

按构造，$x^{\mathsf{os}}_{t=0} = z \sim \mathsf{N}(0, \mathrm{Id})$，而 $x^{\mathsf{os}}_{t=1} = x_1 \sim \rho_1$，因此过程的分布连接了标准正态分布与 ρ_1。容易看出，若令 $I(t, x_0, x_1) = J_t(x_1) + \delta(t)x_0$，并取 $\delta^2(t) + \gamma^2(t) = \alpha^2(t)$，则式（3.14）的单侧随机插值与式（2.1）的随机插值密度相同。限定到这种情况，前述理论结果仍然适用，式（2.10）所定义的速度场 b 变为

$$b(t,x) = \mathbb{E}(\dot\alpha(t)z + \partial_t J(t,x_1) | x^{\mathsf{os}}_t = x), \tag{3.15}$$

S198 · Original / 原文 · PDF p.26 · S3.SS2.p2.2

and the quadratic objective in (2.13) becomes

式（2.13）的二次目标则变为

$$\mathcal{L}_b[\hat b] = \int_0^1 \mathbb{E}\left(\tfrac{1}{2}|\hat b(t, x^{\mathsf{os}}_t)|^2 - (\dot\alpha(t)z + \partial_t J(t,x_1)) \cdot \hat b(t, x^{\mathsf{os}}_t)\right) dt. \tag{3.16}$$

S199 · Original / 原文 · PDF p.26 · S3.SS2.p2.3

In the expression above, x^{os}_t is given by (3.14) and the expectation $\mathbb{E}$ is taken independently over $x_1 \sim \rho_1$ and $z \sim \mathsf{N}(0, Id)$. Similarly, the score is given by

上式中，x^{os}_t 由式（3.14）给出，$\mathbb{E}$ 对相互独立的 $x_1 \sim \rho_1$ 与 $z \sim \mathsf{N}(0, \mathrm{Id})$ 取期望。类似地，得分为

$$s(t,x) = -\alpha^{-1}(t)\eta_z(t,x), \qquad \eta_z(t,x) = \mathbb{E}(z | x^{\mathsf{os}}_t = x), \tag{3.17}$$

S200 · Original / 原文 · PDF p.27 · S3.SS2.p2.4

where $\eta_z(t,x)$ is the equivalent of the denoiser defined in (2.17). These functions are the unique minimizers of the objectives

其中 $\eta_z(t,x)$ 对应于式（2.17）定义的去噪器。这些函数是以下目标的唯一极小化解：

$$\mathcal{L}_s[\hat{s}] = \int_0^1 \mathbb{E}\left(\tfrac{1}{2}|\hat{s}(t,x_t^{\text{os}})|^2 + \gamma^{-1}(t) z \cdot \hat{s}(t,x_t^{\text{os}})\right) dt, \tag{3.18}$$

$$\mathcal{L}_{\eta_z}[\hat{\eta}_z] = \int_0^1 \mathbb{E}\left(\tfrac{1}{2}|\hat{\eta}_z(t,x_t^{\text{os}})|^2 - z \cdot \hat{\eta}_z(t,x_t^{\text{os}})\right) dt. \tag{3.19}$$

S201 · Original / 原文 · PDF p.27 · S3.SS2.p2.5

Moreover, we can weaken Assumption 5 to the following requirement:

此外，还可以把假设 5 放宽为下列要求：

Assumption 33.假设 33

S202 · Original / 原文 · PDF p.27 · Thmtheorem33.p1.2

The density $\rho_1 \in C^2(\mathbb{R}^d)$, satisfies $\rho_1(x) > 0$ for all $x \in \mathbb{R}^d$, and:

密度 $\rho_1 \in C^2(\mathbb{R}^d)$，对一切 $x \in \mathbb{R}^d$ 有 $\rho_1(x) > 0$，并且：

$$\int_{\mathbb{R}^d} |\nabla \log \rho_1(x)|^2 \rho_1(x) dx < \infty. \tag{3.20}$$

S203 · Original / 原文 · PDF p.27 · Thmtheorem33.p1.3

The function J satisfies

函数 J 满足

$$\exists C_1 < \infty : \quad |\partial_t J(t,x_1)| \le C_1 |x_1| \quad \text{for all} \quad (t,x_1) \in [0,1] \times \mathbb{R}^d, \tag{3.21}$$

S204 · Original / 原文 · PDF p.27 · Thmtheorem33.p1.4

and

以及

$$\exists M_1, M_2 < \infty \ : \ \mathbb{E}\left[|\partial_t J(t,x_1)|^4\right] \le M_1; \quad \mathbb{E}\left[|\partial_t^2 J(t,x_1)|^2\right] \le M_2, \quad \text{for all} \quad t \in [0,1], \tag{3.22}$$

S205 · Original / 原文 · PDF p.27 · Thmtheorem33.p1.5

where the expectation is taken over $x_1 \sim \rho_1$.

其中期望按 $x_1 \sim \rho_1$ 计算。

Remark 34.说明 34

S206 · Original / 原文 · PDF p.27 · Thmtheorem34.p1.1

The construction above can easily be generalized to the case where $\rho_0 = \mathsf{N}(0, C_0)$ with C_0 a positive-definite matrix. Without loss of generality, we can then assume that C_0 can be represented as $C_0 = \sigma_0 \sigma_0^\intercal$ where σ_0 is a lower-triangular matrix and replace (3.14)

上述构造可以直接推广到 $\rho_0 = \mathsf{N}(0, C_0)$，其中 C_0 为正定矩阵。不失一般性，可将 C_0 表为 $C_0 = \sigma_0 \sigma_0^\intercal$，其中 σ_0 是下三角矩阵，并将式（3.14）替换为

$$x_t^{\mathsf{os}} = \alpha(t)\sigma_0 z + J(t, x_1), \qquad t \in [0, 1], \tag{3.23}$$

S207 · Original / 原文 · PDF p.27 · Thmtheorem34.p1.2

with J and α satisfying the conditions listed in Definition 32 and where $z \sim \mathsf{N}(0, Id)$.

其中 J 与 α 满足定义 32 的条件，$z \sim \mathsf{N}(0, \mathrm{Id})$。

3.3 Mirror interpolants

3.3 镜像插值

S208 · Original / 原文 · PDF p.27 · S3.SS3.p1.1

Another practically relevant setting is when the base and the target are the same density ρ_1. In this setting we can define a stochastic interpolant as:

另一种具有实际意义的情形是：基分布与目标分布具有同一密度 ρ_1。此时可作如下定义：

Definition 35 (Mirror stochastic interpolant).定义 35（镜像随机插值）

S209 · Original / 原文 · PDF p.27 · Thmtheorem35.p1.1

Given a probability density function $\rho_1 : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$, a mirror stochastic interpolant between ρ_0 and itself is a stochastic process x_t^{mir}

给定概率密度函数 $\rho_1 : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$，在同一密度的两端之间定义镜像随机插值过程 x_t^{mir} 如下（原文此句将两端写作 ρ_0；下述采样条件使用 ρ_1）：

$$x_t^{\mathsf{mir}} = K(t, x_1) + \gamma(t) z, \qquad t \in [0, 1] \tag{3.24}$$

S210 · Original / 原文 · PDF p.27 · Thmtheorem35.p1.2

that fulfills the requirements:

该过程满足：

1. 1.

   S211 · Original / 原文 · PDF p.27 · S3.I3.i1.p1.1

   $K \in C^2([0, 1], C^2(\mathbb{R}^d)^d)$ satisfies the boundary conditions $K(0, x_1) = x_1$ and $K(1, x_1) = x_1$.

   $K \in C^2([0, 1], C^2(\mathbb{R}^d)^d)$，且满足边界条件 $K(0, x_1) = x_1$、$K(1, x_1) = x_1$；

2. 2.

   S212 · Original / 原文 · PDF p.27 · S3.I3.i2.p1.1

   x_1 and z are random variables drawn independently from ρ_1 and $\mathsf{N}(0, Id)$, respectively.

   随机变量 x_1 与 z 分别从 ρ_1 与 $\mathsf{N}(0, \mathrm{Id})$ 独立抽取；

3. 3.

S213 · Original / 原文 · PDF p.27 · S3.I3.i3.p1.1

$\gamma : [0,1] \to \mathbb{R}$ satisfies $\gamma(0) = \gamma(1) = 0$, $\gamma(t) > 0$ for all $t \in (0,1)$, and $\gamma^2 \in C^1([0,1])$.

$\gamma : [0,1] \to \mathbb{R}$ 满足 $\gamma(0) = \gamma(1) = 0$，对一切 $t \in (0,1)$ 有 $\gamma(t) > 0$，且 $\gamma^2 \in C^1([0,1])$。

S214 · Original / 原文 · PDF p.28 · S3.SS3.p2.1

By construction, $x^{\text{mir}}_{t=0} = x^{\text{mir}}_{t=1} = x_1 \sim \rho_1$, so that the distribution of the stochastic process x^{mir}_t bridges ρ_1 to itself. Note that a valid choice is $K(t, x_1) = \alpha(t)x_1$ with $\alpha(0) = \alpha(1) = 1$ (e.g. $\alpha(t) = 1$): in this case, mirror interpolants are related to denoisers, as will be discussed in Section 5.2.

按构造，$x^{\text{mir}}_{t=0} = x^{\text{mir}}_{t=1} = x_1 \sim \rho_1$，故过程 x^{mir}_t 的分布从 ρ_1 出发又回到 ρ_1。一种合法选择是 $K(t, x_1) = \alpha(t)x_1$，其中 $\alpha(0) = \alpha(1) = 1$（例如 $\alpha(t) = 1$）。此时镜像插值与去噪器相关，详见第 5.2 节。

S215 · Original / 原文 · PDF p.28 · S3.SS3.p3.1

It is easy to see that our earlier theoretical results apply where the velocity field b defined in (2.10) becomes

容易看出，此时前述理论仍适用；式（2.10）定义的速度场 b 变为

$$b(t,x) = \mathbb{E}(\partial_t K(t, x_1) + \dot\gamma(t) z | x^{\text{mir}}_t = x), \tag{3.25}$$

S216 · Original / 原文 · PDF p.28 · S3.SS3.p3.2

and the quadratic objective in (2.13) becomes

式（2.13）的二次目标变为

$$\mathcal{L}_b[\hat b] = \int_0^1 \mathbb{E}\left(\tfrac{1}{2}|\hat b(t, x^{\text{mir}}_t)|^2 - (\partial_t K(t, x_1) + \dot\gamma(t) z) \cdot \hat b(t, x^{\text{mir}}_t)\right) dt. \tag{3.26}$$

S217 · Original / 原文 · PDF p.28 · S3.SS3.p3.3

In the expression above, x^{mir}_t is given by (3.24) and the expectation $\mathbb{E}$ is taken independently over $x_1 \sim \rho_1$ and $z \sim \mathsf{N}(0, Id)$. Similarly, the score is given by

上式中 x^{mir}_t 由式（3.24）给出，$\mathbb{E}$ 对相互独立的 $x_1 \sim \rho_1$ 与 $z \sim \mathsf{N}(0, \mathrm{Id})$ 取期望。类似地，得分为

$$s(t,x) = -\gamma^{-1}(t)\eta_z(t,x), \qquad \eta_z(t,x) = \mathbb{E}(z | x^{\text{mir}}_t = x), \tag{3.27}$$

S218 · Original / 原文 · PDF p.28 · S3.SS3.p3.4

which are the unique minimizers of the objective functions

它们是下列目标函数的唯一极小化解：

$$\mathcal{L}_s[\hat s] = \int_0^1 \mathbb{E}\left(\tfrac{1}{2}|\hat s(t, x^{\text{mir}}_t)|^2 + \gamma^{-1}(t) z \cdot \hat s(t, x^{\text{mir}}_t)\right) dt. \tag{3.28}$$

$$\mathcal{L}_{\eta_z}[\hat\eta_z] = \int_0^1 \mathbb{E}\left(\tfrac{1}{2}|\hat\eta_z(t, x^{\text{mir}}_t)|^2 - z \cdot \hat\eta_z(t, x^{\text{mir}}_t)\right) dt. \tag{3.29}$$

S219 · Original / 原文 · PDF p.28 · S3.SS3.p3.5

Moreover, we can weaken Assumption 5 to the following requirement:

此外，可以把假设 5 放宽为下列要求：

Assumption 36.假设 36

S220 · Original / 原文 · PDF p.28 · Thmtheorem36.p1.2

The density $\rho_1 \in C^2(\mathbb{R}^d)$ satisfies $\rho_1(x) > 0$ for all $x \in \mathbb{R}^d$ and

密度 $\rho_1 \in C^2(\mathbb{R}^d)$ 对所有 $x \in \mathbb{R}^d$ 均满足 $\rho_1(x) > 0$，并且

$$\int_{\mathbb{R}^d} |\nabla \log \rho_1(x)|^2 \rho_1(x) dx < \infty. \tag{3.30}$$

S221 · Original / 原文 · PDF p.28 · Thmtheorem36.p1.3

The function K satisfies

函数 K 满足

$$\exists C_1 < \infty : \qquad |\partial_t K(t, x_1)| \le C_1 |x_1| \qquad \text{for all} \quad (t, x_1) \in [0,1] \times \mathbb{R}^d, \tag{3.31}$$

S222 · Original / 原文 · PDF p.28 · Thmtheorem36.p1.4

and

以及

$$\exists M_1, M_2 < \infty \ : \ \mathbb{E}\big[|\partial_t K(t, x_1)|^4\big] \le M_1; \quad \mathbb{E}\big[|\partial_t^2 K(t, x_1)|^2\big] \le M_2, \quad \text{for all} \quad t \in [0,1], \tag{3.32}$$

S223 · Original / 原文 · PDF p.28 · Thmtheorem36.p1.5

where the expectation is taken over $x_1 \sim \rho_1$.

其中期望按 $x_1 \sim \rho_1$ 计算。

Remark 37.说明 37

S224 · Original / 原文 · PDF p.28 · Thmtheorem37.p1.1

Interestingly, if we take $K(t, x_1) = x_1$, then $\partial_t K(t, x_1) = 0$, and the velocity field defined in (3.25) is completely defined by the denoiser η_z

有趣的是，若取 $K(t, x_1) = x_1$，则 $\partial_t K(t, x_1) = 0$，式（3.25）中的速度场完全由去噪器 η_z 决定：

$$b(t, x) = \dot{\gamma}(t) \eta_z(t, x) \tag{3.33}$$

S225 · Original / 原文 · PDF p.28 · Thmtheorem37.p1.2

Since the score s also depends on η_z, this denoiser is the only quantity that needs to be learned.

由于得分 s 也取决于 η_z，只需要学习这一个去噪器。

Remark 38.说明 38

S226 · Original / 原文 · PDF p.28 · Thmtheorem38.p1.1

If ρ_1 is only accessible via empirical samples, mirror interpolants do not enable calculation of the functional form of ρ_1. A notable exception is if we set $K(t, x_1) = 0$ for $t \in [t_1, t_2]$ with $0 < t_1 \le t_2 < 1$: in that case, $x_t^{\text{mir}} = \gamma(t) z \sim \gamma(t) \mathsf{N}(0, Id)$ for

$t \in [t_1, t_2]$, which gives us a reference density for comparison. In this setup, mirror interpolants essentially reduce to two one-sided interpolants glued together (with the second one time-reversed), or in fact a regular stochastic interpolant when $\rho_0 = \rho_1$ and we set $I(t, x_0, x_1) = 0$ for $t \in [t_1, t_2]$.

若只能通过经验样本访问 ρ_1，通常无法用镜像插值计算 ρ_1 的函数表达式。一个值得注意的例外是：取 $0 < t_1 \leq t_2 < 1$，并在 $t \in [t_1, t_2]$ 上令 $K(t, x_1) = 0$。于是该区间内 $x_t^{\text{mir}} = \gamma(t)z$，服从标准正态随机变量经 $\gamma(t)$ 缩放后的分布，从而有一个可供比较的参考密度。在这一设置中，镜像插值实质上是两个单侧插值的拼接，第二段作时间反演；当 $\rho_0 = \rho_1$ 且在该区间取 $I(t, x_0, x_1) = 0$ 时，也可以将其视为普通的随机插值。

3.4 Stochastic interpolants and Schrödinger bridges

3.4 随机插值与 Schrödinger 桥

S227 · Original / 原文 · PDF p.28 · S3.SS4.p1.3

The stochastic interpolant framework can also be used to solve the Schrödinger bridge problem. For background material on this problem, we refer the reader to *Léonard (2014)* and the references therein. Consistent with the overall viewpoint of this paper, we consider the hydrodynamic formulation of the Schrödinger bridge problem, in which the goal is to obtain a pair (ρ, u), that solves the following optimization problem for a fixed $\epsilon > 0$

随机插值框架也可用来求解薛定谔桥问题。背景资料可参见 Léonard（2014）及其参考文献。与全文的视角一致，这里采用薛定谔桥问题的流体动力学表述：对固定的 $\epsilon > 0$，目标是找到求解下列优化问题的一对函数 (ρ, u)：

$$\begin{aligned} &\min_{\hat{u}, \hat{\rho}} \int_0^1 \int_{\mathbb{R}^d} |\hat{u}(t,x)|^2 \hat{\rho}(t,x) dx dt \\ &\text{subject to:} \partial_t \hat{\rho} + \nabla \cdot (\hat{u}\hat{\rho}) = \epsilon \Delta \hat{\rho}, \quad \hat{\rho}(0) = \rho_0 \quad \hat{\rho}(1) = \rho_1 \end{aligned} \tag{3.34}$$

S228 · Original / 原文 · PDF p.29 · S3.SS4.p1.4

Under our assumptions on ρ_0 and ρ_1 listed in Assumption 5, it is known (see e.g. Proposition 4.1 in *Léonard (2014)*) that (3.34) has a unique minimizer $(\rho, u = \nabla\lambda)$, with (ρ, λ) classical solutions of the Euler-Lagrange equations:

若 ρ_0 与 ρ_1 满足假设 5 所列条件，则已知式（3.34）存在唯一极小化解 $(\rho, u = \nabla\lambda)$，其中 (ρ, λ) 是以下 Euler–Lagrange 方程的经典解（例如，见 Léonard，2014，命题 4.1）：

$$\begin{aligned} &\partial_t \rho + \nabla \cdot (\nabla\lambda \rho) = \epsilon \Delta \rho, \quad \rho(0) = \rho_0, \quad \rho(1) = \rho_1, \\ &\partial_t \lambda + \tfrac{1}{2}|\nabla\lambda|^2 = -\epsilon \Delta \lambda. \end{aligned} \tag{3.35}$$

S229 · Original / 原文 · PDF p.29 · S3.SS4.p1.5

To proceed we will make the additional assumption that the solution ρ to (3.35) can be reversibly mapped to a standard Gaussian:

下面还需假设式（3.35）的解 ρ 可以通过可逆映射与标准高斯分布相互转换：

Assumption 39.假设 39

S230 · Original / 原文 · PDF p.29 · Thmtheorem39.p1.1

There exists a reversible map $T : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ with $T, T^{-1} \in C^1([0,1], (C^d(\mathbb{R}^d))^d)$ such that:

存在可逆映射 $T : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$，其中 $T, T^{-1} \in C^1([0,1], (C^d(\mathbb{R}^d))^d)$，使得：

$$\forall t \in [0,1] \quad : \quad z \sim \mathsf{N}(0, Id) \;\Rightarrow\; T(t,z) \sim \rho(t); \quad x_t \sim \rho(t) \;\Rightarrow\; T^{-1}(t, x_t) \sim \mathsf{N}(0, Id), \tag{3.36}$$

S231 · Original / 原文 · PDF p.29 · Thmtheorem39.p1.2

where ρ is the solution to (3.35).

其中 ρ 是式（3.35）的解。

S232 · Original / 原文 · PDF p.29 · S3.SS4.p2.1

We stress that the actual form of the map T is not important for the arguments below. Assumption 39 can be used to show the existence of a stochastic interpolant whose density solves (3.35):

强调一下，以下论证不需要知道映射 T 的具体形式。利用假设 39 可以证明，存在一个随机插值，其密度恰好求解式（3.35）：

Lemma 40.引理 40

S233 · Original / 原文 · PDF p.29 · Thmtheorem40.p1.1

If Assumption (39) holds, then the solution $\rho(t)$ to (3.35) is the density of the stochastic interpolant

若假设 39 成立，则式（3.35）的解 $\rho(t)$ 是下列随机插值的密度：

$$x_t = T(t, \alpha(t)T^{-1}(0, x_0) + \beta(t)T^{-1}(1, x_1)) + \gamma(t)z, \tag{3.37}$$

S234 · Original / 原文 · PDF p.29 · Thmtheorem40.p1.2

as long as $\alpha^2(t) + \beta^2(t) + \gamma^2(t) = 1$.

只要 $\alpha^2(t) + \beta^2(t) + \gamma^2(t) = 1$ 即可。

S235 · Original / 原文 · PDF p.29 · S3.SS4.p3.1

The proof is given in Appendix B.6: (3.37) corresponds to choosing $I(t, x_0, x_1) = T(t, \alpha(t)T^{-1}(t, x_0) + \beta(t)T^{-1}(t, x_1))$ in (2.1). With the help of Lemma 40, we can establish the following result, which shows how to optimize over the function I to solve the problem (3.34)

证明见附录 B.6：式（3.37）相当于在式（2.1）中选取 $I(t, x_0, x_1) = T(t, \alpha(t)T^{-1}(t, x_0) + \beta(t)T^{-1}(t, x_1))$。借助引理 40 可以得到下面的结果，它说明如何通过优化函数 I 来求解问题（3.34）。

Theorem 41.定理 41

S236 · Original / 原文 · PDF p.29 · Thmtheorem41.p1.1

Pick some $\gamma : [0, 1] \to [0, 1)$ such that $\gamma(0) = \gamma(1) = 0$, $\gamma(t) > 0$ for $t \in (0, 1)$, $\gamma \in C^2((0, 1))$ and $\gamma^2 \in C^1([0, 1])$, and let $\hat{x}_t = \hat{I}(t, x_0, x_1) + \gamma(t)z$, with $x_0 \sim \rho_0$, $x_1 \sim \rho_1$, and $z \sim \mathsf{N}(0, Id)$ all independent. Consider the max-min problem over $\hat{I} \in C^1([0, 1], (C^1(\mathbb{R}^d \times \mathbb{R}^d))^d)$ and $\hat{u} \in C^0([0, 1], (C^1(\mathbb{R}^d))^d)$:

取 $\gamma : [0, 1] \to [0, 1)$，使得 $\gamma(0) = \gamma(1) = 0$、对 $t \in (0, 1)$ 有 $\gamma(t) > 0$，且 $\gamma \in C^2((0, 1))$、$\gamma^2 \in C^1([0, 1])$。令 $\hat{x}_t = \hat{I}(t, x_0, x_1) + \gamma(t)z$，其中 $x_0 \sim \rho_0$、$x_1 \sim \rho_1$、$z \sim \mathsf{N}(0, \mathrm{Id})$ 相互独立。考虑关于 $\hat{I} \in C^1([0, 1], (C^1(\mathbb{R}^d \times \mathbb{R}^d))^d)$ 和 $\hat{u} \in C^0([0, 1], (C^1(\mathbb{R}^d))^d)$ 的如下极大极小问题：

$$\max_{\hat{I}} \min_{\hat{u}} \int_0^1 \mathbb{E}\left(\tfrac{1}{2}|\hat{u}(t, \hat{x}_t)|^2 - \left(\partial_t \hat{I}(t, x_0, x_1) + (\dot{\gamma}(t) - \epsilon\gamma^{-1}(t))z\right) \cdot \hat{u}(t, \hat{x}_t)\right) dt. \tag{3.38}$$

S237 · Original / 原文 · PDF p.30 · Thmtheorem41.p1.2

> If Assumption 39 holds, then all the optimizers (I, u) of (3.38) are such that the density of the associated $x_t = I(t, x_0, x_1) + \gamma(t)z$ is the solution ρ to (3.35). Moreover, $u = \nabla\lambda$, with λ the solution to (3.35).
>
> 若假设 39 成立，式（3.38）的任一最优解 (I, u) 都具有如下性质：对应的随机插值 $x_t = I(t, x_0, x_1) + \gamma(t)z$ 的密度即式（3.35）的解 ρ。此外，$u = \nabla\lambda$，其中 λ 也是式（3.35）的解。

S238 · Original / 原文 · PDF p.30 · S3.SS4.p4.1

The proof is also given in Appendix B.6. Note that if we fix $\hat{I}$, the velocity u minimizing this objective is the forward drift b_{F} defined in (2.21). Note also that if we set $\epsilon \to 0$, the minimizing velocity field is b as defined in (2.10), and the max-min problem formally reduces to solving the optimal transport problem. In this case, Assumption 39 becomes more stringent, as we need to assume that that system (3.35) with $\epsilon = 0$ (i.e. in the absence of the diffusive terms) has a classical solution. Theorem (41) gives a practical route towards solving the Schrödinger bridge problem with stochastic interpolants, and we leave the numerical investigation of this formulation to future work.

证明同样见附录 B.6。若固定 $\hat{I}$，使该目标最小的速度 u 就是式（2.21）中的前向漂移 b_{F}。另一方面，当 $\epsilon \to 0$ 时，使目标最小的速度场变为式（2.10）定义的 b，极大极小问题在形式上化为最优传输问题。此时假设 39 更强，因为必须假定式（3.35）在 $\epsilon = 0$（即没有扩散项）时仍有经典解。定理 41 给出了利用随机插值求解薛定谔桥问题的可行途径，这一表述的数值研究留待后续工作。

4 Spatially linear interpolants

4 空间线性插值

S239 · Original / 原文 · PDF p.30 · S4.p1.2

In this section, we study the stochastic interpolants that are obtained when we specialize the function I used in (2.1) to be linear in both x_0 and x_1, i.e. we consider

本节考察式（2.1）中的函数 I 对 x_0 和 x_1 都取线性形式时得到的随机插值，即考虑

$$x_t^{\mathsf{lin}} = \alpha(t)x_0 + \beta(t)x_1 + \gamma(t)z, \tag{4.1}$$

S240 · Original / 原文 · PDF p.30 · S4.p1.3

where $(x_0, x_1) \sim \nu$ and $z \sim \mathsf{N}(0, Id)$ with $(x_0, x_1) \perp z$, and $\alpha, \beta, \gamma^2 \in C^2([0,1])$ satisfy the conditions

其中 $(x_0, x_1) \sim \nu$，$z \sim \mathsf{N}(0, \mathrm{Id})$，且 (x_0, x_1) 与 z 独立；函数 $\alpha, \beta, \gamma^2 \in C^2([0,1])$ 满足下列条件：

$$\alpha(0) = \beta(1) = 1; \quad \alpha(1) = \beta(0) = \gamma(0) = \gamma(1) = 0; \quad \forall t \in (0,1) \ : \ \gamma(t) > 0. \tag{4.2}$$

S241 · Original / 原文 · PDF p.30 · S4.p1.4

Despite its simplicity, this setup offers significant design flexibility. The discussion highlights how the presence of the latent variable $\gamma(t)z$ can simplify the structure of the intermediate density $\rho(t)$. Since our ultimate aim is to investigate the properties of practical generative models built upon ODEs or SDEs, we will also study the effect of time-dependent diffusion coefficient $\epsilon(t)$, which controls the amplitude of the noise in a generative SDE. Throughout, to build intuition, we choose ρ_0 and ρ_1 to be Gaussian mixture densities, for which the drift coefficients can be computed analytically (see Appendix A). This enables us to visualize the effect of each choice on the resulting generative models.

尽管设置简单，设计仍颇为灵活。下面着重讨论潜变量项 $\gamma(t)z$ 怎样简化中间密度 $\rho(t)$ 的结构。我们的最终目的是研究基于常微分方程或随机微分方程的实用生成模型，因此也会考察含时扩散系数 $\epsilon(t)$ 的作用：它控制生成随机微分方程中的噪声幅度。为建立直观认识，下文均取 ρ_0、ρ_1 为高斯混合密度，此时漂移系数可以解析求出（见附录 A），因而可以直观显示每种选择对生成模型的影响。

4.1 Factorization of the velocity field

4.1 速度场分解

S242 · Original / 原文 · PDF p.30 · S4.SS1.p1.2

When the stochastic interpolant is of the form (4.1), the velocity b and the score s defined in (2.10) and (2.14) can both be expressed in terms of the following three conditional expectations (the third is the denoiser already defined in (2.17)):

若随机插值具有式（4.1）的形式，则式（2.10）和（2.14）中的速度 b 与得分 s 都可用以下三个条件期望表示，其中第三个就是式（2.17）已经定义的去噪器：

$$\eta_0(t,x) = \mathbb{E}(x_0 | x_t^{\mathsf{lin}} = x), \quad \eta_1(t,x) = \mathbb{E}(x_1 | x_t^{\mathsf{lin}} = x), \quad \eta_z(t,x) = \mathbb{E}(z | x_t^{\mathsf{lin}} = x). \tag{4.3}$$

S243 · Original / 原文 · PDF p.30 · S4.SS1.p1.3

Specifically, we have

具体地，有

$$b(t,x) = \dot\alpha(t)\eta_0(t,x) + \dot\beta(t)\eta_1(t,x) + \dot\gamma(t)\eta_z(t,x), \qquad s(t,x) = -\gamma^{-1}(t)\eta_z(t,x). \tag{4.4}$$

S244 · Original / 原文 · PDF p.30 · S4.SS1.p1.4

The second relation above holds for $t \in (0,1)$ (i.e. when $\gamma(t) \neq 0$). Moreover, because $\mathbb{E}(x_t|x_t = x) = x$ by definition, the functions η_0, η_1, and η_z satisfy

上面第二个等式在 $t \in (0,1)$ 时成立，即要求 $\gamma(t) \neq 0$。此外，根据定义 $\mathbb{E}(x_t \mid x_t = x) = x$，函数 η_0、η_1 和 η_z 因而满足

$$\forall (t,x) \in [0,1] \times \mathbb{R}^d \quad : \quad \alpha(t)\eta_0(t,x) + \beta(t)\eta_1(t,x) + \gamma(t)\eta_z(t,x) = x. \tag{4.5}$$

S245 · Original / 原文 · PDF p.31 · S4.SS1.p1.5

This enables us to reduce computational expense: given two of the η's, the third can always be calculated via (4.5). Finally, it is easy to see that the functions η_0, η_1, and η_z are the unique minimizers of the objectives

这样可以降低计算开销：给定三个 η 中的任意两个，即可由式（4.5）求出第三个。最后不难看出，函数 η_0、η_1 与 η_z 是下列各目标的唯一极小化解：

$$\begin{aligned}
\mathcal{L}_{\eta_0}(\hat{\eta}_0) &= \int_0^1 \mathbb{E}\left[\tfrac{1}{2}|\hat{\eta}_0(t, x_t^{\text{lin}})|^2 - x_0 \cdot \hat{\eta}_0(t, x_t^{\text{lin}})\right] dt, \\
\mathcal{L}_{\eta_1}(\hat{\eta}_1) &= \int_0^1 \mathbb{E}\left[\tfrac{1}{2}|\hat{\eta}_1(t, x_t^{\text{lin}})|^2 - x_1 \cdot \hat{\eta}_1(t, x_t^{\text{lin}})\right] dt, \\
\mathcal{L}_{\eta_z}(\hat{\eta}_z) &= \int_0^1 \mathbb{E}\left[\tfrac{1}{2}|\hat{\eta}_z(t, x_t^{\text{lin}})|^2 - z \cdot \hat{\eta}_z(t, x_t^{\text{lin}})\right] dt,
\end{aligned} \tag{4.6}$$

S246 · Original / 原文 · PDF p.31 · S4.SS1.p1.6

where x_t^{lin} is defined in (4.1) and the expectation is taken independently over $(x_0, x_1) \sim \nu$ and $z \sim \mathsf{N}(0, Id)$.

其中 x_t^{lin} 由式（4.1）定义，期望分别对相互独立的 $(x_0, x_1) \sim \nu$ 和 $z \sim \mathsf{N}(0, \mathrm{Id})$ 计算。

Stochastic Interpolant		$\alpha(t)$	$\beta(t)$	$\gamma(t)$
Arbitrary ρ_0 (two-sided)	linear	$1-t$	t	$\sqrt{at(1-t)}$
	trig	$\cos\frac{\pi}{2}t$	$\sin\frac{\pi}{2}t$	$\sqrt{at(1-t)}$
	enc-dec	$\cos^2(\pi t)1_{[0,\frac{1}{2})}(t)$	$\cos^2(\pi t)1_{(\frac{1}{2},1]}(t)$	$\sin^2(\pi t)$
Gaussian ρ_0 (one-sided)	linear	$1-t$	t	0
	trig	$\cos\frac{\pi}{2}t$	$\sin\frac{\pi}{2}t$	0
	SBDM (VP)	$\sqrt{1-t^2}$	t	0
Mirror		0	1	$\sqrt{at(1-t)}$

Table 1: Spatially linear interpolants. A table suggesting various linear interpolants. In general, this paper describes methods for arbitrary ρ_0 and ρ_1. In Section 4.4, we detail linear interpolants for one-sided generation, where ρ_0 is a Gaussian and the latent variable z can be absorbed into x_0. Later, in Section 5.1, we discuss how to recast score-based diffusion models (SBDM) as linear one-sided interpolants, which leads to the expressions given in the table when using a variance-preserving parameterization. Linear mirror interpolants, where $\rho_0 = \rho_1$ are equal, are defined by (3.24) with the choice $K(t, x_1) = \alpha(t)x_1$.

表 1：空间线性插值的设计建议。一般框架允许任意 ρ₀、ρ₁。第 4.4 节的单侧插值以高斯为基分布，潜变量 z 可并入 x₀；第 5.1 节把 SBDM 改写为单侧插值，表中采用方差保持参数化。镜像情形 ρ₀=ρ₁，使用式（3.24）且 K(t,x₁)=α(t)x₁。

4.2 Some specific design choices

4.2 具体设计选择

S247 · Original / 原文 · PDF p.31 · S4.SS2.p1.1

It is useful to assume that both ρ_0 and ρ_1 have been scaled to have zero mean and identity covariance (which can be achieved in practice, for example, by an affine transformation of the data). In this case, the time-dependent mean and covariance of (4.1)

are given by

一个有用的假设是，ρ_0 和 ρ_1 都已作尺度变换，使其具有零均值和单位协方差；实践中例如可以对数据作仿射变换来实现。此时，式（4.1）的含时均值与协方差为

$$\mathbb{E}[x_t^{\text{lin}}] = 0, \qquad \mathbb{E}[x_t^{\text{lin}}(x_t^{\text{lin}})^{\mathsf{T}}] = (\alpha^2(t) + \beta^2(t) + \gamma^2(t))Id. \tag{4.7}$$

S248 · Original / 原文 · PDF p.32 · S4.SS2.p1.2

Preserving the identity covariance at all times therefore leads to the constraint

因此，若要求所有时刻都保持单位协方差，便得到约束

$$\forall t \in [0,1] \quad : \quad \alpha^2(t) + \beta^2(t) + \gamma^2(t) = 1. \tag{4.8}$$

S249 · Original / 原文 · PDF p.32 · S4.SS2.p1.3

This choice is also sensible if ρ_0 and ρ_1 have covariances that are not the identity but are on a similar scale. In this case we no longer need to enforce (4.8) exactly, and could, for example, take three functions whose sum of squares is of order one. For definiteness, in the sequel we discuss possible choices that satisfy (4.8) exactly, with the understanding that the corresponding functions α, β, and γ could all be slightly modified without significantly affecting the conclusions.

即使 ρ_0、ρ_1 的协方差并非单位阵，只要两者尺度接近，这个选择仍然合理。此时无需严格满足式（4.8）；例如可取三个函数，使其平方和维持在一的量级。为把讨论具体化，下文只讨论精确满足式（4.8）的选择；相应的 α、β、γ 即便略作修改，也不会实质改变结论。

Linear and trigonometric α and β.

线性与三角 α、β

S250 · Original / 原文 · PDF p.32 · S4.SS2.SSS0.Px1.p1.1

One way to ensure that (4.8) holds while maintaining the influence of ρ_0 and ρ_1 everywhere on $[0,1]$ except at the endpoints is to choose

要使式（4.8）成立，同时让 ρ_0 与 ρ_1 的影响除两个端点外贯穿整个 $[0,1]$，一种做法是取

$$\alpha(t) = t, \qquad \beta(t) = 1 - t, \qquad \gamma(t) = \sqrt{2t(1-t)}. \tag{4.9}$$

S251 · Original / 原文 · PDF p.32 · S4.SS2.SSS0.Px1.p1.2

This choice was advocated in *Liu et al. (2023b)*, without the inclusion of the latent variable ($\gamma = 0$). Another possibility that gives more leeway is to pick any $\gamma : [0,1] \to [0,1]$ and set

Liu 等（2023b）提倡采用这一选择，但不加入潜变量（$\gamma = 0$）。另一种自由度更大的方式是任取 $\gamma : [0,1] \to [0,1]$，并令

$$\alpha(t) = \sqrt{1-\gamma^2(t)}\cos(\tfrac{1}{2}\pi t), \qquad \beta(t) = \sqrt{1-\gamma^2(t)}\sin(\tfrac{1}{2}\pi t). \tag{4.10}$$

S252 · Original / 原文 · PDF p.32 · S4.SS2.SSS0.Px1.p1.3

With $\gamma = 0$, this was the choice preferred in *Albergo and Vanden-Eijnden (2023)*. The PDF $\rho(t)$ obtained with the choices (4.9) and (4.10) when ρ_0 and ρ_1 are both Gaussian mixture densities are shown in Figure 4. As this example shows, when ρ_0 and ρ_1 have distinct complex features, these would be duplicated in $\rho(t)$ at intermediate times if not for the smoothing effect of the latent variable; this behavior is seen in Figure 4, where it is most prominent in the first row with $\gamma(t) = 0$. From a statistical

learning perspective, eliminating the formation of spurious features will simplify the estimation of the velocity field b, which becomes smoother as the formation of such features is suppressed.

当 $\gamma = 0$ 时，这正是 Albergo 与 Vanden-Eijnden（2023）优先采用的选择。ρ_0 和 ρ_1 均为高斯混合密度时，选择（4.9）和（4.10）得到的概率密度 $\rho(t)$ 见图 4。这个例子说明，若 ρ_0 与 ρ_1 具有各不相同的复杂特征，在没有潜变量平滑作用时，这些特征会在中间时刻的 $\rho(t)$ 中重复出现。图 4 展示了这一行为，其中第一行 $\gamma(t) = 0$ 时最明显。从统计学习角度看，消除虚假特征的形成能够简化速度场 b 的估计；抑制这些特征形成时，b 也变得更平滑。

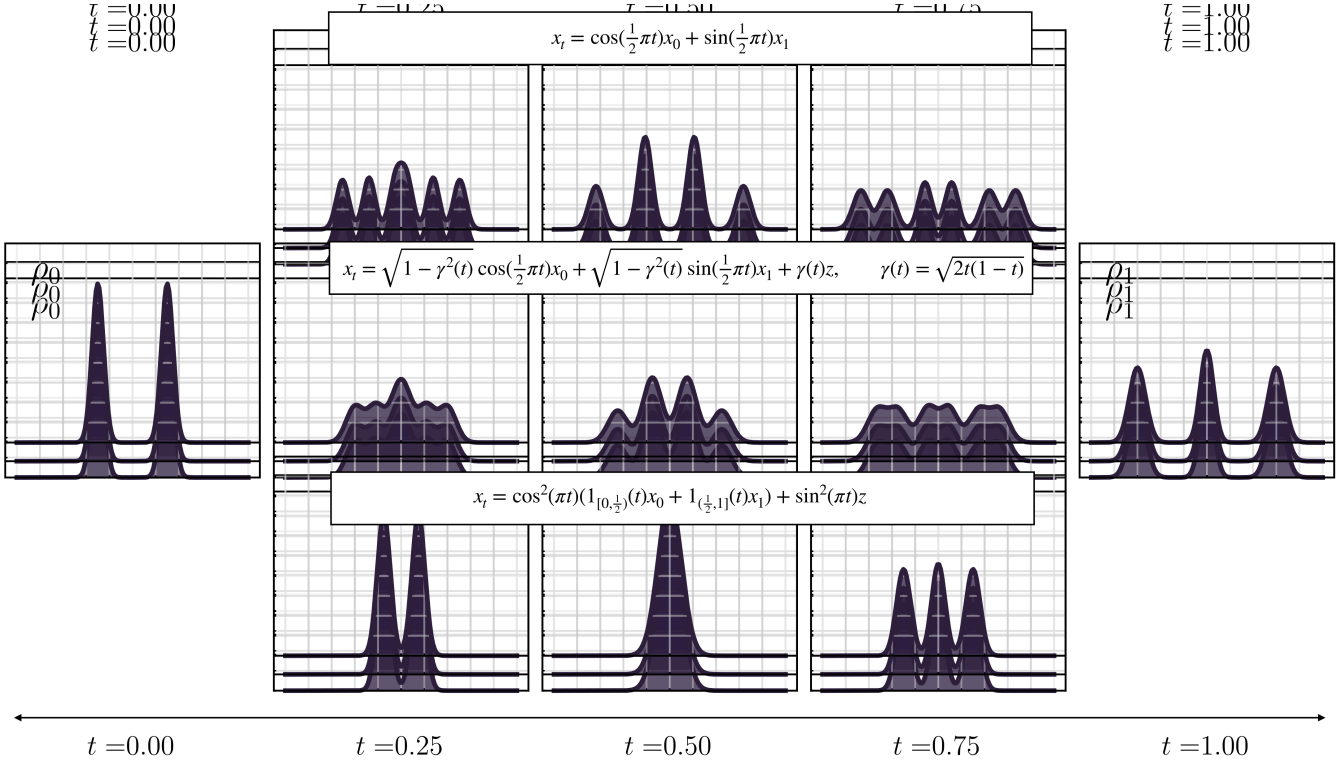

Figure 4: The effect of $\gamma(t)z$ on $\rho(t)$. A visualization of how the choice of $\gamma(t)$ changes the density $\rho(t)$ of $x_t^{\text{lin}} = \alpha(t)x_0 + \beta(t)x_1 + \gamma(t)z$ when ρ_0 and ρ_1 are Gaussian mixture densities with two modes and three modes, respectively. The first row depicts $\gamma(t) = 0$, which reduces to the stochastic interpolant developed in *Albergo and Vanden-Eijnden (2023)*. This case forms a valid transport between ρ_0 and ρ_1, but produces spurious intermediate modes in $\rho(t)$. The second row depicts the choice of $\gamma(t) = \sqrt{2t(1-t)}$. In this case, the spurious modes are partially damped by the addition of the latent variable, leading to a simpler $\rho(t)$. The final row shows the Gaussian encoding-decoding, which smoothly encodes ρ_0 into a standard normal distribution on the time interval $[0, 1/2)$, which is then decoded into ρ_1 on the interval $(1/2, 1]$. In this case, no intermediate modes form in $\rho(t)$: the two modes in ρ_0 collide to form $\mathsf{N}(0, 1)$ at $t = \frac{1}{2}$, which then spreads into the three modes of ρ_1. A visualization of individual sample trajectories from deterministic and stochastic generative models based on ODEs and SDEs whose solutions have density $\rho(t)$ can be seen in Figure 5.

图 4：γ(t)z 对 ρ(t) 的影响。两端分别为双峰、三峰高斯混合，插值为 $\alpha x_0 + \beta x_1 + \gamma z$。首行 γ=0，虽是合法输运，但中间产生伪模态；第二行 $\gamma = \sqrt{2t(1-t)}$，潜变量抑制部分伪模态，使密度更简单。末行为高斯编码—解码：前半程将 ρ_0 平滑编码到标准正态，后半程解码为 ρ_1；两峰在 t=1/2 合成 N(0,1)，再散为三峰，没有额外中间模态。对应 ODE/SDE 的单样本轨迹见图 5。

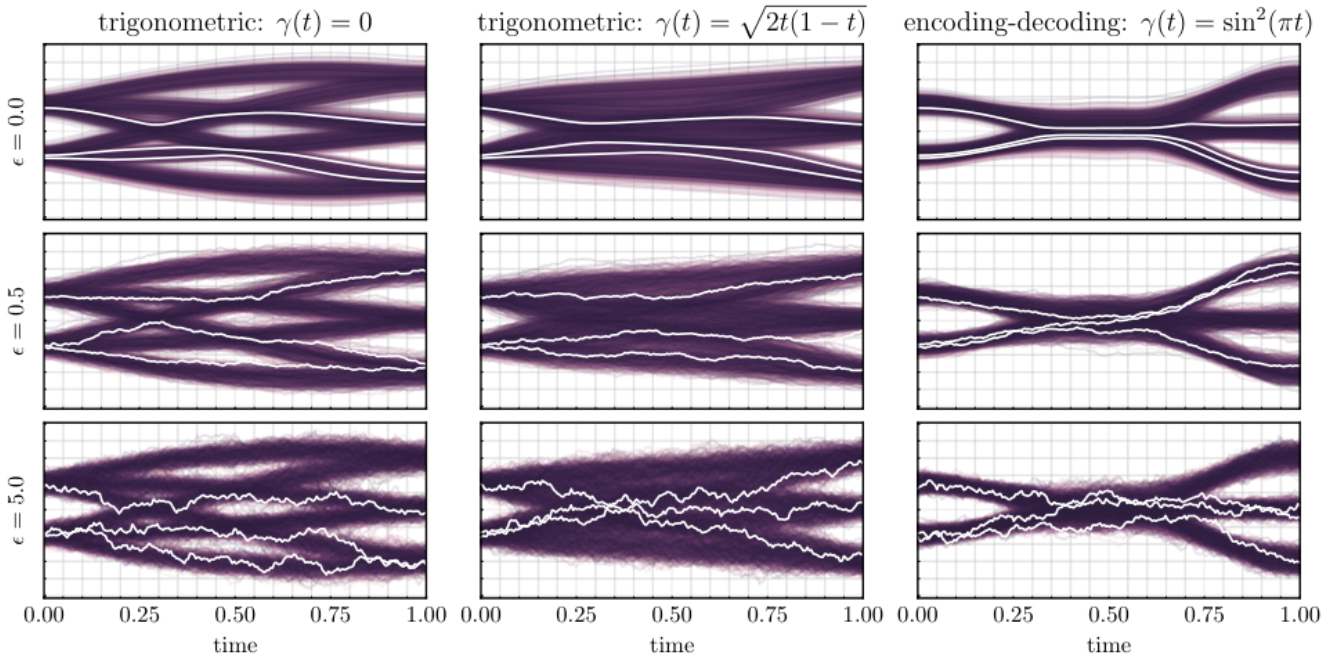

Figure 5: The effect of ϵ on sample trajectories. A visualization of how the choice of ϵ affects the sample trajectories obtained by solving the ODE (2.32) or the forward SDE (2.33). The set-up is the same as in Figure 4: ρ_0 and ρ_1 are taken to be the same Gaussian mixture densities, and the analytical expressions for b and s are used. In the three panels in each column the value of γ is the same, and each panel shows trajectories with different ϵ. Three specific trajectories from the same three initial conditions drawn from ρ_0 are also highlighted in white in every panel. As ϵ increases but γ stays the same, the density $\rho(t)$ is unchanged, but the individual trajectories become increasingly stochastic. While all choices are equivalent with exact b and s, Theorem 23 shows that nonzero values of ϵ provide control on the likelihood in terms of the error in b and s when they are approximate.

图 5：ε 对样本路径的影响。设置与图 4 相同，b、s 使用解析式。每列三幅图的 γ 相同、ε 不同，白色突出相同三个初值的轨迹。固定 γ、增大 ε 时，ρ(t) 不变，但单条轨迹更随机。精确 b、s 下各选择等价；近似时，定理 23 说明非零 ε 可用 b、s 误差控制似然。

Gaussian encoding-decoding.

高斯编码—解码

S253 · Original / 原文 · PDF p.32 · S4.SS2.SSS0.Px2.p1.1

A useful limiting case is to devolve the data from ρ_0 completely into noise by the halfway point $t = \frac{1}{2}$ and to reconstruct ρ_1 completely from noise starting from $t = \frac{1}{2}$. One choice that allows us to do so while satisfying (4.8) is

一个有用的极限情形是：在到达中点 $t = \frac{1}{2}$ 前，把来自 ρ_0 的数据完全退化为噪声；再从 $t = \frac{1}{2}$ 开始，完全由噪声重建 ρ_1。下面这一选择既能实现该过程，又满足（4.8）：

$$\alpha(t) = \cos^2(\pi t)1_{[0,\frac{1}{2})}(t), \qquad \beta(t) = \cos^2(\pi t)1_{(\frac{1}{2},1]}(t), \qquad \gamma(t) = \sin^2(\pi t), \tag{4.11}$$

S254 · Original / 原文 · PDF p.32 · S4.SS2.SSS0.Px2.p1.2

where $1_A(t)$ is the indicator function of A, i.e. $1_A(t) = 1$ if $t \in A$ and $1_A(t) = 0$ otherwise. With this choice, it is easy to see that $x_{t=\frac{1}{2}} = \gamma(\frac{1}{2})z \sim \mathsf{N}(0, \gamma^2(\frac{1}{2}))$, which seamlessly glues together two interpolants: one between ρ_0 and a standard Gaussian, and one between a standard Gaussian and ρ_1.

其中 $1_A(t)$ 是集合 A 的示性函数，即 $t \in A$ 时 $1_A(t) = 1$，否则 $1_A(t) = 0$。采用该选择时容易看出，$x_{t=\frac{1}{2}} = \gamma(\frac{1}{2})z \sim \mathsf{N}(0, \gamma^2(\frac{1}{2}))$；这把两个插值无缝连接起来：一个连接 ρ_0 与标准高斯分布，另一个连接标准高斯分布与 ρ_1。

S255 · Original / 原文 · PDF p.32 · S4.SS2.SSS0.Px2.p2.1

Even though the choice (4.11) encodes ρ_0 into pure noise on the interval $[0, \frac{1}{2}]$, which is then decoded into ρ_1 on the interval $[\frac{1}{2}, 1]$ (and vice-versa when proceeding backwards in time), the resulting velocity b still defines a single continuity equation that

maps ρ_0 to ρ_1 on $[0,1]$. This is most clearly seen at the level of the probability flow (2.32), since its solution X_t is a bijection between the initial and final conditions $X_{t=0}$ and $X_{t=1}$, but a similar pairing can also be observed in the solutions to the forward and backward SDEs (2.33) and (2.34), whose solutions at time $t = 1$ or $t = 0$ remain correlated with the initial or final condition used.

尽管选择（4.11）在区间 $[0, \frac{1}{2}]$ 上把 ρ_0 编码为纯噪声，再在 $[\frac{1}{2}, 1]$ 上将噪声解码成 ρ_1（沿时间反向进行时则相反），所得速度 b 仍定义了一个在整个 $[0,1]$ 上把 ρ_0 映射至 ρ_1 的单一连续性方程。这一点在概率流（2.32）层面最为清楚，因为其解 X_t 在初值 $X_{t=0}$ 与终值 $X_{t=1}$ 之间构成双射。不过，在正向与反向 SDE（2.33）、（2.34）的解中也可观察到类似配对：它们在 $t = 1$ 或 $t = 0$ 时的解，仍与所使用的初始或最终条件保持相关。

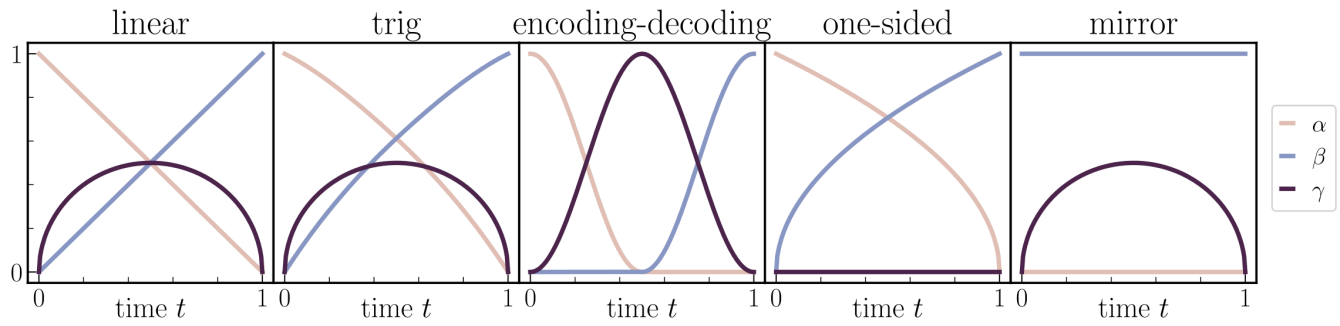

Figure 6: The functions $\alpha(t)$, $\beta(t)$, and $\gamma(t)$ for the linear (4.9), trigonometric (4.10) with $\gamma(t) = \sqrt{2t(1-t)}$, Gaussian encoding-decoding (4.11), one-sided (3.14), and mirror (3.24) interpolants.

图 6：不同插值的 α(t)、β(t)、γ(t)：线性（4.9）、带 γ=√[2t(1−t)] 的三角插值（4.10）、高斯编码—解码（4.11）、单侧（3.14）与镜像（3.24）。

S256 · Original / 原文 · PDF p.33 · S4.SS2.SSS0.Px2.p3.1

This allows for a more direct means of image-to-image translation with diffusions when compared to the recent approach described in *Su et al. (2023)*. The choice (4.11) is depicted in the final row of Figure 4, where no spurious modes form at all; individual sample trajectories of the deterministic and stochastic generative models based on ODEs and SDEs whose solutions have this $\rho(t)$ as density can be seen in the panels forming the third column in Figure 5. We note that the elimination of spurious intermediate modes can also be implemented by use of a data-adapted coupling $\nu(dx_0, dx_1)$, as considered in *Albergo et al. (2023)*.

与 Su 等（2023）近期描述的方法相比，这为利用扩散进行图像到图像转换提供了更直接的途径。图 4 最后一行展示了选择（4.11），其中完全不形成虚假模态。以该 $\rho(t)$ 为解的密度、基于 ODE 与 SDE 的确定性及随机生成模型，其单样本轨迹见图 5 第三列各面板。还应指出，也可使用适应数据的耦合 $\nu(dx_0, dx_1)$ 消除中间虚假模态，Albergo 等（2023）研究了这种方式。

S257 · Original / 原文 · PDF p.34 · S4.SS2.SSS0.Px2.p4.1

Unsurprisingly, it is necessary to have $\gamma(t) > 0$ for the choice (4.11): for $\gamma(t) = 0$, the density $\rho(t)$ collapses to a Dirac measure at $t = \frac{1}{2}$. This consideration highlights that the inclusion of the latent variable $\gamma(t)z$ matters even for the deterministic dynamics (2.32), and its presence is distinct from the stochasticity inherent to the SDEs (2.33) and (2.34).

对于选择（4.11），必须有 $\gamma(t) > 0$，这一点并不意外：若 $\gamma(t) = 0$，密度 $\rho(t)$ 会在 $t = \frac{1}{2}$ 坍缩成 Dirac 测度。这个考量强调，即使对确定性动力学（2.32），加入潜变量 $\gamma(t)z$ 也很重要；它的作用不同于 SDE（2.33）、（2.34）本身所包含的随机性。

4.3 Impact of the latent variable $\gamma(t)z$ and the diffusion coefficient $\epsilon(t)$

4.3 潜变量 γz 与扩散 ε 的影响

S258 · Original / 原文 · PDF p.34 · S4.SS3.p1.1

The stochastic interpolant framework enables us to discern the independent roles of the latent variable $\gamma(t)z$ and the diffusion coefficient $\epsilon(t)$ we use in a generative model. As shown in Theorem 6, the presence of the latent variable $\gamma(t)z$ for $\gamma \neq 0$

smooths both the density $\rho(t)$ and the velocity b defined in (2.10) spatially. This provides a computational advantage at sample generation time because it simplifies the required numerical integration of (2.32), (2.33), and (2.34). Intuitively, this is because the density $\rho(t)$ of x_t can be represented exactly as the density that would be obtained with $\gamma(t) = 0$ convolved with $\mathsf{N}(0, \gamma^2(t)Id)$ at each $t \in (0, 1)$. A comparison between the density $\rho(t)$ obtained with trigonometric interpolants with $\gamma(t) = 0$ and $\gamma(t) = \sqrt{2t(1-t)}$ can be seen in the first and second row of Figure 4.

随机插值框架使我们能区分潜变量 $\gamma(t)z$ 与生成模型中所用扩散系数 $\epsilon(t)$ 的独立作用。由定理 6，当 $\gamma \neq 0$ 时，潜变量 $\gamma(t)z$ 会在空间上同时平滑密度 $\rho(t)$ 和（2.10）定义的速度 b。这简化了对（2.32）、（2.33）、（2.34）所需的数值积分，因此在生成样本时带来计算优势。直观原因是：对每个 $t \in (0,1)$，x_t 的密度 $\rho(t)$ 都可精确表示为 $\gamma(t) = 0$ 时所得密度与 $\mathsf{N}(0, \gamma^2(t)Id)$ 的卷积。图 4 第一、二行比较了三角函数插值在 $\gamma(t) = 0$ 与 $\gamma(t) = \sqrt{2t(1-t)}$ 时的密度 $\rho(t)$。

S259 · Original / 原文 · PDF p.35 · S4.SS3.p2.1

By contrast, the diffusion coefficient $\epsilon(t)$ leaves the density $\rho(t)$ unchanged, and only affects the way we sample it. In particular, the probability flow ODE (2.32) results in a map that pushes every $X_{t=0} = x_0$ onto a single $X_{t=1} = x_1$ and vice-versa. The forward SDE (2.33) maps each $X^{\mathsf{F}}_{t=0} = x_0$ onto an ensemble $X^{\mathsf{F}}_{t=1}$ whose spread is controlled by the amplitude of $\epsilon(t)$ (and similarly for the reversed SDE (2.34) that maps each $X^{\mathsf{B}}_{t=1} = x_1$ onto an ensemble $X^{\mathsf{B}}_{t=0}$). This ensemble is not distributed according to ρ_1 for finite $\epsilon(t)$ – like with the ODE, we need to sample initial conditions from ρ_0 to get solutions at time $t = 1$ that sample ρ_1 – but its density converges towards ρ_1 as $\epsilon(t) \to \infty$. These features are illustrated in Figure 5.

相比之下，扩散系数 $\epsilon(t)$ 不改变密度 $\rho(t)$，只影响我们对该密度采样的方式。具体而言，概率流 ODE（2.32）给出一个映射，把每个 $X_{t=0} = x_0$ 推送到唯一的 $X_{t=1} = x_1$，反之亦然。正向 SDE（2.33）则把每个 $X^{\mathsf{F}}_{t=0} = x_0$ 映射成一组 $X^{\mathsf{F}}_{t=1}$，其离散程度由 $\epsilon(t)$ 的幅值控制；反向 SDE（2.34）也类似，把每个 $X^{\mathsf{B}}_{t=1} = x_1$ 映射成一组 $X^{\mathsf{B}}_{t=0}$。对有限的 $\epsilon(t)$，该样本集合并不服从 ρ_1——如同 ODE，要使 $t = 1$ 的解服从 ρ_1，必须从 ρ_0 采样初态——但其密度会在 $\epsilon(t) \to \infty$ 时趋向 ρ_1。图 5 展示了这些特点。

Remark 42.说明 42

S260 · Original / 原文 · PDF p.35 · Thmtheorem42.p1.2

Another potential advantage of including the latent variable $\gamma(t)z$ is its impact on the velocity b at the end points. Since $x_{t=0} = x_0$ and $x_{t=1} = x_1$, it is easy to see that the velocity b of the linear interpolant x_t defined in (4.1) satisfies

加入潜变量 $\gamma(t)z$ 的另一潜在优势，是它对端点处速度 b 的影响。由于 $x_{t=0} = x_0$ 且 $x_{t=1} = x_1$，容易看出，（4.1）定义的线性插值 x_t 的速度 b 满足

$$\begin{aligned} b(0,x) &= \dot\alpha(0)x + \dot\beta(0)\mathbb{E}[x_1|x_0 = x] - \lim_{t\to 0}\gamma(t)\dot\gamma(t)s_0(x), \\ b(1,x) &= \dot\alpha(1)\mathbb{E}[x_0|x_1 = x] + \dot\beta(1)x - \lim_{t\to 1}\gamma(t)\dot\gamma(t)s_1(x), \end{aligned} \tag{4.12}$$

S261 · Original / 原文 · PDF p.35 · Thmtheorem42.p1.3

where $s_0 = \nabla \log \rho_0$ and $s_1 = \nabla \log \rho_1$. If $\gamma \in C^2([0,1])$, because $\gamma(0) = \gamma(1) = 0$, the terms involving the scores s_0 and s_1 in these expressions vanish. Choosing $\gamma^2 \in C^1([0,1])$ but γ not differentiable at $t = 0$ or $t = 1$ leaves open the possibility that the limits remain nonzero. For example, if we take one of the choices discussed in Section 3.1, i.e.

其中 $s_0 = \nabla \log \rho_0$、$s_1 = \nabla \log \rho_1$。若 $\gamma \in C^2([0,1])$，由于 $\gamma(0) = \gamma(1) = 0$，这些表达式中含得分 s_0、s_1 的项就会消失。若选择 $\gamma^2 \in C^1([0,1])$，但 γ 在 $t = 0$ 或 $t = 1$ 不可微，这些极限仍有可能非零。例如，取 3.1 节讨论过的一种选择，即

$$\gamma(t) = \sqrt{at(1-t)}, \qquad a > 0, \tag{4.13}$$

S262 · Original / 原文 · PDF p.35 · Thmtheorem42.p1.4

we obtain

我们得到

$$\lim_{t\to 0}\gamma(t)\dot\gamma(t) = -\lim_{t\to 1}\gamma(t)\dot\gamma(t) = \frac{a}{2}. \tag{4.14}$$

S263 · Original / 原文 · PDF p.35 · Thmtheorem42.p1.5

As a result, the choice (4.13) ensures that the velocity b encodes information about the score of the densities ρ_0 and ρ_1 at the end points. We stress however that, while the choice of $\gamma(t)$ given in (4.13) is appealing because of its nontrivial influence on the velocity b at the endpoints, the user is free to explore a variety of alternatives. We present some examples in Table 2, specifying the differentiability of γ at $t = 0$ and $t = 1$. The function $\gamma(t)$ specified in (4.13) is the only featured case for which the contribution from the score is non-vanishing in the velocity b at the endpoints. In Section 7, we illustrate on numerical examples that there are tradeoffs between different choices of γ, which might be directly related to this fact. When using the ODE as a generative model, the score is only felt through b, whereas it is explicit when using the SDE as a generative model.

因此，选择（4.13）确保速度 b 在端点编码了密度 ρ_0、ρ_1 的得分信息。不过我们强调，尽管（4.13）的 $\gamma(t)$ 因对端点速度有非平凡影响而很有吸引力，使用者仍可探索多种其他选择。表 2 给出一些例子，并注明 γ 在 $t=0$ 和 $t=1$ 的可微性。在所列情形中，只有（4.13）的 $\gamma(t)$ 使端点速度 b 中的得分贡献不消失。第 7 节通过数值例子说明，不同 γ 选择之间存在折中，这可能与上述事实直接相关。采用 ODE 生成模型时，得分仅通过 b 起作用；采用 SDE 生成模型时，得分则显式出现。

$\gamma(t)$:	$\sqrt{at(1-t)}$	$t(1-t)$	$\hat\sigma(t)$	$\sin^2(\pi t)$
C^1 at $t=0,1$	✗	✓	✓	✓

Table 2: Differentiability of $\gamma(t)z$. A characterization of the possible choices of $\gamma(t)$ with respect to their differentiability. The column specified by $\hat\sigma(t)$ is sum of sigmoid functions, made compact by the notation $\hat\sigma(t) = \sigma(f(t-\frac{1}{2})+1) - \sigma(f(t-\frac{1}{2})-1) - \sigma(-\frac{f}{2}+1) + \sigma(-\frac{f}{2}-1)$, where $\sigma(t) = e^t/(1+e^t)$ and f is a scaling factor.

表 2：γ(t)z 的可微性。比较不同 γ 在端点的可微性。σ̂(t) 是 sigmoid 函数的组合：σ(f(t−1/2)+1)−σ(f(t−1/2)−1)−σ(−f/2+1)+σ(−f/2−1)，其中 σ(t)=eᵗ/(1+eᵗ)，f 为尺度因子。

4.4 Spatially linear one-sided interpolants

4.4 空间线性单侧插值

S264 · Original / 原文 · PDF p.35 · S4.SS4.p1.1

Much of the discussion above generalizes to one-sided interpolants if we take the function $J(t, x_1)$ in (3.14) to be linear in x_1 and define

若令（3.14）中的函数 $J(t,x_1)$ 对 x_1 为线性，并定义下面的插值，上述大部分讨论也可推广至单侧插值：

$$x_t^{\text{os,lin}} = \alpha(t)z + \beta(t)x_1, \qquad t\in[0,1] \tag{4.15}$$

S265 · Original / 原文 · PDF p.36 · S4.SS4.p1.2

where $\alpha^2, \beta \in C^2([0,1])$ and $\alpha(0) = \beta(1) = 1$, $\alpha(1) = \beta(0) = 0$, and $\alpha(t) > 0$ for all $t \in [0,1)$. The velocity b and the score s defined in (2.10) and (2.14) can now be expressed as

其中 $\alpha^2, \beta \in C^2([0,1])$，$\alpha(0)=\beta(1)=1$、$\alpha(1)=\beta(0)=0$，且对所有 $t\in[0,1)$ 有 $\alpha(t)>0$。此时，（2.10）和（2.14）定义的速度 b 与得分 s 可表示为

$$b(t,x) = \dot{\alpha}(t)\eta_z^{\text{os}}(t,x) + \dot{\beta}(t)\eta_1^{\text{os}}(t,x), \qquad s(t,x) = -\alpha^{-1}(t)\eta_z^{\text{os}}(t,x), \tag{4.16}$$

S266 · Original / 原文 · PDF p.36 · S4.SS4.p1.3

where the second expression holds for all $t \in [0,1)$ and we defined:

其中第二个表达式对所有 $t \in [0,1)$ 成立，并定义：

$$\eta_z^{\text{os}}(t,x) = \mathbb{E}(z|x_t^{\text{os,lin}} = x), \qquad \eta_1^{\text{os}}(t,x) = \mathbb{E}(x_1|x_t^{\text{os,lin}} = x). \tag{4.17}$$

S267 · Original / 原文 · PDF p.36 · S4.SS4.p1.4

Note that, by definition of the conditional expectation, η_z^{os} and η_1^{os} satisfy

由条件期望定义，η_z^{os} 与 η_1^{os} 满足

$$\forall (t,x) \in [0,1] \times \mathbb{R}^d \quad : \quad \alpha(t)\eta_z^{\text{os}}(t,x) + \beta(t)\eta_1^{\text{os}}(t,x) = x. \tag{4.18}$$

S268 · Original / 原文 · PDF p.36 · S4.SS4.p1.5

As a result, only one of them needs to be estimated. For example, we can express η_1^{os} as a function of η_z^{os} for all t such that $\beta(t) \; / = 0$, and use the result to express the velocity (4.16) as

因此，只需估计其中一个。例如，对所有满足 $\beta(t) \neq 0$ 的 t，可把 η_1^{os} 表示为 η_z^{os} 的函数，再据此将速度（4.16）写成

$$b(t,x) = \dot{\beta}(t)\beta^{-1}(t)x + \big(\dot{\alpha}(t) - \alpha(t)\dot{\beta}(t)\beta^{-1}(t)\big)\eta_z^{\text{os}}(t,x) \qquad \forall t \; : \; \beta(t) \neq 0. \tag{4.19}$$

S269 · Original / 原文 · PDF p.36 · S4.SS4.p1.6

Assuming that $\beta(t) \neq 0$ for all $t \in (0,1]$, this formula only needs to be supplemented at $t = 0$ with

若对所有 $t \in (0,1]$ 都有 $\beta(t) \neq 0$，则只需在 $t = 0$ 为该公式补充

$$b(0,x) = \dot{\alpha}(0)x + \dot{\beta}(0)\mathbb{E}[x_1] \tag{4.20}$$

S270 · Original / 原文 · PDF p.36 · S4.SS4.p1.7

which follows from (4.16) since $x_{t=0}^{\text{os,lin}} = z$. Later in Section 5.2 we will show that using the velocity b in (4.19) to solve the probability flow ODE (2.32) can be seen as using a denoiser to construct a generative model.

这由（4.16）和 $x_{t=0}^{\text{os,lin}} = z$ 得出。稍后的 5.2 节将说明，用（4.19）中的速度 b 求解概率流 ODE（2.32），可视为使用去噪器构造生成模型。

S271 · Original / 原文 · PDF p.36 · S4.SS4.p2.2

Finally note that η_z and/or η_1 can be estimated using the following two objective functions, respectively:

最后，η_z 和 / 或 η_1 可分别通过下列两个目标函数估计：

$$\begin{aligned}
\mathcal{L}_{\eta_z}(\hat{\eta}_z^{\text{os}}) &= \int_0^1 \mathbb{E}\left[\tfrac{1}{2}|\hat{\eta}_z^{\text{os}}(t, x_t^{\text{os,lin}})|^2 - z \cdot \hat{\eta}_z^{\text{os}}(t, x_t^{\text{os,lin}})\right] dt, \\
\mathcal{L}_{\eta_1}(\hat{\eta}_1^{\text{os}}) &= \int_0^1 \mathbb{E}\left[\tfrac{1}{2}|\hat{\eta}_1^{\text{os}}(t, x_t^{\text{os,lin}})|^2 - x_1 \cdot \hat{\eta}_1^{\text{os}}(t, x_t^{\text{os,lin}})\right] dt.
\end{aligned} \tag{4.21}$$

5 Connections with other methods

5 与其他方法的联系

S272 · Original / 原文 · PDF p.36 · S5.p1.1

In this section, we discuss connections between the stochastic interpolant framework and the score-based diffusion method *(Song et al., 2021b)*, the stochastic localization framework *(Eldan, 2013; El Alaoui and Montanari, 2022; Montanari, 2023)*), denoising methods *(Simoncelli and Adelson, 1996; Hyvärinen, 1999; Kadkhodaie and Simoncelli, 2021; Ho et al., 2020)*, and the rectified flow method *(Liu et al., 2023b)*.

本节讨论随机插值框架与以下方法的联系：基于得分的扩散方法（Song 等，2021b）、随机局部化框架（Eldan，2013；El Alaoui 与 Montanari，2022；Montanari，2023）、去噪方法（Simoncelli 与 Adelson，1996；Hyvärinen，1999；Kadkhodaie 与 Simoncelli，2021；Ho 等，2020），以及整流流方法（Liu 等，2023b）。

5.1 Score-based diffusion models and stochastic localization

5.1 得分扩散与随机局部化

S273 · Original / 原文 · PDF p.37 · S5.SS1.p1.1

Score-based diffusion models (SBDM) are based on variants of the Ornstein-Uhlenbeck process

基于得分的扩散模型（SBDM）以 Ornstein–Uhlenbeck 过程的变体为基础：

$$dZ_\tau = -Z_\tau dt + \sqrt{2} dW_\tau, \qquad Z_{\tau=0} \sim \rho_1, \tag{5.1}$$

S274 · Original / 原文 · PDF p.37 · S5.SS1.p1.2

which has the property that the marginal density of its solution at time τ converges to a standard normal as τ tends towards infinity. By learning the score of the density of Z_τ, we can write the associated backward SDE for (5.1), which can then be used as a generative model – this backwards SDE is also the one that is used in the stochastic localization process, see *Montanari (2023)*.

该过程具有如下性质：当 $\tau \to \infty$ 时，其解在时刻 τ 的边缘密度收敛到标准正态分布。学习 Z_τ 密度的得分后，可以写出与（5.1）对应的反向 SDE，再用它作为生成模型。这个反向 SDE 也是随机局部化过程所用的方程，见 Montanari（2023）。

S275 · Original / 原文 · PDF p.37 · S5.SS1.p2.1

To see the connection with stochastic interpolants, notice that the solution of (5.1) from the initial condition $Z_{\tau=0} = x_1 \sim \rho_1$ can be written exactly as

为看清与随机插值的联系，注意（5.1）从初态 $Z_{\tau=0} = x_1 \sim \rho_1$ 出发的解可精确写成

$$Z_\tau = x_1 e^{-\tau} + \sqrt{2} \int_0^\tau e^{-\tau+s} dW_s. \tag{5.2}$$

S276 · Original / 原文 · PDF p.37 · S5.SS1.p2.2

As a result, the law of Z_τ conditioned on $Z_{\tau=0} = x_1$ is given by

因此，在条件 $Z_{\tau=0} = x_1$ 下，Z_τ 的分布为

$$Z_\tau \sim \mathsf{N}(x_1 e^{-\tau}, (1 - e^{-2\tau}) Id), \tag{5.3}$$

S277 · Original / 原文 · PDF p.37 · S5.SS1.p2.3

for any time $\tau \in [0, \infty)$. This is also the law of the process

对任意时刻 $\tau \in [0, \infty)$ 都成立。这也是下列过程的分布：

$$y_\tau = x_1 e^{-\tau} + \sqrt{1 - e^{-2\tau}}\, z, \qquad z \sim \mathsf{N}(0, Id), \qquad \tau \in [0, \infty). \tag{5.4}$$

S278 · Original / 原文 · PDF p.37 · S5.SS1.p2.4

If we let $x_1 \sim \rho_1$ with $x_1 \perp z$, the process y_τ is similar to a one-sided stochastic interpolant, except the density of y_τ only converges to $\mathsf{N}(0, Id)$ as $\tau \to \infty$; by contrast, the one-sided interpolants we introduced in Section 3.2 converge on the finite interval $[0, 1]$. In SBDM, this is handled by capping the evolution of Z_τ to a finite time interval $[0, T]$ with $T < \infty$, and then by using the backward SDE associated with (5.1) restricted to $[0, T]$. However, this introduces a bias that is not present with one-sided stochastic interpolants, because the final condition used for the backwards SDE in SBDM is drawn from $\mathsf{N}(0, Id)$ even though the density of the process (5.1) is not Gaussian at time T.

令 $x_1 \sim \rho_1$ 且 $x_1 \perp z$，过程 y_τ 与单侧随机插值类似，区别是 y_τ 的密度仅在 $\tau \to \infty$ 时才收敛到 $\mathsf{N}(0, Id)$，而 3.2 节引入的单侧插值在有限区间 $[0, 1]$ 上即完成收敛。SBDM 的处理方式是把 Z_τ 的演化截断到有限区间 $[0, T]$，其中 $T < \infty$，然后使用（5.1）对应反向 SDE 在 $[0, T]$ 上的限制。但这会引入单侧随机插值所没有的偏差：SBDM 反向 SDE 使用的终态从 $\mathsf{N}(0, Id)$ 采样，然而过程（5.1）在时刻 T 的密度并非高斯。

S279 · Original / 原文 · PDF p.37 · S5.SS1.p3.1

We can, however, turn (5.4) into a one-sided linear stochastic interpolant by defining $t = e^{-\tau}$ and by choosing $\alpha(t)$ and $\beta(t)$ in (4.15) to have a specific form. More precisely, evaluating (5.4) at $\tau = -\log t$,

不过，通过定义 $t = e^{-\tau}$ 并为（4.15）中的 $\alpha(t)$、$\beta(t)$ 选择特定形式，可把（5.4）变成单侧线性随机插值。具体地，在 $\tau = -\log t$ 处计算（5.4），得到

$$y_{\tau = -\log t} = \sqrt{1 - t^2} z + t x_1 \equiv x_t^{\mathsf{os,lin}} \quad \text{for} \quad \alpha(t) = \sqrt{1 - t^2}, \quad \beta(t) = t. \tag{5.5}$$

S280 · Original / 原文 · PDF p.37 · S5.SS1.p3.2

With this choice of $\alpha(t)$ and $\beta(t)$, from (4.16) we get the velocity field

对于这样的 $\alpha(t)$ 和 $\beta(t)$，由（4.16）得到速度场

$$b(t, x) = -\frac{t}{\sqrt{1 - t^2}} \eta_z^{\mathsf{os}}(t, x) + \eta_1^{\mathsf{os}}(t, x) \equiv t s(t, x) + \eta_1^{\mathsf{os}}(t, x) \tag{5.6}$$

S281 · Original / 原文 · PDF p.37 · S5.SS1.p3.3

where η_z^{os} and η_1^{os} are defined in (4.17). This expression shows that the velocity b used in the probability flow ODE (2.32) is well-behaved at all times, including at $t = 1$ where $\dot\alpha(t)$ is singular. The same is true for the drift $b_{\mathsf{F}}(t, x) = b(t, x) + \epsilon(t) s(t, x)$ used in the forward SDE (2.33), regardless of the choice of $\epsilon \in C^0([0, 1])$ with $\epsilon(t) \geq 0$. This shows that casting SBDM into a one-sided linear stochastic interpolant (4.15) allows the construction of unbiased generative models that operate on $t \in [0, 1]$. This comes at no extra computational cost, since only one of the two functions defined in (4.17) needs to be estimated, which is akin to estimating the score in SBDM.

其中 η_z^{os}、η_1^{os} 由（4.17）定义。这个表达式表明，概率流 ODE（2.32）所用速度 b 在所有时刻均良好，包括 $\dot\alpha(t)$ 奇异的 $t = 1$。对正向 SDE（2.33）的漂移 $b_{\mathsf{F}}(t, x) = b(t, x) + \epsilon(t) s(t, x)$ 也一样，无论选择何种满足 $\epsilon(t) \geq 0$ 的 $\epsilon \in C^0([0, 1])$。这说明，把 SBDM 写成单侧线性随机插值（4.15），即可构造在 $t \in [0, 1]$ 上运行的无偏生成模型，并不增加计算成本，因为只需估计（4.17）两个函数中的一个，这与 SBDM 估计得分类似。

S282 · Original / 原文 · PDF p.38 · S5.SS1.p4.1

It is worth comparing the above procedure to an equivalent change of time at the level of the diffusion process (5.1), which we now show leads to singular terms that pose numerical and analytical difficulties. Indeed, if we define $Z_t^{\mathsf{B}} = Z_{\tau=-\log t}$, from (5.1) we obtain

值得把上述过程与直接在扩散过程（5.1）层面作等价时间变换比较。下面说明，后者会产生奇异项，带来数值和分析困难。事实上，若定义 $Z_t^{\mathsf{B}} = Z_{\tau=-\log t}$，由（5.1）可得

$$dZ_t^{\mathsf{B}} = t^{-1} Z_t^{\mathsf{B}} dt + \sqrt{2t^{-1}} dW_t^{\mathsf{B}}, \quad Z_{t=1}^{\mathsf{B}} \sim \rho_1, \tag{5.7}$$

S283 · Original / 原文 · PDF p.38 · S5.SS1.p4.2

to be solved backwards in time. Because of the factor t^{-1}, this SDE cannot easily be solved until $t = 0$, which corresponds to $\tau = \infty$ in the original (5.1). For the same reason, the forward SDE associated with (5.7)

该方程需沿时间反向求解。由于存在 t^{-1} 因子，这个 SDE 很难一直求解到 $t = 0$；该时刻对应原式（5.1）中的 $\tau = \infty$。出于同一原因，与（5.7）对应的正向 SDE

$$dZ_t^{\mathsf{F}} = t^{-1} Z_t^{\mathsf{F}} dt + 2t^{-1} s(t, Z_t^{\mathsf{F}}) dt + \sqrt{2t^{-1}} dW_t, \tag{5.8}$$

S284 · Original / 原文 · PDF p.38 · S5.SS1.p4.3

cannot be solved from $t = 0$, where formally $Z_{t=0}^{\mathsf{F}} \overset{d}{=} Z_{t=0}^{\mathsf{B}} = Z_{\tau=\infty} \sim \mathsf{N}(0, Id)$. This means it cannot be used as a generative model unless we start from some $t > 0$, which introduces a bias. Importantly, this problem does not arise with the stochastic interpolant framework, because the construction of the density $\rho(t)$ connecting ρ_0 and ρ_1 is handled separately from the construction of the process that generates samples from $\rho(t)$. By contrast, SBDM combines these two operations into one, leading to the singularity at $t = 0$ in the coefficients in (5.7) and (5.8).

无法从 $t = 0$ 开始求解；形式上，此处 $Z_{t=0}^{\mathsf{F}} \overset{d}{=} Z_{t=0}^{\mathsf{B}} = Z_{\tau=\infty} \sim \mathsf{N}(0, Id)$。这意味着，除非从某个 $t > 0$ 开始、从而引入偏差，否则它不能用作生成模型。重要的是，随机插值框架没有这个问题，因为构造连接 ρ_0 与 ρ_1 的密度 $\rho(t)$，与构造从 $\rho(t)$ 生成样本的过程，是分别处理的。相比之下，SBDM 将两步合一，导致（5.7）、（5.8）的系数在 $t = 0$ 处奇异。

Remark 43.说明 43

S285 · Original / 原文 · PDF p.38 · Thmtheorem43.p1.1

To emphasize the last point made above, we stress that there is no contradiction between having a singular drift and diffusion coefficient in (5.8), and being able to write a nonsingular SDE with stochastic interpolants. To see why, notice that the stochastic interpolant tells us that we can change the diffusion coefficient in (5.8) to any nonsingular $\epsilon \in C^0([0,1])$ with $\epsilon(t) \geq 0$ and replace this SDE with

为强调上一点，我们指出，（5.8）的漂移和扩散系数奇异，与随机插值能够给出一个非奇异 SDE 之间并不矛盾。原因是，随机插值告诉我们，可以把（5.8）的扩散系数改为任意非奇异、满足 $\epsilon(t) \geq 0$ 的 $\epsilon \in C^0([0,1])$，并用下列方程替换该 SDE：

$$dX_t^{\mathsf{F}} = t^{-1} Z_t^{\mathsf{F}} dt + (t^{-1} + \epsilon(t)) s(t, X_t^{\mathsf{F}}) dt + \sqrt{2\epsilon(t)} dW_t, \tag{5.9}$$

S286 · Original / 原文 · PDF p.38 · Thmtheorem43.p1.2

This SDE has the property that $X_{t=1}^{\mathsf{F}} \sim \rho_1$ if $X_{t=0}^{\mathsf{F}} \sim \rho_0$, and its drift is also nonsingular at $t = 0$ and given precisely by (5.6). Indeed, using the constraint (4.18), which here reads $x = \sqrt{1-t^2}\eta_z^{\mathsf{os}}(t,x) + t\eta_1^{\mathsf{os}}(t,x) \equiv -(1-t^2)s(t,x) + t\eta_1^{\mathsf{os}}(t,x)$, it is easy to see that

这个 SDE 满足：若 $X_{t=0}^{\mathsf{F}} \sim \rho_0$，则 $X_{t=1}^{\mathsf{F}} \sim \rho_1$。它的漂移在 $t = 0$ 也非奇异，恰由（5.6）给出。事实上，使用约束（4.18），其在此写为 $x = \sqrt{1-t^2}\eta_z^{\mathsf{os}}(t,x) + t\eta_1^{\mathsf{os}}(t,x) \equiv -(1-t^2)s(t,x) + t\eta_1^{\mathsf{os}}(t,x)$，容易看出

$$t^{-1}x + t^{-1}s(t,x) = ts(t,x) + \eta_1^{\text{os}}(t,x), \tag{5.10}$$

S287 · Original / 原文 · PDF p.38 · Thmtheorem43.p1.3

which is nonsingular at $t = 0$.

它在 $t = 0$ 处非奇异。

5.2 Denoising methods

5.2 去噪方法

S288 · Original / 原文 · PDF p.38 · S5.SS2.p1.1

Consider the spatially-linear one-sided stochastic interpolant defined in (4.15). By solving this equation for x_1, we obtain

考虑（4.15）定义的空间线性单侧随机插值。由该式解出 x_1，可得

$$x_1 = \beta^{-1}(t)\left(x_t^{\text{os,lin}} - \alpha(t)z\right) \qquad t \in (0,1]. \tag{5.11}$$

S289 · Original / 原文 · PDF p.38 · S5.SS2.p1.2

Taking a conditional expectation at fixed $x_t^{\text{os,lin}}$ and using (4.17) implies that

固定 $x_t^{\text{os,lin}}$ 取条件期望，并使用（4.17），可得

$$\mathbb{E}(x_1|x_t^{\text{os,lin}}) = \eta_1^{\text{os}}(t, x_t^{\text{os,lin}}) = \beta^{-1}(t)\left(x_t^{\text{os,lin}} - \alpha(t)\eta_z^{\text{os}}(t, x_t^{\text{os,lin}})\right) \qquad t \in (0,1] \tag{5.12}$$

S290 · Original / 原文 · PDF p.39 · S5.SS2.p1.3

while trivially $\mathbb{E}(x_1|x_{t=0}^{\text{os,lin}}) = \mathbb{E}[x_1]$ since $x_{t=0}^{\text{os,lin}} = z$. This expression is commonly used in denoising methods *(Simoncelli and Adelson, 1996; Kadkhodaie and Simoncelli, 2021)*, and it is Stein's unbiased risk estimator (SURE) for x_1 given the noisy information in $x_t^{\text{os,lin}}$ *Stein (1981)*. Rather than considering the conditional expectation of x_1, we can consider an analogous quantity for $x_s^{\text{os,lin}}$ for any $s \in [0,1]$; this leads to the following result.

而由于 $x_{t=0}^{\text{os,lin}} = z$，显然有 $\mathbb{E}(x_1 \mid x_{t=0}^{\text{os,lin}}) = \mathbb{E}[x_1]$。这个表达式常见于去噪方法（Simoncelli 与 Adelson，1996；Kadkhodaie 与 Simoncelli，2021），原文将其称为给定 $x_t^{\text{os,lin}}$ 的含噪信息时针对 x_1 的 Stein 无偏风险估计（SURE；Stein，1981）。除了 x_1 的条件期望，还可考虑任意 $s \in [0,1]$ 时 $x_s^{\text{os,lin}}$ 的类似量，由此得到以下结果。

Lemma 44 (SURE).引理 44（SURE）

S291 · Original / 原文 · PDF p.39 · Thmtheorem44.p1.1

For $s \in [0,1]$, we have

对 $s \in [0,1]$，有

$$\mathbb{E}(x_s^{\text{os,lin}}|x_t^{\text{os,lin}}) = \frac{\beta(s)}{\beta(t)}x_t^{\text{os,lin}} + \left(\alpha(s) - \frac{\alpha(t)\beta(s)}{\beta(t)}\right)\eta_z^{\text{os}}(t, x_t^{\text{os,lin}}) \qquad t \in (0,1] \tag{5.13}$$

S292 · Original / 原文 · PDF p.39 · Thmtheorem44.p1.2

and $\mathbb{E}(x_s^{\text{os,lin}}|x_{t=0}^{\text{os,lin}}) = \alpha(s)x_{t=0}^{\text{os,lin}} + \beta(s)\mathbb{E}[x_1]$.

以及 $\mathbb{E}(x_s^{\text{os,lin}} \mid x_{t=0}^{\text{os,lin}}) = \alpha(s)x_{t=0}^{\text{os,lin}} + \beta(s)\mathbb{E}[x_1]$。

Proof.

证明

S293 · Original / 原文 · PDF p.39 · S5.SS2.p2.1

(5.13) follows from inserting (5.11) in the expression for $x_s^{\text{os,lin}}$ and taking the conditional expectation using the definition of η_1^{os} and η_z^{os} in (4.17). $\mathbb{E}(x_s^{\text{os,lin}}|x_{t=0}^{\text{os,lin}}) = \alpha(s)x_{t=0}^{\text{os,lin}} + \beta(s)\mathbb{E}[x_1]$ follows from $x_{t=0}^{\text{os,lin}} = z$ together with (5.11). ■

将（5.11）代入 $x_s^{\text{os,lin}}$ 的表达式，并利用（4.17）中 η_1^{os}、η_z^{os} 的定义取条件期望，即得（5.13）。等式 $\mathbb{E}(x_s^{\text{os,lin}} \mid x_{t=0}^{\text{os,lin}}) = \alpha(s)x_{t=0}^{\text{os,lin}} + \beta(s)\mathbb{E}[x_1]$ 则由 $x_{t=0}^{\text{os,lin}} = z$ 及（5.11）得出。证毕。

S294 · Original / 原文 · PDF p.39 · S5.SS2.p3.1

At this stage, equations (5.11) and (5.13) cannot be used as generative models: the random variable $\mathbb{E}(x_1|x_t^{\text{os,lin}})$ is not a sample of ρ_1, and the random variable $\mathbb{E}(x_s^{\text{os,lin}}|x_t^{\text{os,lin}})$ is not a sample from $\rho(s)$, the density of $x_s^{\text{os,lin}}$. However, the following result shows that if we iterate upon formula (5.13) by taking infinitesimal steps, we obtain a generative model consistent with the probability flow equation (2.32) associated with $x_t^{\text{os,lin}}$.

此时，（5.11）与（5.13）还不能用作生成模型：随机变量 $\mathbb{E}(x_1 \mid x_t^{\text{os,lin}})$ 不是 ρ_1 的样本，随机变量 $\mathbb{E}(x_s^{\text{os,lin}} \mid x_t^{\text{os,lin}})$ 也不是 $x_s^{\text{os,lin}}$ 的密度 $\rho(s)$ 的样本。但下一个结果表明，若对公式（5.13）以无穷小步长反复迭代，就能得到一个与 $x_t^{\text{os,lin}}$ 所对应概率流方程（2.32）一致的生成模型。

Theorem 45.定理 45

S295 · Original / 原文 · PDF p.39 · Thmtheorem45.p1.1

Let $t_j = j/N$ with $j \in \{1, \dots, N\}$, set $X_1^{\text{den}} = z$, and define for $j = 1, \dots, N-1$,

令 $t_j = j/N$，其中 $j \in \{1, \dots, N\}$，设 $X_1^{\text{den}} = z$，并对 $j = 1, \dots, N-1$ 定义

$$X_{j+1}^{\text{den}} = \frac{\beta(t_{j+1})}{\beta(t_j)} X_j^{\text{den}} + \left(\alpha(t_{j+1}) - \frac{\alpha(t_j)\beta(t_{j+1})}{\beta(t_j)}\right) \eta_z^{\text{os}}(t_j, X_j^{\text{den}}). \tag{5.14}$$

S296 · Original / 原文 · PDF p.39 · Thmtheorem45.p1.2

Then, (5.14) is a consistent integration scheme for the probability flow equation (2.32) associated with the velocity field (4.16) expressed as in (4.19). That is, if $N, j \to \infty$ with $j/N \to t \in [0,1]$, then $X_j^{\text{den}} \to X_t$ where

则（5.14）是概率流方程（2.32）的一个相容积分格式；其对应速度场由（4.16）定义，并写成（4.19）的形式。即当 $N, j \to \infty$ 且 $j/N \to t \in [0,1]$ 时，有 $X_j^{\text{den}} \to X_t$，其中

$$\dot{X}_t = b(t, X_t) = \frac{\dot{\beta}(t)}{\beta(t)} X_t + \left(\dot{\alpha}(t) - \frac{\alpha(t)\dot{\beta}(t)}{\beta(t)}\right) \eta_z^{\text{os}}(t, X_t), \quad X_{t=0} = z. \tag{5.15}$$

S297 · Original / 原文 · PDF p.39 · Thmtheorem45.p1.3

In particular, if $z \sim \mathsf{N}(0, Id)$, then $X_N^{\text{den}} \to x_1 \sim \rho_1$ in this limit.

特别地，若 $z \sim \mathsf{N}(0, Id)$，则在这一极限下，$X_N^{\text{den}} \to x_1 \sim \rho_1$。

S298 · Original / 原文 · PDF p.39 · S5.SS2.p4.1

The proof of this theorem is given in Appendix B.7, and proceeds by Taylor expansion of the right-hand side of (5.14).

该定理的证明见附录 B.7，方法是对（5.14）右端作 Taylor 展开。

5.3 Rectified flows

5.3 矫正流

S299 · Original / 原文 · PDF p.39 · S5.SS3.p1.1

We now discuss how stochastic interpolants can be rectified according to the procedure proposed in *Liu et al. (2023b)*. Suppose that we have perfectly learned the velocity field b in the probability flow equation (2.32) for a given stochastic interpolant. Denote by $X_t(x)$ the solution to this ODE with the initial condition $X_{t=0}(x) = x$, i.e.

下面讨论如何按 Liu 等（2023b）提出的过程，对随机插值进行整流。假设对某个给定随机插值，我们已完美学得概率流方程（2.32）中的速度场 b。以 $X_t(x)$ 表示该 ODE 在初始条件 $X_{t=0}(x) = x$ 下的解，即

$$\frac{d}{dt}X_t(x) = b(t, X_t(x)), \qquad X_{t=0}(x) = x. \tag{5.16}$$

S300 · Original / 原文 · PDF p.40 · S5.SS3.p1.2

We can use the map $X_{t=1} : \mathbb{R}^d \to \mathbb{R}^d$ to define a new stochastic interpolant

可以利用映射 $X_{t=1} : \mathbb{R}^d \to \mathbb{R}^d$ 定义一个新的随机插值：

$$x_t^{\text{rec}} = \alpha(t)x_0 + \beta(t)X_{t=1}(x_0), \tag{5.17}$$

S301 · Original / 原文 · PDF p.40 · S5.SS3.p1.3

where $\alpha^2, \beta \in C^2([0,1])$ satisfy $\alpha(0) = \beta(1) = 1$, $\alpha(1) = \beta(0)w = 0$, and $\alpha(t) > 0$ for all $t \in [0,1)$. Clearly, we then have $x_{t=0}^{\text{rec}} = x_0 \sim \rho_0$ since $X_{t=0}(x_0) = x_0$ and $x_{t=1}^{\text{rec}} = X_{t=1}(x_0) \sim \rho_1$ by definition of the probability flow equation. We can define a new probability flow equation associated with the velocity field

其中，$\alpha^2, \beta \in C^2([0,1])$ 满足 $\alpha(0) = \beta(1) = 1$、$\alpha(1) = \beta(0)w = 0$，且对所有 $t \in [0,1)$ 都有 $\alpha(t) > 0$。显然，由于 $X_{t=0}(x_0) = x_0$，有 $x_{t=0}^{\text{rec}} = x_0 \sim \rho_0$；根据概率流方程的定义，又有 $x_{t=1}^{\text{rec}} = X_{t=1}(x_0) \sim \rho_1$。我们可以定义与下述速度场关联的一个新的概率流方程：

$$b^{\text{rec}}(t,x) = \mathbb{E}[\dot{x}_t^{\text{rec}} | x_t^{\text{rec}} = x] = \dot{\alpha}(t)\mathbb{E}[x_0 | x_t^{\text{rec}} = x] + \dot{\beta}(t)\mathbb{E}[X_{t=1}(x_0) | x_t^{\text{rec}} = x]. \tag{5.18}$$

S302 · Original / 原文 · PDF p.40 · S5.SS3.p1.4

It is easy to see that this velocity field is amenable to estimation, since it is the unique minimizer of

容易看出，这一速度场可以通过估计得到，因为它是下列目标的唯一极小化解：

$$\mathcal{L}_{b^{\text{rec}}}[\hat{b}^{\text{rec}}] = \int_0^1 \mathbb{E}\left[\tfrac{1}{2}|\hat{b}^{\text{rec}}(t, x_t^{\text{rec}})|^2 - (\dot{\alpha}(t)x_0 + \dot{\beta}(t)X_{t=1}(x_0)) \cdot \hat{b}^{\text{rec}}(t, x_t^{\text{rec}})\right] dt, \tag{5.19}$$

S303 · Original / 原文 · PDF p.40 · S5.SS3.p1.5

where x_t^{rec} is given in (5.17) and the expectation is now only on $x_0 \sim \rho_0$. Our next result show that the probability flow equation associated with the velocity field (5.18) has straight line solutions, but ultimately it leads to a generative model that is identical to the one based on (5.16). To phrase this result, we first make an assumption on the invertibility of x_t^{rec}.

其中，x_t^{rec} 由式（5.17）给出，而此时的期望仅对 $x_0 \sim \rho_0$ 取。下一个结果表明，与速度场（5.18）关联的概率流方程具有直线解，但最终得到的生成模型与基于式（5.16）的模型相同。为表述这一结果，我们首先对 x_t^{rec} 的可逆性作出假设。

> Assumption 46.假设 46
>
> S304 · Original / 原文 · PDF p.40 · Thmtheorem46.p1.1

The map $x \to M(t,x)$ where $M(t,x) = \alpha(t)x + \beta(t)X_{t=1}(x)$ with $X_t(x)$ solution to (5.16) is invertible for all $(t,x) \in [0,1] \times \mathbb{R}^d$, i.e. $\exists N(t,\cdot,) : \mathbb{R}^d \to \mathbb{R}^d$ such that

令 $M(t,x) = \alpha(t)x + \beta(t)X_{t=1}(x)$，其中 $X_t(x)$ 是式（5.16）的解。假设对所有 $(t,x) \in [0,1] \times \mathbb{R}^d$，映射 $x \to M(t,x)$ 都可逆，即 $\exists N(t,\cdot,) : \mathbb{R}^d \to \mathbb{R}^d$，使得：

$$\forall (t,x) \in [0,1] \times \mathbb{R}^d \quad : \quad N(t, M(t,x)) = M(t, N(t,x)) = x. \tag{5.20}$$

S305 · Original / 原文 · PDF p.40 · S5.SS3.p2.1

This is equivalent to requiring that the determinant of the Jacobian of $M(t,x)$ is nonzero for all $(t,x) \in [0,1] \times \mathbb{R}^d$; put differently, Assumption 46 requires that the Jacobian of $X_{t=1}(x)$ never has eigenvalues precisely equal to $-\alpha(t)/\beta(t)$, which is generic. Under this assumption, we state the following theorem.

这等价于要求，对所有 $(t,x) \in [0,1] \times \mathbb{R}^d$，$M(t,x)$ 的 Jacobian 行列式均非零。换言之，假设 46 要求 $X_{t=1}(x)$ 的 Jacobian 矩阵从不具有恰好等于 $-\alpha(t)/\beta(t)$ 的特征值；这属于一般位置下的情形。在这一假设下，我们给出以下定理。

Theorem 47.定理 47

S306 · Original / 原文 · PDF p.40 · Thmtheorem47.p1.1

Consider the probability flow equation associated with (5.18),

考虑与式（5.18）关联的概率流方程：

$$\frac{d}{dt} X_t^{\text{rec}}(x) = b^{\text{rec}}(t, X_t^{\text{rec}}(x)), \qquad X_{t=0}^{\text{rec}}(x) = x. \tag{5.21}$$

S307 · Original / 原文 · PDF p.40 · Thmtheorem47.p1.2

Then, all solutions are such that $X_{t=1}^{\text{rec}}(x_0) \sim \rho_1$ if $x_0 \sim \rho_0$. In addition, if Assumption (46) holds, the velocity field defined in (5.18) reduces to

那么，若 $x_0 \sim \rho_0$，所有解均满足 $X_{t=1}^{\text{rec}}(x_0) \sim \rho_1$。此外，若假设（46）成立，式（5.18）定义的速度场可简化为：

$$b^{\text{rec}}(t,x) = \dot{\alpha}(t)N(t,x) + \dot{\beta}(t)X_{t=1}(N(t,x)) \tag{5.22}$$

S308 · Original / 原文 · PDF p.40 · Thmtheorem47.p1.3

and the solution to the probability flow ODE (5.21) is simply

而概率流 ODE（5.21）的解就是：

$$X_t^{\text{rec}}(x) = \alpha(t)x + \beta(t)X_{t=1}(x). \tag{5.23}$$

S309 · Original / 原文 · PDF p.40 · S5.SS3.p3.1

The proof is given in Appendix B.8.

证明见附录 B.8。

S310 · Original / 原文 · PDF p.40 · S5.SS3.p4.1

Theorem 47 implies that $X_t^{\text{rec}}(x)$ is a simpler flow than $X_t(x)$, but we stress that they give the same map, $X_{t=1}^{\text{rec}} = X_{t=1}$. In particular, $X_t^{\text{rec}}(x)$ reduces to a straight line between x and $X_{t=1}(x)$ for $\alpha(t) = 1 - t$ and $\beta(t) = t$. We also note that the approach can be used to learn a single-step map, since (5.22) and $N(t = 0, x) = x$ give

定理 47 意味着，$X_t^{\text{rec}}(x)$ 是一个比 $X_t(x)$ 更简单的流；但我们强调，两者给出相同的映射，即 $X_{t=1}^{\text{rec}} = X_{t=1}$。特别地，取 $\alpha(t) = 1 - t$、$\beta(t) = t$ 时，$X_t^{\text{rec}}(x)$ 简化为连接 x 与 $X_{t=1}(x)$ 的直线。我们还注意到，这一方法可以用来学习单步映射，因为式（5.22）与 $N(t = 0, x) = x$ 给出：

$$b^{\text{rec}}(t = 0, x) = \dot{\alpha}(0)x + \dot{\beta}(0)X_{t=1}(x), \tag{5.24}$$

S311 · Original / 原文 · PDF p.40 · S5.SS3.p4.2

which expresses $X_{t=1}(x)$ in terms of known quantities as long as $\dot{\beta}(0) \neq 0$. For example, if $\dot{\alpha}(0) = 0$ and $\dot{\beta}(0) = 1$, we obtain $b^{\text{rec}}(t = 0, x) = X_{t=1}(x)$.

只要 $\dot{\beta}(0) \neq 0$，该式就用已知量表达了 $X_{t=1}(x)$。例如，若 $\dot{\alpha}(0) = 0$、$\dot{\beta}(0) = 1$，则得到 $b^{\text{rec}}(t = 0, x) = X_{t=1}(x)$。

Remark 48 (Optimal transport).说明 48（最优输运）

S312 · Original / 原文 · PDF p.41 · Thmtheorem48.p1.1

The discussion above highlights the fact that a probability flow equation can have straight line solutions and lead to a map that exactly pushes ρ_0 onto ρ_1 but is not the optimal transport map. That is, straight line solutions is a necessary condition for optimal transport, but it is not sufficient.

上述讨论突出表明：一个概率流方程可以具有直线解，并产生将 ρ_0 精确推送到 ρ_1 的映射，但这个映射不一定是最优传输映射。也就是说，解沿直线运动是最优传输的必要条件，但不是充分条件。

Remark 49 (Gradient fields).说明 49（梯度场）

S313 · Original / 原文 · PDF p.41 · Thmtheorem49.p1.1

The map is unaffected by the rectification procedure because we do not impose that the velocity $b^{\text{rec}}(t, x)$ be a gradient field. If we do impose this structure by setting $b^{\text{rec}}(t, x) = \nabla\phi(t, x)$ for some $\phi : \mathbb{R}^d \to \mathbb{R}$, then $X_{t=1}^{\text{rec}} \neq X_{t=1}$. As shown in *Liu (2022)*, iterating over this procedure eventually gives the optimal transport map. That is, implemented over gradient fields and iterated infinitely, rectification computes Brenier's polar decomposition of the map *(Brenier, 1991)*.

矫正过程不改变该映射，是因为我们没有要求速度 $b^{\text{rec}}(t, x)$ 为梯度场。若施加这一结构，令 $b^{\text{rec}}(t, x) = \nabla\phi(t, x)$，其中 $\phi : \mathbb{R}^d \to \mathbb{R}$，则有 $X_{t=1}^{\text{rec}} \neq X_{t=1}$。如 Liu（2022）所示，反复迭代这一过程最终会得到最优传输映射。也就是说，在梯度场上实施并无限次迭代的矫正过程，计算的正是映射的 Brenier 极分解（Brenier，1991）。

Remark 50 (Consistency models).说明 50（一致性模型）

S314 · Original / 原文 · PDF p.41 · Thmtheorem50.p1.1

Recent work has introduced the notion of consistency models *(Song et al., 2023)*, which distill a velocity field learned via score-based diffusion into a single-step map. Section 5.1 and the previous discussion provide an alternative perspective

on consistency models, and show how they may be computed in the framework of stochastic interpolants via rectification.

近期工作提出了一致性模型的概念（Song 等，2023），将通过基于得分的扩散方法学到的速度场蒸馏为单步映射。第 5.1 节及前面的讨论为一致性模型提供了另一种理解视角，并表明可通过矫正在随机插值框架中计算这类模型。

6 Algorithmic aspects
6 算法细节

S315 · Original / 原文 · PDF p.41 · S6.p1.1

The methods described in the previous sections have efficient numerical realizations. Here, we detail algorithms and practical recommendations for an implementation. These suggestions can be split into two complementary tasks: learning the drift coefficients, and sampling with an ODE or an SDE.

前述各节的方法都可以高效地数值实现。这里，我们详细介绍实现算法与实践建议。这些建议可分为两项互补的任务：学习漂移系数，以及使用 ODE 或 SDE 进行采样。

6.1 Learning
6.1 学习

S316 · Original / 原文 · PDF p.41 · S6.SS1.p1.1

As described in Section 2.2, there are a variety of algorithmic choices that can be made when learning the drift coefficients in (2.9), (2.20), and (2.22). While all choices lead to exact generative models in the absence of numerical and statistical errors, in practice, the presence of these errors ensures that different choices lead to different generative models, some of which may perform better for specific applications. Here, we describe the various realizations explicitly.

如第 2.2 节所述，学习式（2.9）、（2.20）与（2.22）中的漂移系数时，可以作出多种算法选择。在不存在数值误差与统计误差时，所有选择都能得到精确的生成模型；但在实践中，这些误差的存在使不同选择产生不同的生成模型，其中一些可能更适合特定应用。这里，我们明确说明各种实现方式。

Deterministic generative modeling: Learning b versus learning v and s.
确定性建模：学习 b 或学习 v、s

S317 · Original / 原文 · PDF p.41 · S6.SS1.SSS0.Px1.p1.1

Recall from Section 2.2 that the drift b of the transport equation (2.9) can be written as $b(t,x) = v(t,x) - \gamma(t)\dot\gamma(t)s(t,x)$. This raises the practical question of whether it would be better to learn an estimate $\hat b$ of b by minimizing the empirical risk

回顾第 2.2 节，输运方程（2.9）的漂移 b 可以写成 $b(t,x) = v(t,x) - \gamma(t)\dot\gamma(t)s(t,x)$。这引出一个实践问题：是通过最小化下列经验风险，直接学习 b 的估计 $\hat b$ 更好，

$$\hat{\mathcal{L}}_b(\hat b) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{2}|\hat b(t_i, x_{t_i}^i)|^2 - \hat b(t_i, x_{t_i}^i)\cdot\left(\partial_t I(t_i, x_0^i, x_1^i) + \dot\gamma(t_i)z^i\right)\right), \tag{6.1}$$

S318 · Original / 原文 · PDF p.41 · S6.SS1.SSS0.Px1.p1.2

or to learn estimates of $\hat v$ and $\hat s$ by minimizing the empirical risks

还是通过最小化下列经验风险，分别学习估计 $\hat v$ 与 $\hat s$ 更好：

$$\hat{\mathcal{L}}_v(\hat v) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{2}|\hat v(t_i, x_{t_i}^i)|^2 - \hat v(t_i, x_{t_i}^i)\cdot\partial_t I(t_i, x_0^i, x_1^i)\right) \tag{6.2}$$

S319 · Original / 原文 · PDF p.41 · S6.SS1.SSS0.Px1.p1.3

and

以及

$$\hat{\mathcal{L}}_s(\hat{s}) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{2}|\hat{s}(t_i, x_{t_i}^i)|^2 + \gamma(t_i)^{-1}\hat{s}(t_i, x_{t_i}^i)\cdot z^i\right) \tag{6.3}$$

Algorithm 1 Learning b with arbitrary ρ_0 and ρ_1.

算法 1：任意端点密度下学习 b。

Input: Batch size N, interpolant function $I(t, x_0, x_1)$, coupling $\nu(dx_0, dx_1)$ to sample x_0, x_1, noise function $\gamma(t)$, initial parameters θ_b, gradient-based optimization algorithm to minimize $\mathcal{L}_b$ defined in (2.13), number of gradient steps N_g.
Returns: An estimate $\hat{b}$ of b.
for $j = 1, \ldots, N_g$ do
  Draw N samples $(t_i, x_0^i, x_1^i, z^i) \sim \mathsf{Unif}([0,1]) \times \nu \times \mathsf{N}(0,1)$, for $i = 1, \ldots, N$.
  Construct samples $x_t^i = I(t_i, x_0^i, x_1^i) + \gamma(t_i)z^i$, for $i = 1, \ldots, N$.
  Take gradient step with respect to $\mathcal{L}_b\left(\theta_b, \{x_t^i\}_{i=1}^N\right)$.
Return: $\hat{b}$.

输入：批量 N、插值函数 I、端点耦合 ν、噪声 γ、初始参数 θb、最小化式（2.13）的梯度优化器、梯度步数 Ng。输出 b 的估计 $\hat{b}$。对 j=1,…,Ng：抽取 N 组独立样本 $(t_i, x_0^i, x_1^i, z^i) \sim \mathrm{Unif}([0,1]) \times \nu \times N(0,1)$；构造 $x_t^i = I(t_i, x_0^i, x_1^i) + \gamma(t_i)z^i$；对批数据的 Lb 做一步梯度更新。最后返回 $\hat{b}$。

S320 · Original / 原文 · PDF p.41 · S6.SS1.SSS0.Px1.p2.1

and construct the estimate $\hat{b}(t,x) = \hat{v}(t,x) - \gamma(t)\dot{\gamma}(t)\hat{s}(t,x)$. Above, $x_{t_i}^i = I(t_i, x_0^i, x_1^i) + \gamma(t_i)z^i$, and N denotes the number of samples t^i, x_0^i, and x_1^i. If the practitioner is interested in a deterministic generative model (for example, to exploit adaptive integration or exact likelihood computation), learning the estimate $\hat{b}$ directly only requires learning a single model, and hence will typically lead to greater efficiency. This recommendation is captured in Algorithm 1. If both stochastic and deterministic generative models are of interest, it is necessary to learn two models for most choices of the interpolant; we discuss more suggestions for stochastic case below.

然后构造估计 $\hat{b}(t,x) = \hat{v}(t,x) - \gamma(t)\dot{\gamma}(t)\hat{s}(t,x)$。上式中，$x_{t_i}^i = I(t_i, x_0^i, x_1^i) + \gamma(t_i)z^i$，$N$ 表示样本 t^i、x_0^i 和 x_1^i 的数量。如果使用者关心的是确定性生成模型，例如希望利用自适应积分或精确似然计算，那么直接学习估计 $\hat{b}$ 只需要训练一个模型，因而通常更加高效。算法 1 体现了这一建议。若同时关心随机与确定性生成模型，则对大多数插值选择而言，都必须学习两个模型；下面我们将进一步讨论随机情形的建议。

Antithetic sampling and capping.

对偶采样与截幅

S321 · Original / 原文 · PDF p.42 · S6.SS1.SSS0.Px2.p1.3

In practice, the losses for b (2.13) and s (2.16) can become high-variance near the endpoints $t = 0$ and $t = 1$ due to the presence of the singular term $1/\gamma(t)$ and the (potentially) singular term $\dot{\gamma}(t)$. This issue can be eliminated by using antithetic sampling, which we found necessary for stable training of objectives involving $\gamma^{-1}(t)$. To show why, we consider the loss (2.16) for s, but an analogous calculation can be performed for the loss (2.13) for b or (2.19) for η_z (even though it is not necessary for this last quantity). We first observe that, by definition of x_t and by Taylor expansion, as $t \to 0$ or $t \to 1$

在实践中，由于存在奇异项 $1/\gamma(t)$ 以及可能奇异的项 $\dot{\gamma}(t)$，b 的损失（2.13）和 s 的损失（2.16）在端点 $t = 0$、$t = 1$ 附近可能具有很高的方差。采用对偶采样可以消除这一问题；我们发现，对于含 $\gamma^{-1}(t)$ 的目标，对偶采样是实现稳定训练所必需的。为说明原因，我们考虑 s 的损失（2.16）；对 b 的损失（2.13）或 η_z 的损失（2.19）也可以进行类似计算，尽管最后一个量并不需要这种处理。首先，由 x_t 的定义及 Taylor 展开可见，当 $t \to 0$ 或 $t \to 1$ 时，

$$
\begin{aligned}
\frac{1}{\gamma(t)} z \cdot s(t, x_t) &= \frac{1}{\gamma(t)} z \cdot s(t, I(t, x_0, x_1) + \gamma(t) z), \\
&= \frac{1}{\gamma(t)} z \cdot \left(s(t, I(t, x_0, x_1)) + \gamma(t) \nabla s(t, I(t, x_0, x_1)) z + o(\gamma(t)) \right), \\
&= \frac{1}{\gamma(t)} z \cdot s(t, I(t, x_0, x_1)) + z \cdot \nabla s(t, I(t, x_0, x_1)) z + o(1).
\end{aligned} \tag{6.4}
$$

S322 · Original / 原文 · PDF p.42 · S6.SS1.SSS0.Px2.p1.4

Even though the conditional mean of the first term at the right-hand side is finite in the limit as $t \to 0$ or $t \to 1$, its variance diverges. By contrast, let $x_t^+ = I(t, x_0, x_1) + \gamma(t) z$ and $x_t^- = I(t, x_0, x_1) - \gamma(t) z$ with x_0, x_1, and z fixed. Then,

虽然右端第一项的条件均值在 $t \to 0$ 或 $t \to 1$ 的极限下有限，其方差却会发散。相较之下，固定 x_0, x_1 与 z，令 $x_t^+ = I(t, x_0, x_1) + \gamma(t) z$、$x_t^- = I(t, x_0, x_1) - \gamma(t) z$，则有：

$$
\begin{aligned}
&\frac{1}{2\gamma(t)} \left(z \cdot s(t, x_t^+) - z \cdot s(t, x_t^-) \right) \\
&= \frac{1}{2\gamma(t)} \left(z \cdot s(t, I(t, x_0, x_1) + \gamma(t) z) \right) - z \cdot s(t, I(t, x_0, x_1) - \gamma(t) z) \right), \\
&= \frac{1}{2\gamma(t)} z \cdot \left(s(t, I(t, x_0, x_1)) + \gamma(t) \nabla s(t, I(t, x_0, x_1)) z + o(\gamma(t)) \right) \\
&- \frac{1}{2\gamma(t)} z \cdot \left(s(t, I(t, x_0, x_1)) - \gamma(t) \nabla s(t, I(t, x_0, x_1)) z + o(\gamma(t)) \right), \\
&= z \cdot \nabla s(t, I(t, x_0, x_1)) z + o(1),
\end{aligned} \tag{6.5}
$$

S323 · Original / 原文 · PDF p.43 · S6.SS1.SSS0.Px2.p1.5

so that both the conditional mean and variance are finite in the limit as $t \to 0$ or $t \to 1$ despite the singularity of $1/\gamma(t)$. In practice, this can be implemented by using x_t^+ and x_t^- for every draw of x_0, x_1, and z in the empirical discretization of the population loss.

因此，尽管 $1/\gamma(t)$ 存在奇点，条件均值与条件方差在 $t \to 0$ 或 $t \to 1$ 的极限下都保持有限。实践中，在对总体损失作经验离散化时，每抽取一组 x_0, x_1 和 z，同时使用 x_t^+ 与 x_t^- 即可实现这一方法。

Algorithm 2 Learning η_z with arbitrary ρ_0 and ρ_1.

算法 2：任意端点密度下学习 ηz。

Input: Batch size N, interpolant function $I(t, x_0, x_1)$, a coupling $\nu(dx_0, dx_1)$ to sample x_0, x_1, noise function $\gamma(t)$, initial parameters θ_{η_z}, gradient-based optimization algorithm to minimize $\mathcal{L}_{\eta_z}$ defined in (2.19), number of gradient steps N_g.
Returns: An estimate $\hat{\eta}_z$ of η_z.
for $j = 1, \ldots, N_g$ do
  Draw N samples $(t_i, x_0^i, x_1^i, z^i) \sim \mathsf{Unif}([0,1]) \times \nu \times \mathsf{N}(0,1)$, for $i = 1, \ldots, N$.
  Construct samples $x_t^i = I(t_i, x_0^i, x_1^i) + \gamma(t_i) z^i$, for $i = 1, \ldots, N$.
  Take gradient step with respect to $\mathcal{L}_{\eta_z}\left(\theta_{\eta_z}, \{x_t^i\}_{i=1}^N\right)$.
Return: $\hat{\eta}_z$.

输入：批量 N、I、端点耦合 ν、γ、去噪器初始参数 θηz、最小化式（2.19）的梯度优化器、梯度步数 Ng。输出 ηz 的估计。每步抽取 N 组 $(t_i, x_0^i, x_1^i, z^i) \sim \mathrm{Unif}([0,1]) \times \nu \times \mathrm{N}(0,1)$，构造 $x_t^i = I(t_i, x_0^i, x_1^i) + \gamma(t_i) z^i$，再对 Lηz 做梯度更新，最终返回去噪器估计。

Learning the score s versus learning a denoiser η_z.

学习得分 s 或去噪器 ηz

S324 · Original / 原文 · PDF p.43 · S6.SS1.SSS0.Px3.p1.1

When learning s, an alternative to antithetic sampling is to consider learning the denoiser η_z defined in (2.17), which is related to the score by a factor of γ. Note that the objective function for the denoiser in (2.19) is well behaved for all $t \in [0,1]$, and can be thought of as a generalization of the DDPM loss introduced in *Ho et al. (2020)*. The empirical risk associated with this loss reads

学习 s 时，对偶采样的一种替代方法，是学习式（2.17）定义的去噪器 η_z；它与得分之间相差一个 γ 因子。注意，式（2.19）中去噪器的目标函数对所有 $t \in [0,1]$ 都具有良好的性质，可视为 Ho 等（2020）提出的 DDPM 损失的推广。与这一损失对应的经验风险为：

$$\hat{\mathcal{L}}_{\eta_z}(\hat{\eta}_z) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{2}|\hat{\eta}_z(t_i, x_{t_i}^i)|^2 - \hat{\eta}_z(t_i, x_{t_i}^i)\cdot z^i\right) \tag{6.6}$$

S325 · Original / 原文 · PDF p.43 · S6.SS1.SSS0.Px3.p1.2

A detailed procedure for learning the denoiser η_z, e.g. for its use in an SDE-based generative model is given in Algorithm 2. For the case of one-sided spatially-linear interpolants, the procedure becomes particularly simple, which is highlighted in Algorithm 3.

算法 2 给出了学习去噪器 η_z 的详细步骤，例如可将其用于基于 SDE 的生成模型。对单侧空间线性插值，这一过程尤其简单，算法 3 突出了这一点。

Algorithm 3 Learning η_z^{os} with Gaussian ρ_0.

算法 3：高斯基分布下学习单侧去噪器。

Input: Batch size N, interpolant function $x_t^{\mathrm{os,lin}}$, a coupling $\nu(dx_0, dx_1)$ to sample z, x_1, initial parameters $\theta_{\eta_z^{\mathrm{os}}}$, gradient-based optimization algorithm to minimize $\mathcal{L}_{\eta_z^{\mathrm{os}}}$ defined in (2.19), number of gradient steps N_g.
Returns: An estimate $\hat{\eta}_z^{\mathrm{os}}$ of η_z^{os}.
for $j = 1, \ldots, N_g$ do
  Draw N samples $(t_i, z^i, x_1^i) \sim \mathsf{Unif}([0,1]) \times \nu$, for $i = 1, \ldots, N$.
  Construct samples $x_t^i = \alpha(t_i)z + \beta(t_i)x_1^i$, for $i = 1, \ldots, N$.
  Take gradient step w.r.t $\mathcal{L}_{\eta_z^{\mathrm{os}}}\left(\theta_{\eta_z^{\mathrm{os}}}, \{x_t^i\}_{i=1}^N\right)$.
Return: $\hat{\eta}_z^{\mathrm{os}}$.

输入：批量 N、空间线性单侧插值、用于采样 z 与 x_1 的耦合 ν、单侧去噪器初始参数、相应梯度优化器与 Ng。输出单侧去噪器估计。每步采样 N 组 $(t_i, z^i, x_1^i) \sim \mathrm{Unif}([0,1]) \times \nu$，构造 $x_t^i = \alpha(t_i)z^i + \beta(t_i)x_1^i$，对单侧去噪损失做梯度更新，返回估计。译注：原算法构造式中的 z 未写批下标，按该步采样的 z^i 理解。

6.2 Sampling

6.2 采样

S326 · Original / 原文 · PDF p.43 · S6.SS2.p1.1

We now discuss several practical aspects of sampling generative models based on stochastic interpolants. These are intimately related to the choice of objects that are learned, as well as to the specific interpolant used to build a path between ρ_0 and ρ_1. A general algorithm for sampling models built on either ordinary or stochastic differential equations is presented in Algorithm 4.

现在讨论基于随机插值的生成模型在采样时的几个实践问题。这些问题与所学习对象的选择密切相关，也与用于构建 ρ_0 到 ρ_1 路径的具体插值有关。算法 4 给出了对基于常微分方程或随机微分方程的模型进行采样的通用算法。

Algorithm 4 Sampling general stochastic interpolants.

算法 4：一般随机插值的采样。

Input: Number of samples n, timestep Δt, drift estimates $\hat{b}$ and $\hat{\eta}_z$, initial time t_0, final time t_f, noise function $\gamma(t)$, diffusion coefficient $\epsilon(t)$, SDE or ODE timestepper TakeStep.
Returns: $\{\hat{x}_1^{(i)}\}_{i=1}^n$, a batch of model samples.
Initialize
Set time $t = t_0$.
Draw initial conditions $\hat{x}_{t_0}^{(i)} \sim \rho_0$ for $i = 1, \ldots, n$.
Construct $\hat{s}(t, x) = -\hat{\eta}_z(t, x)/\gamma(t)$.
Construct $\hat{b}_{\mathsf{F}}(t, x) = \hat{b}(t, x) + \epsilon(t)\hat{s}(t, x)$. // Reduces to $\hat{b}$ for $\epsilon(t) = 0$ (ODE).
while $t < t_f$ do
Propagate $\hat{x}_{t+\Delta t}^{(i)} = \texttt{TakeStep}(t, \hat{x}_t^{(i)}, b_{\mathsf{F}}, \epsilon, \Delta t)$ for $i = 1, \ldots, n$. // ODE or SDE integrator.
Update $t = t + \Delta t$.
Return: $\{\hat{x}^{(i)}\}_{i=1}^n$.

输入：样本数 n、步长 Δt、漂移 $\hat{b}$ 与去噪器 $\hat{\eta}_z$、初末时间 t_0、t_f、$\gamma(t)$、$\epsilon(t)$ 以及 ODE/SDE 单步积分器 TakeStep。输出一批模型样本。初始化 $t=t_0$，逐个采样 $\hat{x}_{t_0}^{i}\sim\rho_0$；构造 $\hat{s}=-\hat{\eta}_z/\gamma$，$\hat{b}_F=\hat{b}+\epsilon\hat{s}$（$\epsilon=0$ 时退化为 ODE 漂移 $\hat{b}$）。只要 $t<t_f$，就对所有样本调用 TakeStep$(t,\hat{x}_t^{i},\hat{b}_F,\epsilon,\Delta t)$，再令 $t\leftarrow t+\Delta t$；最后返回样本。

Using the denoiser η_z instead of the score s.

以 ηz 替代得分 s

S327 · Original / 原文 · PDF p.44 · S6.SS2.SSS0.Px1.p1.1

We remarked in Section 6.1 that learning the denoiser η_z is more numerically stable than learning the score s directly. We note that while the objective for η_z is well-behaved for all $t \in [0, 1]$, the resulting drifts can become singular at $t = 0$ and $t = 1$ when using $s(t, x) = -\eta_z(t, x)/\gamma(t)$. There are several ways to avoid this singularity in practice. One method is to choose a time-varying $\epsilon(t)$ that vanishes in a small interval around the endpoints $t = 0$ and $t = 1$, which avoids this numerical instability. An alternative option is to integrate the SDE up to a final time t_f with $t_f < 1$, and then to perform a step of denoising using (5.13). We use this approach in Section 7 below when sampling the SDE.

第 6.1 节曾指出，学习去噪器 η_z 比直接学习得分 s 的数值稳定性更好。我们注意到，虽然 η_z 的目标对所有 $t \in [0,1]$ 都具有良好性质，但在使用 $s(t,x) = -\eta_z(t,x)/\gamma(t)$ 时，得到的漂移可能在 $t=0$、$t=1$ 处出现奇点。实践中有多种方法可以避免这一奇异性。一种方法是选取随时间变化的 $\epsilon(t)$，使其在端点 $t=0$、$t=1$ 附近的一个小区间内为零，从而避免这种数值不稳定。另一种方法是把 SDE 积分到小于 1 的终止时刻 t_f，随后利用式（5.13）执行一次去噪。下面第 7 节对 SDE 进行采样时采用的就是后一种方法。

A denoiser is all you need for spatially-linear one-sided interpolants.

空间线性单侧插值只需去噪器

Algorithm 5 Sampling spatially-linear one-sided interpolants with Gaussian ρ_0.

算法 5：高斯基分布、空间线性单侧插值的采样。

Input: Number of samples n, timestep Δt, denoiser estimate $\hat{\eta}_z$, initial time t_0, final time t_f, noise function $\gamma(t)$, diffusion coefficient $\epsilon(t)$, interpolant functions $\alpha(t)$ and $\beta(t)$, SDE or ODE timestepper TakeStep.
Returns: $\{\hat{x}_1^{(i)}\}_{i=1}^n$, a batch of model samples.
Initialize
Set time $t = t_0$.
Draw initial conditions $\hat{x}_{t_0}^{(i)} \sim \rho_0$ for $i = 1, \ldots, n$.
Construct $\hat{s}(t,x) = -\hat{\eta}_z(t,x)/\alpha(t)$.
Construct $\hat{b}(t,x) = \dot{\alpha}(t)\hat{\eta}_z^{\text{os}}(t,x) + \frac{\dot{\beta}(t)}{\beta(t)}(x - \alpha(t)\hat{\eta}_z^{\text{os}}(t,x))$.
Construct $\hat{b}_{\text{F}}(t,x) = \hat{b}(t,x) + \epsilon(t)\hat{s}(t,x)$. // Reduces to $\hat{b}$ for $\epsilon(t) = 0$ (ODE).
while $t < t_f$ do
Propagate $\hat{x}_{t+\Delta t}^{(i)} = \texttt{TakeStep}(t, \hat{x}_t^{(i)}, b_{\text{F}}, \epsilon, \Delta t)$ for $i = 1, \ldots, n$. // ODE or SDE integrator.
Update $t = t + \Delta t$.
Return: $\{\hat{x}^{(i)}\}_{i=1}^n$.

输入：样本数 n、Δt、单侧去噪器估计、t₀、tf、γ、ε、α、β 及 TakeStep。输出一批模型样本。初始化 t=t₀，采样 x̂t₀ⁱ~ρ₀；构造 ŝ=−ηẑ/α，以及 b̂=α̇ηẑos+(β̇/β)(x−αηẑos)，再构造 b̂F=b̂+εŝ（ε=0 为 ODE）。循环至 tf：对每个样本调用同一单步积分器，再增加 t；最后返回样本。

S328 · Original / 原文 · PDF p.44 · S6.SS2.SSS0.Px2.p1.1

As shown in (4.19), and as considered in Section 5.2, the denoiser η_z^{os} is sufficient to represent the velocity field b appearing in the probability flow equation (2.32).

如式（4.19）所示，并如第 5.2 节所讨论，单侧去噪器 η_z^{os} 足以表示概率流方程（2.32）中出现的速度场 b。

S329 · Original / 原文 · PDF p.45 · S6.SS2.SSS0.Px2.p2.1

Using this definition for b and the relationship $s(t,x) = -\eta_z(t,x)/\gamma(t)$, we state the following ordinary and stochastic differential equations for sampling

利用这一 b 的定义以及关系 $s(t,x) = -\eta_z(t,x)/\gamma(t)$，可得到下列用于采样的常微分方程与随机微分方程：

$$\begin{aligned}
\text{ODE}: &\dot{X}_t = \dot{\alpha}(t)\eta_z^{\text{os}}(t,X_t) + \frac{\dot{\beta}(t)}{\beta(t)}\big(X_t - \alpha(t)\eta_z^{\text{os}}(t,X_t)\big) \\
\text{SDE}: &dX_t^{\text{F}} = \big(\dot{\alpha}(t)\eta_z^{\text{os}}(t,X_t^{\text{F}}) + \frac{\dot{\beta}(t)}{\beta(t)}\big(X_t^{\text{F}} - \alpha(t)\eta_z^{\text{os}}(t,X_t^{\text{F}})\big) - \frac{\epsilon(t)}{\alpha(t)}\eta_z^{\text{os}}(t,X_t^{\text{F}})\big)dt \\
&+\sqrt{2\epsilon(t)}dW_t.
\end{aligned} \tag{6.7}$$

S330 · Original / 原文 · PDF p.45 · S6.SS2.SSS0.Px2.p2.2

Because $\beta(0) = 0$, the drift is numerically singular in both equations. However, $b(t=0,x)$ has a finite limit

由于 $\beta(0) = 0$，两个方程中的漂移在数值上都存在奇点。然而，$b(t=0,x)$ 具有有限极限：

$$b(t=0,x) = \dot{\alpha}(0)x + \dot{\beta}(0)\mathbb{E}[x_1], \tag{6.8}$$

S331 · Original / 原文 · PDF p.45 · S6.SS2.SSS0.Px2.p2.3

as originally given in (4.20). Equation (6.8) can be estimated using available data, which means that when learning a one-sided interpolant, ODE and SDE-based generative models can be defined exactly on the interval $t \in [0,1]$ using only a score or a

denoiser without singularity.

这正是式（4.20）最初给出的结果。式（6.8）可以利用现有数据估计，这意味着，在学习单侧插值时，只使用得分或去噪器，就能在整个区间 $t \in [0,1]$ 上精确定义无奇异性的 ODE 与 SDE 生成模型。

S332 · Original / 原文 · PDF p.45 · S6.SS2.SSS0.Px2.p3.1

The factor of $\alpha(t)^{-1}$ in the final term of the SDE could pose numerical problems at $t = 1$, as $\alpha(1) = 0$. As discussed in the paragraph above, a choice of $\epsilon(t)$ which is such that $\epsilon(t)/\alpha(t) \to C$ for some constant C as $t \to 1$ avoids any issue.

SDE 最后一项中的 $\alpha(t)^{-1}$ 因子可能在 $t = 1$ 处造成数值问题，因为 $\alpha(1) = 0$。如前一段所讨论，选择 $\epsilon(t)$ 使得当 $t \to 1$ 时，$\epsilon(t)/\alpha(t) \to C$，其中 C 为某个常数，即可避免这一问题。

S333 · Original / 原文 · PDF p.45 · S6.SS2.SSS0.Px2.p4.1

An algorithm for sampling with only the denoiser η_z^{os} is given in Algorithm 5.

算法 5 给出了仅使用去噪器 η_z^{os} 进行采样的算法。

7 Numerical results

7 数值结果

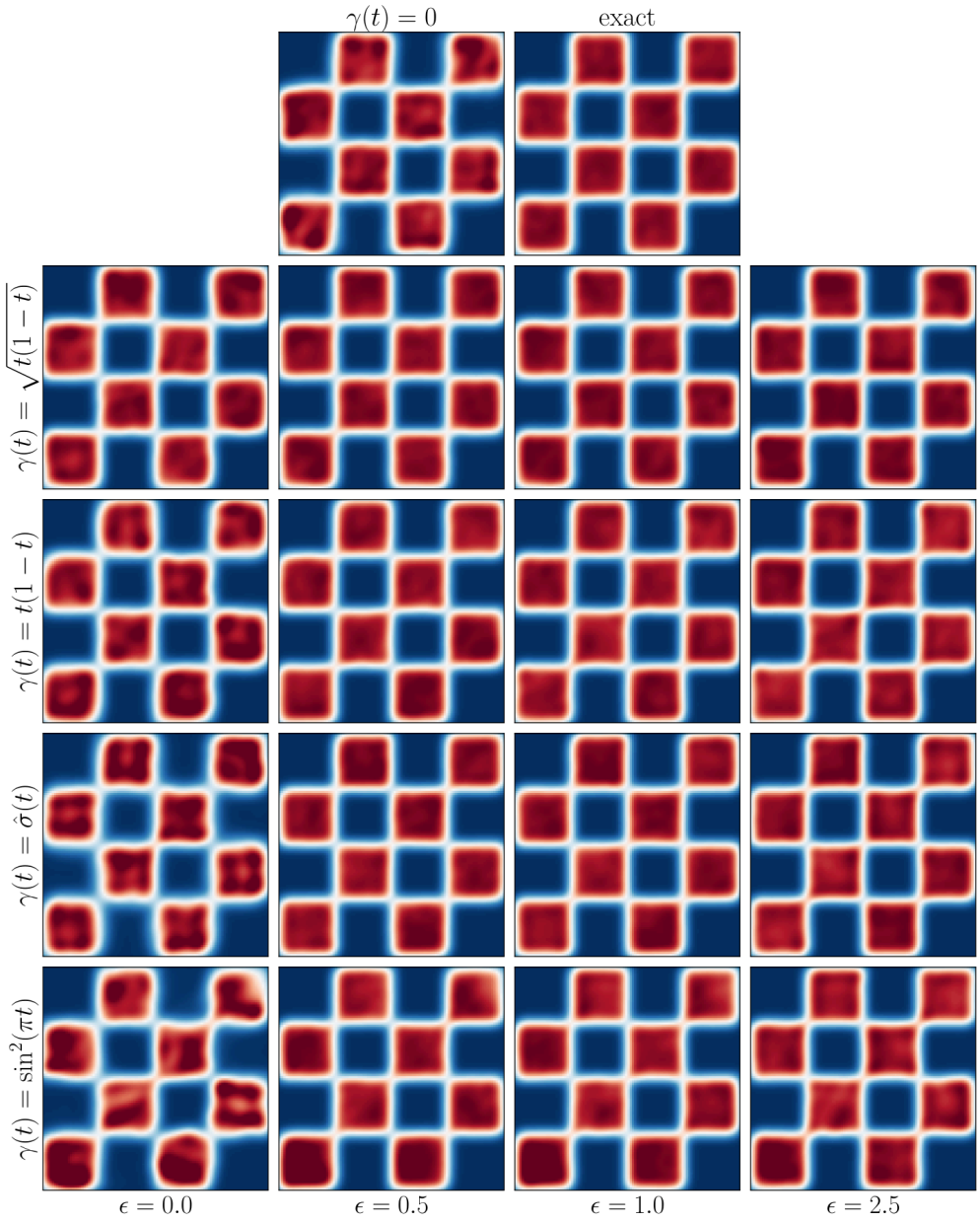

Figure 7: The effects of $\gamma(t)$ and ϵ on sample quality: qualitative comparison. Kernel density estimates of $\hat{\rho}(1)$ for models with different choices of γ and ϵ. Sampling with $\epsilon = 0$ corresponds to using the probability flow with the learned drift $\hat{b} = \hat{v} - \gamma\dot{\gamma}\hat{s}$, whereas sampling with $\epsilon > 0$ corresponds to using the SDE with the learned drift $\hat{b}$ and score $\hat{s}$. We find empirically that SDE sampling is generically better than ODE sampling for this target density, though the gap is smallest for the probability flow specified with $\gamma(t) = \sqrt{t(1-t)}$, in agreement with the Remark 42 regarding the influence of γ on b at the endpoints. The SDE performs well at any noise level, though numerically integrating it for higher ϵ requires a smaller step size.

图 7：γ 与 ε 对样本质量的定性影响。图示不同 γ、ε 下模型终态密度的核密度估计。ε=0 使用学得漂移 $\hat{b}=\hat{v}-\gamma\dot{\gamma}\hat{s}$ 的概率流；ε>0 使用 $\hat{b}$、$\hat{s}$ 的 SDE。对此目标，SDE 通常优于 ODE；γ=√[t(1−t)] 时差距最小，与说明 42 中端点得分作用相符。SDE 在各噪声水平都表现良好，但 ε 越大，数值积分步长越小。

S334 · Original / 原文 · PDF p.46 · S7.p1.1

So far, we have been focused on the impact of α, β, and γ in (4.1) on the density $\rho(t)$, which we illustrated analytically. In this section, we study examples where the drift coefficients must be learned over parametric function classes. In particular, we explore numerically the tradeoffs between generative models based on ODEs and SDEs, as well as the various design choices introduced in Sections 3, 4, and 6. In Section 7.1, we consider simple two-dimensional distributions that can be visualized

easily. In Section 7.2, we consider high-dimensional Gaussian mixtures, where we can compare our learned models to analytical solutions. Finally in Section 7.3 we perform some experiments in image generation.

到目前为止，我们关注的是式（4.1）中 α、β 与 γ 对密度 $\rho(t)$ 的影响，并以解析方式加以说明。本节研究必须在参数化函数类上学习漂移系数的例子。具体而言，我们通过数值实验探究基于 ODE 与 SDE 的生成模型之间的取舍，以及第 3、4、6 节引入的各种设计选择。第 7.1 节考虑易于可视化的简单二维分布。第 7.2 节考虑高维高斯混合分布，可将学到的模型与解析解比较。最后，第 7.3 节开展一些图像生成实验。

7.1 Deterministic versus stochastic models: 2D

7.1 二维确定性与随机模型

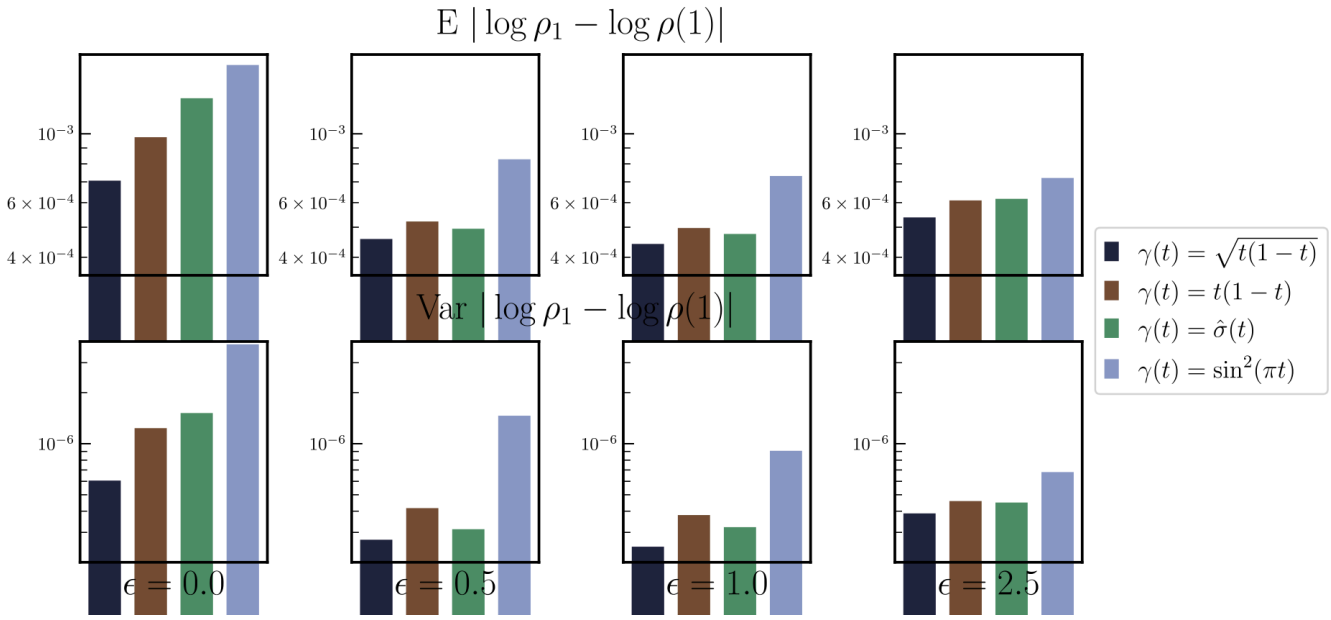

Figure 8: The effects of $\gamma(t)$ and ϵ on sample quality: quantitative comparison. For each γ and each ϵ specified in Figure 7, we compute the mean and variance of the absolute value of the difference of $\log \rho_1$ (exact) and $\log \hat{\rho}(1)$ (model). The model specified with $\gamma(t) = \sqrt{t(1-t)}$ is the best performing probability flow ($\epsilon = 0$). At large ϵ, SDE sampling with the same learned drift $\hat{b}$ and score $\hat{s}$ performs better, complementing the observations in the previous figure.

图 8：图 7 的定量比较。每个 γ、ε 下，计算真实 logρ₁ 与模型 logρ̂(1) 之差绝对值的均值、方差。γ=√[t(1−t)] 的概率流（ε=0）最佳；较大 ε 时，使用同样 b̂、ŝ 的 SDE 更好，与前图互补。

S335 · Original / 原文 · PDF p.47 · S7.SS1.p1.1

As shown in Section 2.2, the evolution of $\rho(t)$ can be captured exactly by either the transport equation (2.9) or by the forward and backward Fokker-Planck equations (2.20) and (2.22). These perspectives lead to generative models that are either based on the deterministic dynamics (2.32) or the forward and backward stochastic dynamics (2.33) and (2.34), where the level of stochasticity can be tuned by varying the diffusion coefficient $\epsilon(t)$. We showed in Section 2.4 that setting a constant $\epsilon(t) = \epsilon > 0$ can offer better control on the likelihood when using an imperfect velocity b and an imperfect score s. Moreover, the optimal choice of ϵ is determined by the relative accuracy of the estimates $\hat{b}$ and $\hat{s}$. Having laid out the evolution of $\rho(t)$ for different choices of γ in the previous section, we now show how different values of ϵ can build these densities up from individual trajectories. The stochasticity intrinsic to the sampling process increases with ϵ, but by construction, the marginal density $\rho(t)$ for fixed α, β and γ is independent of ϵ.

如第 2.2 节所示，密度 $\rho(t)$ 的演化既可由输运方程（2.9）精确刻画，也可由前向与后向 Fokker–Planck 方程（2.20）、（2.22）精确刻画。这些视角分别导出基于确定性动力学（2.32）或前向、后向随机动力学（2.33）、（2.34）的生成模型，其中随机性水平可通过改变扩散系数 $\epsilon(t)$ 来调节。第 2.4 节表明，采用常数 $\epsilon(t) = \epsilon > 0$，在速度 b 和得分 s 不精确时，可以更好地控制似然。此外，最优 ϵ 的选择取决于估计 $\hat{b}$ 与 $\hat{s}$ 的相对精度。上一节已说明不同 γ 选择下 $\rho(t)$ 的演化；现在我们展示不同 ϵ 值如何通过单条轨迹构建这些密度。采样过程内在的随机性随 ϵ 增大而增强，但按照构造，对于固定的 α、β 和 γ，边缘密度 $\rho(t)$ 与 ϵ 无关。

The roles of $\gamma(t)$ and ϵ for 2D density estimation.

二维密度估计中 γ、ε 的作用

S336 · Original / 原文 · PDF p.47 · S7.SS1.SSS0.Px1.p1.1

To explore the roles of γ and ϵ, we consider a target density ρ_1 whose mass concentrates on a two-dimensional checkerboard and a base density $\rho_0 = \mathsf{N}(0, Id)$; here, the target was chosen to highlight the ability of the method to learn a challenging density with sharp boundaries. The same model architecture and training procedure was used to learn both v and s for several choices of γ given in Table 2. The feed-forward network was defined with 4 layers, each of size 512, and with the ReLU *Nair and Hinton (2010)* as an activation function.

为探究 γ 与 ϵ 的作用，我们考虑质量集中于二维棋盘格上的目标密度 ρ_1，以及基准密度 $\rho_0 = \mathsf{N}(0, Id)$。选择这一目标，是为了突出该方法学习具有尖锐边界的复杂密度的能力。对于表 2 给出的若干 γ 选择，我们采用相同的模型架构和训练步骤学习 v 与 s。前馈网络包含 4 层，每层宽度为 512，并以 ReLU（Nair 与 Hinton，2010）作为激活函数。

S337 · Original / 原文 · PDF p.48 · S7.SS1.SSS0.Px1.p2.1

After training, we draw 300,000 samples using either an ODE ($\epsilon = 0$) or an SDE with $\epsilon = 0.5$, $\epsilon = 1.0$, or $\epsilon = 2.5$. We compute kernel density estimates for each resulting density, which we compare to the exact density and to the original stochastic interpolant from *Albergo and Vanden-Eijnden (2023)* (obtained by setting $\gamma = 0$). Results are given in Figure 7 for each γ and each ϵ. Sampling with $\epsilon > 0$ empirically performs better, though the gap is smallest when using the γ specified in (4.13). Moreover, even when $\epsilon = 0$, using the probability flow with γ given by (4.13) performs better than the original interpolant from *Albergo and Vanden-Eijnden (2023)*. Numerical comparisons of the mean and variance of the absolute value of the difference of $\log \rho_1$ (exact) from $\log \hat{\rho}(1)$ (model) for the various configurations are given in Figure 8, which corroborate the above observations.

训练完成后，我们使用 ODE（$\epsilon = 0$）或扩散系数分别为 $\epsilon = 0.5$、$\epsilon = 1.0$、$\epsilon = 2.5$ 的 SDE 抽取 300,000 个样本。我们对得到的各个密度计算核密度估计，并与精确密度及 Albergo 与 Vanden-Eijnden（2023）的原始随机插值进行比较，后者通过设置 $\gamma = 0$ 得到。图 7 给出了每种 γ 与 ϵ 组合的结果。经验上，$\epsilon > 0$ 时采样效果更好，但采用式（4.13）规定的 γ 时，两者差距最小。此外，即使 $\epsilon = 0$，采用式（4.13）中 γ 的概率流，也优于 Albergo 与 Vanden-Eijnden（2023）的原始插值。图 8 给出了各配置下 $\log \rho_1$（精确值）与 $\log \hat{\rho}(1)$（模型值）之差的绝对值的均值和方差比较，定量支持了上述观察。

7.2 Deterministic versus stochastic models: 128D Gaussian mixtures

7.2 128维高斯混合模型

S338 · Original / 原文 · PDF p.48 · S7.SS2.p1.1

We now study the performance of the stochastic interpolant method in the case where the target is a high-dimensional Gaussian mixture. Gaussian mixtures (GMs) are a convenient class of target distributions to study, because they can be made arbitrarily complex by increasing the number of modes, their separation, and the overall dimensionality. Moreover, by considering low-dimensional projections, we can compute quantitative error metrics such as the KL-divergence between the target and the model as a function of the (constant) diffusion coefficient ϵ. This enables us to quantify the tradeoffs of ODE and SDE-based samplers.

现在研究目标为高维高斯混合分布时随机插值方法的性能。高斯混合（GM）是一类便于研究的目标分布，因为通过增加模态数、模态间距和整体维度，可以使其复杂程度任意提高。此外，考察低维投影后，我们可以计算目标与模型之间的 KL 散度等定量误差指标，并将其视为常数扩散系数 ϵ 的函数。这使我们能够量化 ODE 与 SDE 采样器之间的取舍。

Experimental details.

实验细节

S339 · Original / 原文 · PDF p.48 · S7.SS2.SSS0.Px1.p1.1

We consider the problem of mapping $\rho_0 = \mathsf{N}(0, Id)$ to a Gaussian mixture with five modes in dimension $d = 128$. The mean $m_i \in \mathbb{R}^d$ of each mode is drawn i.i.d. $m_i \sim \mathsf{N}(0, \sigma^2 Id)$ with $\sigma = 7.5$. To maximize performance at high ϵ, the timestep should be adapted to ϵ; here, we chose to use a fixed computational budget that performs well for moderate levels of ϵ to avoid computational effort that may become unreasonable in practice.

我们考虑将 $\rho_0 = \mathsf{N}(0, Id)$ 映射到维度为 $d = 128$、具有五个模态的高斯混合分布。各模态的均值 $m_i \in \mathbb{R}^d$ 独立同分布地抽取自 $m_i \sim \mathsf{N}(0, \sigma^2 Id)$，其中 $\sigma = 7.5$。要在较大的 ϵ 下最大化性能，应当使时间步长适应 ϵ；这里我们选择一个在中等 ϵ 水平下表现良好的固定计算预算，以避免投入实践中可能变得不合理的计算量。

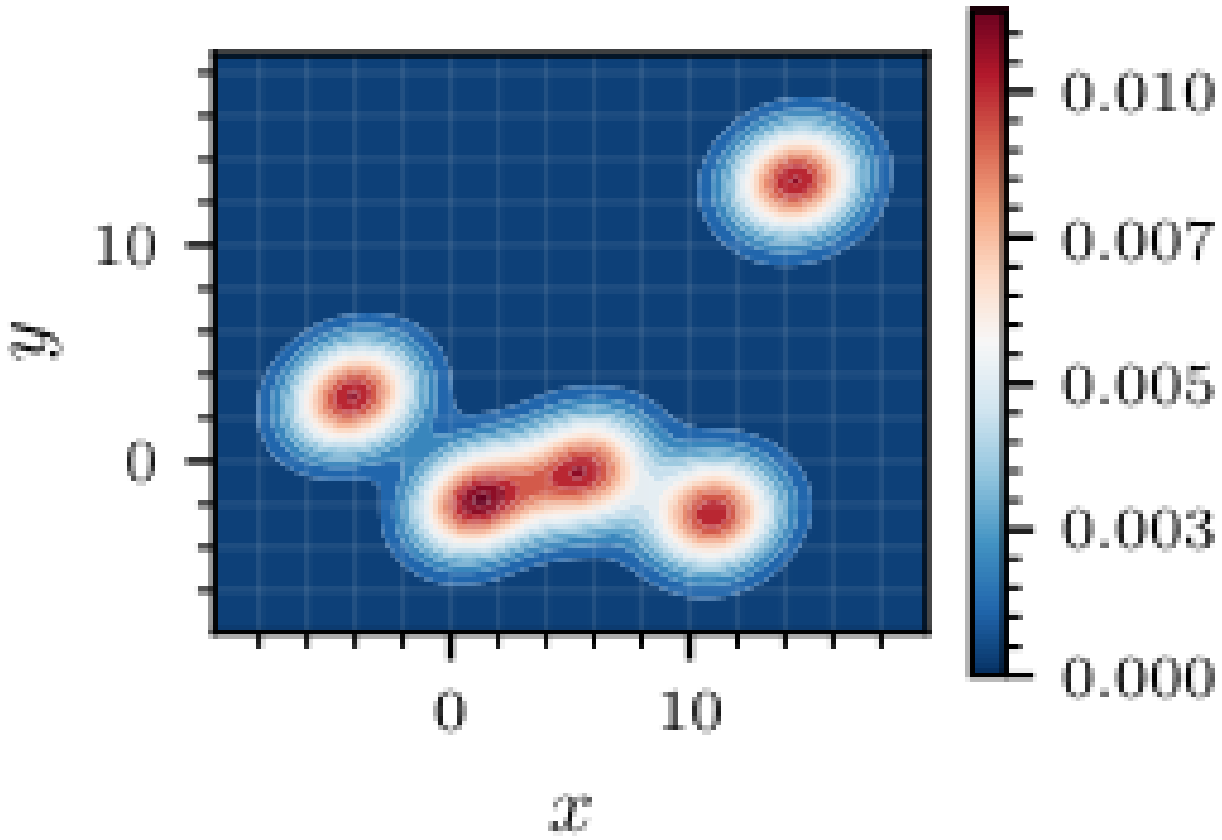

Figure 9: Gaussian mixtures: target projection. Low-dimensional marginal of the target density ρ_1 for the Gaussian mixture experiment, visualized via KDE.

图 9：高斯混合实验目标密度的低维边缘，用 KDE 可视化。

S340 · Original / 原文 · PDF p.48 · S7.SS2.SSS0.Px1.p2.1

Each covariance $C_i \in \mathbb{R}^{d \times d}$ is also drawn randomly with $C_i = \frac{1}{d} W_i^{\mathsf{T}} W_i + Id$ and $(W_i)_{kl} \sim \mathsf{N}(0, 1)$ for $k, l = 1, \ldots, d$; this choice ensures that each covariance is a positive definite perturbation of the identity with diagonal entries that are $O(1)$ with respect to dimension d. For a fixed random draw of the means $\{m_i\}_{i=1}^5$ and covariances $\{C_i\}_{i=1}^5$, we study the four combinations of learning b or v and s or η_z to form a stochastic interpolant from ρ_0 to ρ_1. In each case, we consider a linear interpolant with $\alpha(t) = 1 - t$, $\beta(t) = t$, and $\gamma(t) = \sqrt{t(1-t)}$. For visual reference, a projection of the target density ρ_1 onto the first two coordinates is depicted in Figure 9 – it contains significant multimodality, several modes that are difficult to distinguish, and one mode that is well-separated from the others, which requires nontrivial transport to resolve. In the following experiments, all samples were generated with the fourth-order Dormand-Prince adaptive ODE solver (dopri5) for $\epsilon = 0$ and by using one thousand timesteps of the Heun SDE integrator introduced in *Karras et al. (2022)* for $\epsilon \neq 0$. When learning η_z, to avoid singularity at $t = 0$ and $t = 1$ when dividing by $\gamma(t)$ in the formula $s(t, x) = -\eta(t, x)/\gamma(t)$, we set $t_0 = 10^{-4}$ and $t_f = 1 - t_0$ in Algorithm 4. For all other cases, we set $t_0 = 0$ and $t_f = 1$.

各协方差 $C_i \in \mathbb{R}^{d \times d}$ 也随机生成，取 $C_i = \frac{1}{d} W_i^{\mathsf{T}} W_i + Id$，其中对 $k, l = 1, \ldots, d$，有 $(W_i)_{kl} \sim \mathsf{N}(0, 1)$。这一选择保证各协方差都是单位矩阵的正定扰动，且其对角元素相对于维度 d 为 $O(1)$。固定随机抽得的均值 $\{m_i\}_{i=1}^5$ 与协方差

$\{C_i\}_{i=1}^5$ 后，我们研究学习 b 或 v，以及学习 s 或 η_z 的四种组合，以构造从 ρ_0 到 ρ_1 的随机插值。所有情形都采用线性插值，其中 $\alpha(t) = 1 - t$、$\beta(t) = t$、$\gamma(t) = \sqrt{t(1-t)}$。为便于直观参照，图 9 绘出了目标密度 ρ_1 在前两个坐标上的投影：它具有明显的多模态性，其中若干模态难以区分，而有一个模态与其余模态相距较远，需要非平凡的传输才能正确恢复。在以下实验中，$\epsilon = 0$ 时，所有样本均由四阶 Dormand–Prince 自适应 ODE 求解器（dopri5）生成；$\epsilon \neq 0$ 时，则采用 Karras 等（2022）引入的 Heun SDE 积分器，以一千个时间步生成样本。学习 η_z 时，为避免在公式 $s(t,x) = -\eta(t,x)/\gamma(t)$ 中除以 $\gamma(t)$ 导致 $t = 0$、$t = 1$ 处出现奇点，我们在算法 4 中设置 $t_0 = 10^{-4}$、$t_f = 1 - t_0$。其余情形均设置 $t_0 = 0$、$t_f = 1$。

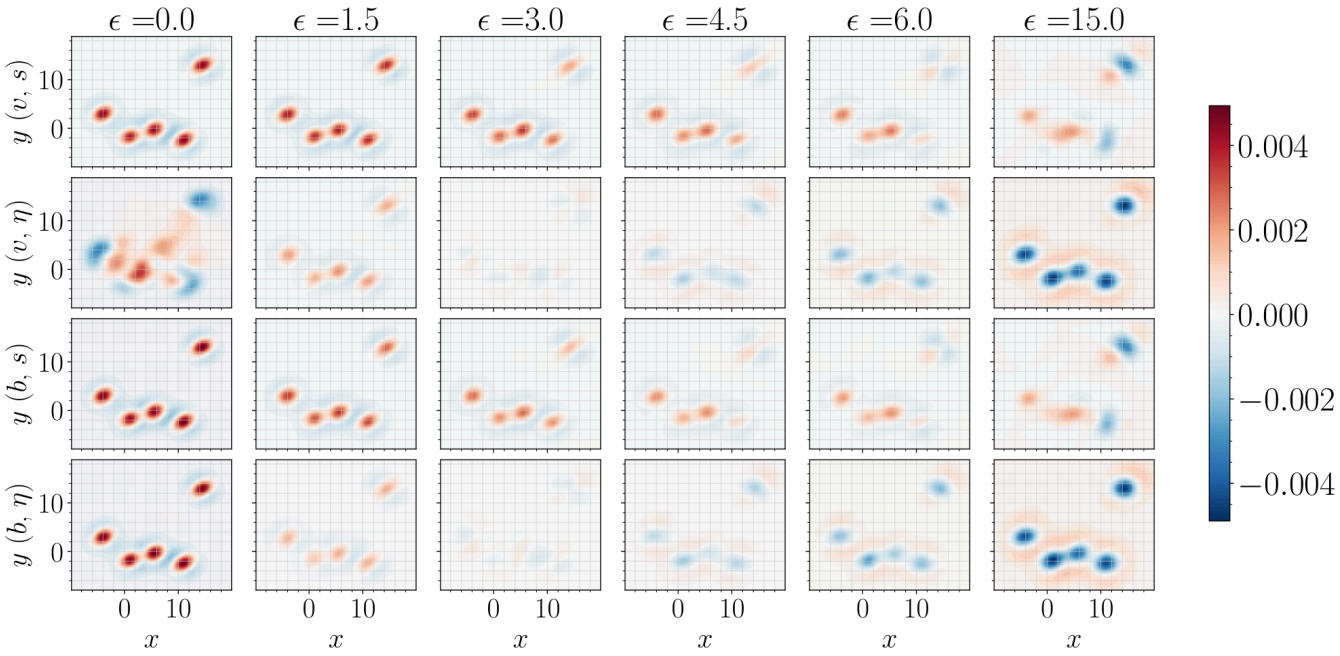

Figure 10: Gaussian mixtures: density errors. Errors $\hat{\rho}_1(x,y) - \rho_1(x,y)$ in the marginals over the first two coordinates for all four variations of learning b or v and s or η, computed via kernel density estimation. For small ϵ, the model densities tend to be overly-concentrated, and overestimate the density within the modes and underestimate the densities in the tails. As ϵ increases, the model becomes less concentrated and more accurately represents the target density. For ϵ too large, the model becomes overly spread out and under-estimates the density within the modes. Note: visualizations show two-dimensional slices of a 128-dimensional density.

图 10：高斯混合的密度误差。四种学习组合（b 或 v，s 或 η）的前两坐标边缘误差 $\hat{\rho}_1(x,y) - \rho_1(x,y)$，以 KDE 估计。小 ε 时过度集中，峰内高估、尾部低估；增大 ε 改善这一问题；过大时又过分分散、低估峰内密度。图示为 128 维密度的二维切片。

Quantitative metric

定量指标

S341 · Original / 原文 · PDF p.49 · S7.SS2.SSS0.Px2.p1.1

To quantify performance, we make use of an error metric given by a KL-divergence between kernel density estimates (KDE) of low-dimensional marginals of ρ_1 and the model density $\hat{\rho}_1$; this error metric was chosen for computational tractability and interpretability. To compute it, we draw $50,000$ samples from ρ_1 and each $\hat{\rho}_1$. We obtain samples from the marginal density over the first two coordinates by projection, and then compute a Gaussian KDE with bandwidth parameter chosen by Scott's rule. We then draw a fresh set of $N_e = 50,000$ samples $\{x_i\}_{i=1}^{N_e}$ with each $x_i \sim \rho_1$ for evaluation. To compute the KL-divergence, we form a Monte-Carlo estimate with control variate

为量化性能，我们使用一个误差指标：目标密度 ρ_1 与模型密度 $\hat{\rho}_1$ 的低维边缘分布的核密度估计（KDE）之间的 KL 散度。选择这一误差指标，是出于计算可行性与可解释性的考虑。计算时，我们分别从 ρ_1 与各个 $\hat{\rho}_1$ 中抽取 50,000 个样本；通过投影获得前两个坐标边缘密度的样本，再以 Scott 法则选择带宽，计算高斯 KDE。随后，我们另外抽取一组新的 $N_e = 50,000$ 个样本 $\{x_i\}_{i=1}^{N_e}$ 用于评估，其中每个 $x_i \sim \rho_1$。为计算 KL 散度，我们构造含控制变量的 Monte Carlo 估计：

$$\mathrm{KL}(\rho_1 \,\|\, \hat{\rho}_1) \approx \frac{1}{N_e} \sum_{i=1}^{N_e} \left(\log \rho_1(x_i) - \log \hat{\rho}_1(x_i) - \left(\frac{\hat{\rho}_1(x_i)}{\rho_1(x_i)} - 1 \right) \right). \tag{7.1}$$

We found use of the control variate $\hat{\rho}_1/\rho_1 - 1$ helpful to reduce variance in the Monte-Carlo estimate; moreover, by concavity of the logarithm, use of the control variate ensures that the Monte-Carlo estimate cannot become negative.

我们发现，采用控制变量 $\hat{\rho}_1/\rho_1 - 1$ 有助于减小 Monte Carlo 估计的方差；此外，由于对数函数的凹性，使用这一控制变量可保证 Monte Carlo 估计不会变为负值。

Results.

结果

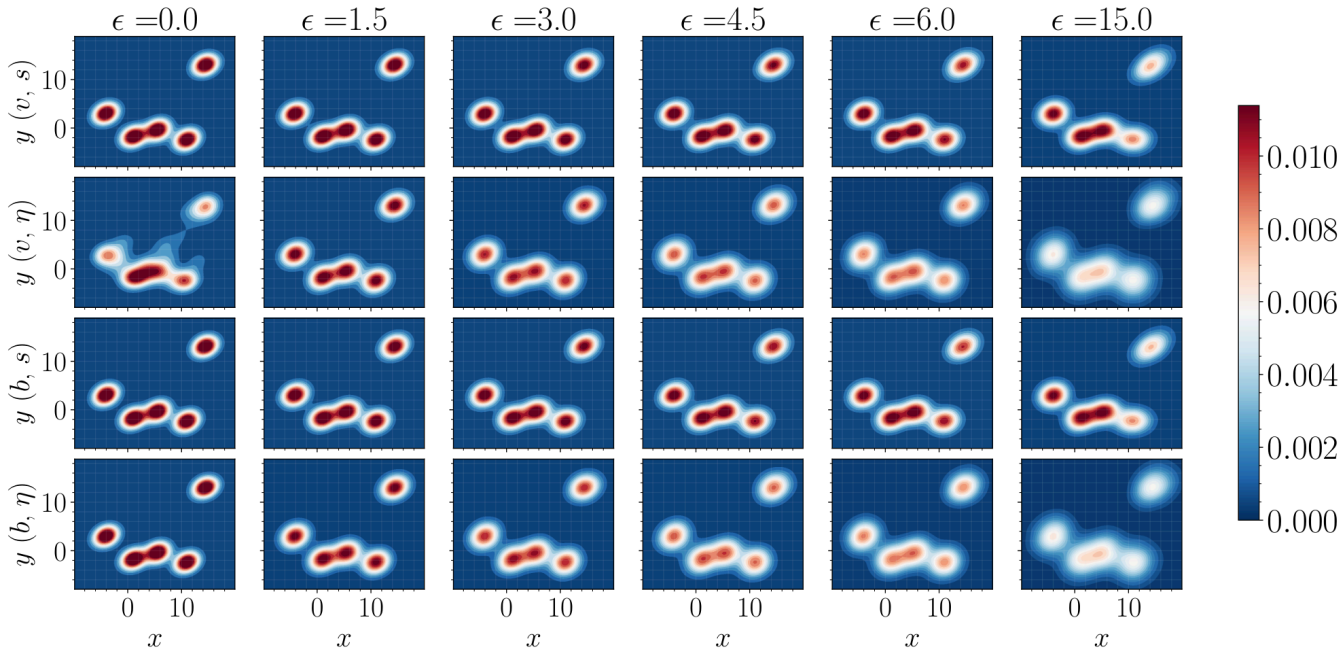

Figure 11: Gaussian mixtures: densities. A visualization of the marginal densities in the first two variables of the model density $\hat{\rho}_1$ computed for all four variations of learning b or v and s or η, computed via kernel density estimation. Note: visualizations show two-dimensional slices of a 128-dimensional density.

图 11：高斯混合密度。四种学习组合所得模型密度之前两变量边缘，使用 KDE。图示为 128 维密度的二维切片。

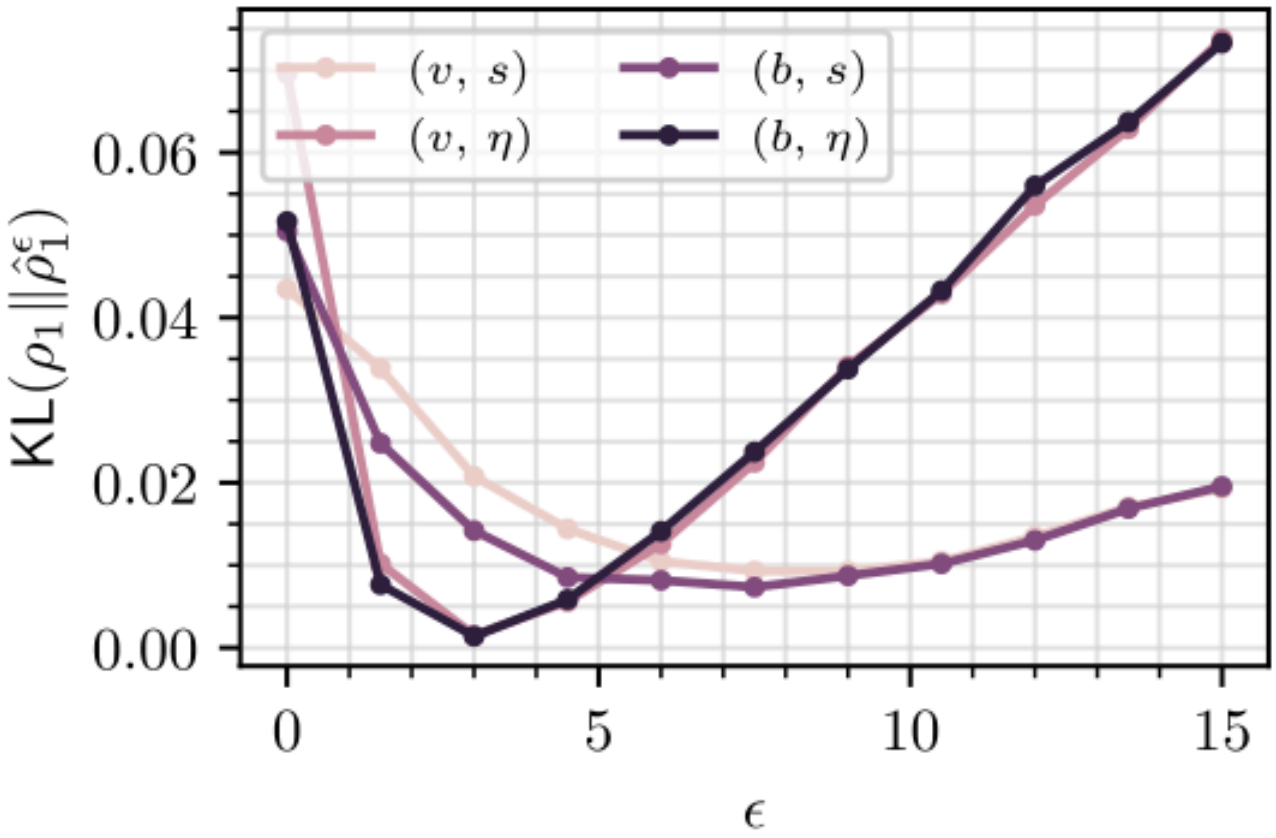

Figure 12: Gaussian mixtures: quantitative comparison. $\mathrm{KL}(\rho_1 \parallel \hat{\rho}_1^\epsilon)$ as a function of ϵ when learning each of the four possible sets of drift coefficients (b, s), (b, η), (v, s), (v, η). The best performance is achieved when learning b and η, along with a proper choice of $\epsilon > 0$.

图 12：高斯混合的定量比较。分别学习 (b,s)、(b,η)、(v,s)、(v,η) 后，KL(ρ₁ ‖ ρ̂₁ᵋ) 随 ε 变化。学习 b、η 并恰当选择 ε>0 时最佳。

S343 · Original / 原文 · PDF p.49 · S7.SS2.SSS0.Px3.p1.1

Figures 10 and 11 display two-dimensional projections (computed via KDE) of the model density error $\hat{\rho}_1 - \rho_1$ and the model density $\hat{\rho}_1$ itself, respectively, for different instantiations of Algorithms 1 and 2 and different choices of ϵ in Algorithm 4. Taken together with Figure 9, these results demonstrate qualitatively that small values of ϵ tend to over-estimate the density within the modes and under-estimate the density in the tails. Conversely, when ϵ is taken too large, the model tends to under-estimate the modes and over-estimate the tails. Somewhere in between (and for differing levels of ϵ), every model obtains its optimal performance. Figure 12 makes these observations quantitative, and displays the KL-divergence from the target marginal to the model marginal $\mathrm{KL}(\rho_1 \parallel \hat{\rho}_1^\epsilon)$ as a function of ϵ, with each data point on the curve matching the models depicted in Figures 10 and 11. We find that for each case, there is an optimal value of $\epsilon \neq 0$, in line with the qualitative picture put forth by Figures 10 and 11. Moreover, we find that learning b generically performs better than learning v, and that learning η generically performs better than learning s (except when ϵ is taken large enough that performance starts to degrade). With proper treatment of the singularity in the sampling algorithm when using the denoiser in the construction of $s(t, x) = -\eta(t, x)/\gamma(t)$ – either by capping $t_0 \neq 0$ and $t_f \neq 1$ or by properly tuning $\epsilon(t)$ as discussed in Section 6.2 – our results suggest that learning the denoiser is best practice.

图 10 与图 11 分别展示模型密度误差 $\hat{\rho}_1 - \rho_1$ 和模型密度 $\hat{\rho}_1$ 本身的二维投影，投影均通过 KDE 计算；图中比较了算法 1、2 的不同实现，以及算法 4 中不同的 ϵ 选择。结合图 9，这些结果定性表明：较小的 ϵ 往往会高估模态内部的密度、低估尾部密度；相反，ϵ 过大时，模型往往低估模态密度、高估尾部密度。在两者之间，各模型会在不同的 ϵ 水平上达到各自最优性能。图 12 对这些观察作了定量说明，绘出目标边缘分布到模型边缘分布的 KL 散度 $\mathrm{KL}(\rho_1 \parallel \hat{\rho}_1^\epsilon)$ 随 ϵ 的变化；曲线上每个数据点都对应图 10、11 所示的模型。我们发现，每种情形都存在最优的非零值 $\epsilon \neq 0$，与图 10、11 所呈现的定性图景一致。此外，学习 b 总体上优于学习 v，学习 η 总体上优于学习 s，但当 ϵ 大到性能开始退化时，后一比较有例外。用去噪器构造 $s(t, x) = -\eta(t, x)/\gamma(t)$ 时，只要妥善处理采样算法中的奇

点——或将起止时刻截为 $t_0 \neq 0$、$t_f \neq 1$，或按照第 6.2 节所述适当调节 $\epsilon(t)$——我们的结果就表明，学习去噪器是最佳实践。

7.3 Image generation

7.3 图像生成

S344 · Original / 原文 · PDF p.51 · S7.SS3.p1.1

In the following, we demonstrate that the proposed method scales straightforwardly to high-dimensional problems like image generation. To this end, we illustrate the use of our approach on the 128×128 Oxford flowers dataset *Nilsback and Zisserman (2006)* by testing two different variations of the interpolant for image generation: the one-sided interpolant, using $\rho_0 = \mathsf{N}(0, Id)$, as well as the mirror interpolant, where $\rho_0 = \rho_1$ both represent the data distribution. The purpose of this section is to demonstrate that our theory is well-motivated, and that it provides a framework that is both scalable and flexible. In this regard, image generation is a convenient exercise, but is not the main focus of this work, and we will leave a more thorough study on other datasets such as ImageNet with standard benchmarks such as the Frechet Inception Distance (FID) for a future study.

下面展示，所提出的方法可以直接扩展到图像生成等高维问题。为此，我们在 128×128 的 Oxford flowers 数据集（Nilsback 与 Zisserman，2006）上说明本方法的应用，并测试用于图像生成的两种插值变体：其一是采用 $\rho_0 = \mathsf{N}(0, Id)$ 的单侧插值；其二是镜像插值，其中 $\rho_0 = \rho_1$ 都表示数据分布。本节旨在表明我们的理论有充分动机，并提供了一个既可扩展又灵活的框架。就这一目标而言，图像生成是一个方便的检验任务，而不是本工作的主要重点；因此，在 ImageNet 等其他数据集上采用 Fréchet Inception Distance（FID）等标准基准开展更全面的研究，留待未来工作。

Generation from Gaussian ρ_0.

从高斯基分布生成

S345 · Original / 原文 · PDF p.51 · S7.SS3.SSS0.Px1.p1.1

We train spatially-linear one-sided interpolants $x_t = (1-t)z + tx_1$ and $x_t = \cos(\frac{\pi}{2}t)z + \sin(\frac{\pi}{2}t)x_1$ on the 128×128 Oxford flowers dataset, where we take $z \sim \mathsf{N}(0, Id)$ and x_1 from the data distribution. Based on our results for Gaussian mixtures, we learn the drift $b(t,x)$, the score $s(t,x)$, and the denoiser $\eta_z(t,x)$ to benchmark our generative models based on ODEs or SDEs. In all cases, we parameterize the networks representing $\hat{\eta}$, $\hat{s}$ and $\hat{b}$ using the U-Net architecture used in *Ho et al. (2020)*. Minimization of the objective functions given in Section 3.2 is performed using the Adam optimizer. Details of the architecture in both cases and all training hyperparameters are provided in Appendix C.1.

我们在 128×128 的 Oxford flowers 数据集上训练空间线性的单侧插值 $x_t = (1-t)z + tx_1$ 与 $x_t = \cos(\frac{\pi}{2}t)z + \sin(\frac{\pi}{2}t)x_1$，其中取 $z \sim \mathsf{N}(0, Id)$，x_1 来自数据分布。根据高斯混合实验的结果，我们学习漂移 $b(t,x)$、得分 $s(t,x)$ 和去噪器 $\eta_z(t,x)$，以比较基于 ODE 或 SDE 的生成模型。所有情形中，表示 $\hat{\eta}$、$\hat{s}$ 和 $\hat{b}$ 的网络都采用 Ho 等（2020）使用的 U-Net 架构进行参数化。第 3.2 节给出的目标函数使用 Adam 优化器最小化。两种情形的架构细节及全部训练超参数见附录 C.1。

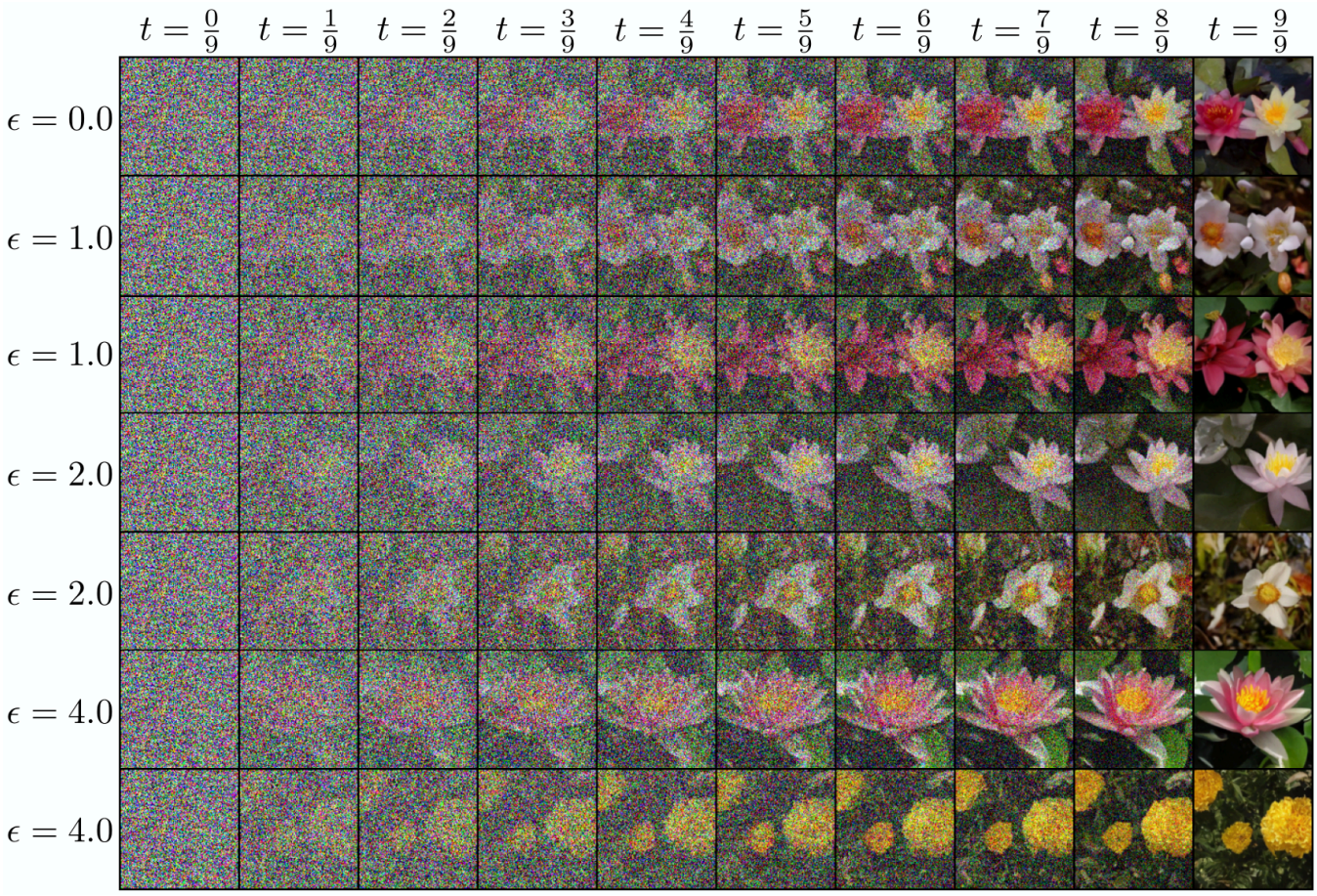

Figure 13: Image generation: Oxford flowers. Example generated flowers from the same initial condition x_0 using either the ODE with $\epsilon = 0$ and learned b or the SDE for various increasing values of ϵ with learned b and s. For $\epsilon = 0$, sampling is done using the dopri5 solver and therefore the number of steps is adaptive. Otherwise, 2000, 2500, and 4000 steps were taken using the Heun solver for $\epsilon =$ 1.0, 2.0, and 4.0 respectively.

图 13：Oxford flowers 图像生成。固定同一初值 x₀，比较 ε=0、学习 b 的 ODE，与不同增大 ε、学习 b 和 s 的 SDE。ODE 用 dopri5，自适应选择步数；SDE 的 ε=1、2、4 分别用 Heun 积分 2000、2500、4000 步。

S346 · Original / 原文 · PDF p.51 · S7.SS3.SSS0.Px1.p2.1

Like in the case of learning Gaussian mixtures, we use the fourth-order dopri5 solver when sampling with the ODE and the Heun method for the SDE, as detailed in Algorithm 4. When learning a denoiser η_z, we found it beneficial to complete the image generation with a final denoising step, in which we set $\epsilon = 0$ and switch the integrator to the one given in (5.14).

与学习高斯混合分布时一样，我们在使用 ODE 采样时采用四阶 dopri5 求解器，在使用 SDE 采样时采用 Heun 方法，具体见算法 4。学习去噪器 η_z 时，我们发现，最后增加一次去噪步骤以完成图像生成是有益的；在这一步中，设置 $\epsilon = 0$，并将积分器切换为式（5.14）给出的积分器。

S347 · Original / 原文 · PDF p.51 · S7.SS3.SSS0.Px1.p3.1

Exemplary images generated from the model using the ODE and the SDE with various value of the diffusion coefficient ϵ are shown in Figure 13, starting from the same sample from ρ_0. The result illustrates that different images can be generated from the same sample when using the SDE, and their diversity increases as we increase the diffusion coefficient ϵ. To highlight that the model does not memorize the training set, in Figure 14 we compare an example generated image to its five nearest neighbors in the training set (measured in ℓ_1 norm), which appear visually quite different.

图 13 展示从相同的 ρ_0 样本出发，分别使用 ODE 和取不同扩散系数 ϵ 的 SDE，由模型生成的图像示例。结果表明，使用 SDE 时，同一个初始样本可以生成不同图像，而且随着扩散系数 ϵ 增大，图像多样性也会增加。为突出表明模型没有记忆训练集，图 14 将一张生成图像与其在训练集中的五个最近邻作比较，最近邻以 ℓ_1 范数度量；这些邻居在视觉上与生成图像很不相同。



Figure 14: No memorization. Top: An example generated image from a trained linear one-sided interpolant. Bottom: Five nearest neighbors (in ℓ_1 norm) to the generated image in the dataset. The nearest neighbors are visually distinct, highlighting that the interpolant does not overfit on the dataset.

图 14：未记忆样例。上图是训练好的线性单侧插值生成图；下图是数据集中按 L1 范数最近的五个邻居。邻居在视觉上明显不同，作者以此说明模型未对数据集过拟合。

Mirror interpolant.

镜像插值

S348 · Original / 原文 · PDF p.51 · S7.SS3.SSS0.Px2.p1.1

We consider the mirror interpolant $x_t = x_1 + \gamma(t)z$, for which (3.33) shows that the drift b is given in terms the denoiser η_z by $b(t,x) = \dot{\gamma}(t)\eta_z(t)$; this means that it is sufficient to only learn an estimate $\hat{\eta}_z$ to construct a generative model. Similar to the previous section, we demonstrate this on the Oxford flowers dataset, again making use of a U-Net parameterization for $\hat{\eta}_z(t,x)$. Further experimental details can be found in Appendix C.1. In this setup the output image at time $t = 1$ is the same as the input image if we use the ODE (2.32); with the SDE, however, we can generate new images from the same input. This is illustrated in Figure 15, where we show how a sample image from the dataset ρ_1 is pushed forward through the SDE (2.33) with $\epsilon(t) = \epsilon = 10$. As can be seen the original image is resampled to a proximal flower not seen in the dataset.

我们考虑镜像插值 $x_t = x_1 + \gamma(t)z$。式（3.33）表明，其漂移 b 可以用去噪器 η_z 表示为 $b(t,x) = \dot{\gamma}(t)\eta_z(t)$；这意味着，只学习一个估计 $\hat{\eta}_z$ 就足以构建生成模型。与上一节类似，我们在 Oxford flowers 数据集上演示这一方法，并再次用 U-Net 对 $\hat{\eta}_z(t,x)$ 参数化。更多实验细节见附录 C.1。在这一设置下，如果采用 ODE（2.32），时刻 $t = 1$ 的输出图像与输入图像相同；而使用 SDE 时，则可以由同一个输入生成新图像。图 15 对此作了说明：图中展示数据集 ρ_1 中的一张样本图像如何经由 $\epsilon(t) = \epsilon = 10$ 的 SDE（2.33）向前推送。可以看到，原图像被重新采样为一朵与其相近、但数据集中未见的花。

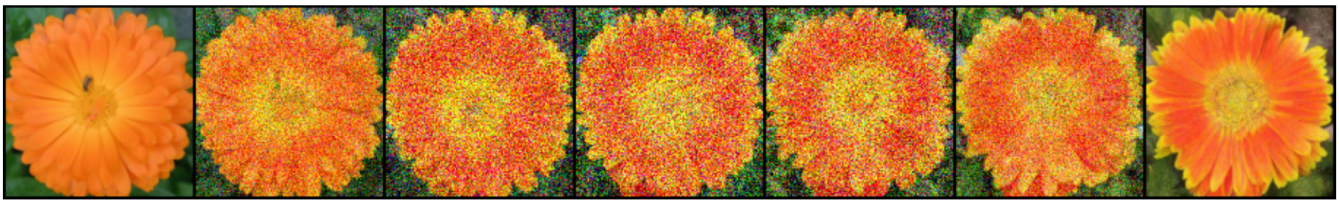

Figure 15: Mirror interpolant. Example trajectory from a learned denoiser model $\hat{b} = \dot{\gamma}(t)\hat{\eta}_z$ for the mirror interpolant $x_t = x_1 + \gamma(t)z$ on the 128×128 Oxford flowers dataset with $\gamma(t) = \sqrt{t(1-t)}$. The parametric form for $\hat{\eta}_z$ is the U-Net from *Ho et al. (2020)*, with hyperparameter details given in Appendix C. The choice of ϵ in the generative SDE given in (2.33) influences the extent to which the generated image differs from the original. Here, $\epsilon(t) = 10.0$.

图 15：镜像插值。在 128×128 Oxford flowers 上，$x_t = x_1 + \gamma(t)z$，$\gamma = \sqrt{t(1-t)}$，用学习的去噪器构造 $\hat{b} = \dot{\gamma}\hat{\eta}_z$。网络为 Ho 等（2020）的 U-Net，超参数见附录 C。式（2.33）中 ε 决定生成图偏离原图的程度，此处 $\epsilon(t) = 10$。

8 Conclusion

8 结论

S349 · Original / 原文 · PDF p.53 · S8.p1.1

The above exposition provides a full treatment of the stochastic interpolant method, as well as a careful consideration of its relation to existing literature. Our goal is to provide a general framework that can be used to devise generative models built upon dynamical transport of measure. To this end, we have detailed mathematical theory and efficient algorithms for constructing both deterministic and stochastic generative models that map between two densities exactly in finite time. Along the way, we have illustrated the various design parameters that can be used to shape this process, with connections, for example, to optimal transport and Schrödinger bridges. While we detail specific instantiations, such as the mirror and one-sided interpolants, we highlight that there is a much broader space of possible designs that may be relevant for future applications. Several candidate application domains include the solution of inverse problems such as image inpainting and super-resolution, spatiotemporal forecasting of dynamical systems, and scientific problems such as sampling of molecular configurations and machine learning-assisted Markov chain Monte-Carlo.

以上内容完整介绍了随机插值方法，并仔细讨论了它与现有文献的关系。我们的目标是提供一个通用框架，用于设计基于测度动态输运的生成模型。为此，我们详细给出了数学理论和高效算法，用以构造能够在有限时间内精确映射两个密度的确定性与随机生成模型。在这一过程中，我们说明了可用于塑造该过程的多种设计参数，并展示了它与最优传输、Schrödinger 桥等问题的联系。虽然我们详细介绍了镜像插值和单侧插值等具体形式，但我们强调，潜在的设计空间远比这些例子广阔，可能对未来应用具有价值。候选应用领域包括：图像修补、超分辨率等逆问题的求解，动力系统的时空预测，以及分子构型采样、机器学习辅助的 Markov 链 Monte Carlo 等科学问题。

Acknowledgements

致谢

S350 · Original / 原文 · PDF p.53 · Sx1.p1.1

We thank Joan Bruna, Jonathan Niles-Weed, Loucas Pillaud-Vivien, and Cédric Gerbelot for helpful discussions regarding stability estimates of dynamical transport. We are also grateful to Qiang Liu, Ricky Chen, and Yaron Lipman for feedback on previous and related work, and to Kyle Cranmer and Michael Lindsey for discussions on transport costs. We thank Mark Goldstein for insightful comments regarding practical considerations when training denoising-diffusion models. MSA is supported by the National Science Foundation under the award PHY-2141336. EVE is supported by the National Science Foundation under awards DMR-1420073, DMS-2012510, and DMS-2134216, by the Simons Collaboration on Wave Turbulence, Grant No. 617006, and by a Vannevar Bush Faculty Fellowship.

感谢 Joan Bruna、Jonathan Niles-Weed、Loucas Pillaud-Vivien 和 Cédric Gerbelot 就动态输运的稳定性估计所作的有益讨论。也感谢 Qiang Liu、Ricky Chen 和 Yaron Lipman 对先前工作及相关工作提出反馈，并感谢 Kyle Cranmer 和 Michael Lindsey 围绕输运成本展开讨论。感谢 Mark Goldstein 就去噪扩散模型训练的实践问题提出富有洞见的意见。MSA 获美国国家科学基金会项目 PHY-2141336 资助。EVE 获美国国家科学基金会项目 DMR-1420073、DMS-2012510 与 DMS-2134216，Simons 波湍流合作项目（项目号 617006），以及 Vannevar Bush Faculty Fellowship 的资助。

Appendix A Bridging two Gaussian mixture densities
附录 A 连接高斯混合密度

S351 · Original / 原文 · PDF p.54 · A1.p1.2

In this appendix, we consider the case where ρ_0 and ρ_1 are both Gaussian mixture densities. We denote by

本附录考虑 ρ_0 与 ρ_1 均为高斯混合密度的情形。我们用下述记号表示

$$\begin{aligned}\mathsf{N}(x|m,C) &= (2\pi)^{-d/2}[\det C]^{-1/2}\exp\left(-\tfrac{1}{2}(x-m)^{\mathsf{T}}C^{-1}(x-m)\right),\\ &= (2\pi)^{-d}\int_{\mathbb{R}^d} e^{ik\cdot(x-m)-\frac{1}{2}k^{\mathsf{T}}Ck}dk,\end{aligned} \tag{A.1}$$

S352 · Original / 原文 · PDF p.54 · A1.p1.3

the Gaussian probability density with mean vector $m\in\mathbb{R}^d$ and positive-definite symmetric covariance matrix $C=C^{\mathsf{T}}\in\mathbb{R}^{d\times d}$. We assume that

均值向量为 $m\in\mathbb{R}^d$、对称正定协方差矩阵为 $C=C^{\mathsf{T}}\in\mathbb{R}^{d\times d}$ 的高斯概率密度。假设

$$\rho_0(x)=\sum_{i=1}^{N_0}p_i^0\mathsf{N}(x|m_i^0,C_i^0),\qquad \rho_1(x)=\sum_{i=1}^{N_1}p_i^1\mathsf{N}(x|m_i^1,C_i^1) \tag{A.2}$$

S353 · Original / 原文 · PDF p.54 · A1.p1.4

where $N_0,N_1\in\mathbb{N}$, $p_i^0>0$ with $\sum_{i=1}^{N_0}p_i^0=1$, $m_i^0\in\mathbb{R}^d$, $C_i^0=(C_i^0)^{\mathsf{T}}\in\mathbb{R}^{d\times d}$ positive-definite, and similarly for p_i^1, m_i^1, and C_i^1. We have:

其中，$N_0,N_1\in\mathbb{N}$，$p_i^0>0$ 且 $\sum_{i=1}^{N_0}p_i^0=1$，$m_i^0\in\mathbb{R}^d$，$C_i^0=(C_i^0)^{\mathsf{T}}\in\mathbb{R}^{d\times d}$ 为正定矩阵；p_i^1、m_i^1 与 C_i^1 也满足类似条件。于是有：

Proposition 51.命题 51

S354 · Original / 原文 · PDF p.54 · Thmtheorem51.p1.1

Consider the process x_t defined in (2.1) using the probability densities in (A.2) and the interpolant in (4.1), i.e.

考虑式（2.1）定义的过程 x_t，其中使用式（A.2）的概率密度和式（4.1）的插值，即

$$x_t^{\mathsf{lin}}=\alpha(t)x_0+\beta(t)x_1+\gamma(t)z, \tag{4.1}$$

S355 · Original / 原文 · PDF p.54 · Thmtheorem51.p1.2

where $x_0,\sim\rho_0$, $x_1\sim\rho_1$, and $z\sim\mathsf{N}(0,Id)$ with $x_0\perp x_1\perp z$, and $\alpha,\beta,\gamma^2\in C^2([0,1])$ satisfy the conditions in (4.2). Denote

其中，$x_0,\sim\rho_0$、$x_1\sim\rho_1$、$z\sim\mathsf{N}(0,Id)$，且 $x_0\perp x_1\perp z$；$\alpha,\beta,\gamma^2\in C^2([0,1])$ 满足式（4.2）的条件。记

$$m_{ij}(t)=\alpha(t)m_i^0+\beta(t)m_j^1,\quad C_{ij}(t)=\alpha^2(t)C_i^0+\beta^2(t)C_j^1+\gamma^2(t)Id, \tag{A.3}$$

S356 · Original / 原文 · PDF p.54 · Thmtheorem51.p1.3

where $i=1,\dots,N_0$, $j=1,\dots,N_1$. Then the probability density ρ of x_t is the Gaussian mixture density

其中 $i=1,\dots,N_0$、$j=1,\dots,N_1$。那么，x_t 的概率密度 ρ 是下述高斯混合密度：

$$\rho(t,x)=\sum_{i=1}^{N_0}\sum_{j=1}^{N_1}p_i^0p_j^1\mathsf{N}(x|m_{ij}(t),C_{ij}(t)) \tag{A.4}$$

S357 · Original / 原文 · PDF p.54 · Thmtheorem51.p1.4

and the velocity b and the score s defined in (2.10) and (2.14) are

而式（2.10）与（2.14）定义的速度 b 和得分 s 分别为

$$b(t,x)=\frac{\sum_{i=1}^{N_0}\sum_{j=1}^{N_1}p_i^0p_j^1\left(\dot m_{ij}(t)+\frac{1}{2}\dot C_{ij}(t)C_{ij}^{-1}(t)(x-m^{ij}(t))\right)\mathsf{N}(x|m_{ij}(t),C_{ij}(t))}{\sum_{i=1}^{N_0}\sum_{j=1}^{N_1}p_i^0p_j^1\mathsf{N}(x|m_{ij}(t),C_{ij}(t))}, \tag{A.5}$$

S358 · Original / 原文 · PDF p.54 · Thmtheorem51.p1.5

and

以及

$$s(t,x)=-\frac{\sum_{i=1}^{N_0}\sum_{j=1}^{N_1}p_i^0p_j^1C_{ij}^{-1}(t)(x-m_{ij}(t))\mathsf{N}(x|m_{ij}(t),C_{ij}(t))}{\sum_{i=1}^{N_0}\sum_{j=1}^{N_1}p_i^0p_j^1\mathsf{N}(x|m_{ij}(t),C_{ij}(t))}. \tag{A.6}$$

S359 · Original / 原文 · PDF p.55 · A1.p2.1

This proposition implies that b and s grow at most linearly in x, and are approximately linear in regions where the modes of $\rho(t,x)$ remain well-separated. In particular, if ρ_0 and ρ_1 are both Gaussian densities, $\rho_0=\mathsf{N}(m_0,C_0)$ and $\rho_1=\mathsf{N}(m_1,C_1)$, we have

这一命题意味着，b 与 s 随 x 至多线性增长，并且在 $\rho(t,x)$ 的模态保持良好分离的区域中近似为线性。特别地，若 ρ_0 与 ρ_1 都是高斯密度，即 $\rho_0=\mathsf{N}(m_0,C_0)$、$\rho_1=\mathsf{N}(m_1,C_1)$，则有

$$b(t,x)=\dot m(t)+\frac{1}{2}\dot C(t)C^{-1}(t)(x-m(t)), \tag{A.7}$$

S360 · Original / 原文 · PDF p.55 · A1.p2.2

and

以及

$$s(t,x)=-C^{-1}(t)(x-m(t)), \tag{A.8}$$

S361 · Original / 原文 · PDF p.55 · A1.p2.3

where

其中

$$m(t)=\alpha(t)m_0+\beta(t)m_1,\quad C(t)=\alpha^2(t)C_0+\beta^2(t)C_1+\gamma^2(t)Id. \tag{A.9}$$

S362 · Original / 原文 · PDF p.55 · A1.p2.4

Note that the probability flow ODE (2.32) associated with the velocity (A.7) is the linear ODE

注意，与速度（A.7）关联的概率流 ODE（2.32）是如下线性 ODE：

$$\frac{d}{dt}X_t = \dot{m}(t) + \tfrac{1}{2}\dot{C}(t)C^{-1}(t)(X_t - m(t)). \tag{A.10}$$

S363 · Original / 原文 · PDF p.55 · A1.p2.5

This equation can only be solved analytically if $\dot{C}(t)$ and $C(t)$ commute (which is the case e.g. if $C_0 = Id$), but it is easy to see that it always guarantees that

只有在 $\dot{C}(t)$ 与 $C(t)$ 可交换时，这一方程才能解析求解，例如 $C_0 = Id$ 时即属此种情形。但容易看出，它始终保证

$$\mathbb{E}_0 X_t(x_0) = m(t), \qquad \mathbb{E}_0\left[(X_t(x_0) - m(t))(X_t(x_0) - m(t))^\mathsf{T}\right] = C(t), \tag{A.11}$$

S364 · Original / 原文 · PDF p.55 · A1.p2.6

where $X_t(x_0)$ denotes the solution to (A.10) for the initial condition $X_{t=0}(x_0) = x_0$ and $\mathbb{E}_0$ denotes expectation over $x_0 \sim \rho_0$. A similar statement is true if we solve (A.10) with final conditions at $t = 1$ drawn from ρ_1. Similarly, the forward SDE (2.33) associated with the velocity (A.7) and the score (A.8) is the linear SDE

其中，$X_t(x_0)$ 表示式（A.10）在初始条件 $X_{t=0}(x_0) = x_0$ 下的解，$\mathbb{E}_0$ 表示对 $x_0 \sim \rho_0$ 取期望。若在 $t = 1$ 从 ρ_1 抽取终值条件并求解式（A.10），类似结论也成立。同样，与速度（A.7）及得分（A.8）关联的前向 SDE（2.33）是如下线性 SDE：

$$dX_t^\mathsf{F} = \dot{m}(t)dt + \left(\tfrac{1}{2}\dot{C}(t) - \epsilon\right)C^{-1}(t)(X_t^\mathsf{F} - m(t))dt + \sqrt{2\epsilon}dW_t. \tag{A.12}$$

S365 · Original / 原文 · PDF p.55 · A1.p2.7

and its solutions are such that

并且它的解满足

$$\mathbb{E}_0\mathbb{E}_\mathsf{F}^{x_0} X_t^\mathsf{F} = m(t), \qquad \mathbb{E}_0\mathbb{E}_\mathsf{F}^{x_0}\left[(X_t^\mathsf{F} - m(t))(X_t^\mathsf{F} - m(t))^\mathsf{T}\right] = C(t), \tag{A.13}$$

S366 · Original / 原文 · PDF p.55 · A1.p2.8

where $\mathbb{E}_\mathsf{F}^{x_0}$ denotes expectation over the solution of (A.12) conditional on the event $X_{t=0}^\mathsf{F} = x_0$ and $\mathbb{E}_0$ denotes expectation over $x_0 \sim \rho_0$. A similar statement also holds for the backward SDE (2.34).

其中，$\mathbb{E}_\mathsf{F}^{x_0}$ 表示在事件 $X_{t=0}^\mathsf{F} = x_0$ 条件下，对式（A.12）的解取期望；$\mathbb{E}_0$ 表示对 $x_0 \sim \rho_0$ 取期望。后向 SDE（2.34）也满足类似结论。

Proof.

证明

S367 · Original / 原文 · PDF p.55 · A1.p3.1

The characteristic function of $\rho(t, x)$ is given by

$\rho(t, x)$ 的特征函数为

$$g(t, k) = \mathbb{E}e^{ik\cdot x_t} = \sum_{i=1}^{N_0}\sum_{j=1}^{N_1} p_i^0 p_j^1 e^{ik\cdot m_{ij}(t) - \frac{1}{2}k^\mathsf{T}C_{ij}(t)k}. \tag{A.14}$$

S368 · Original / 原文 · PDF p.55 · A1.p3.2

whose inverse Fourier transform is (A.4). This automatically implies (A.6) since we know from 2.14 that $s = \nabla \log \rho$. To derive (A.7) use the function m defined below in (B.12):

其逆 Fourier 变换就是式（A.4）。由于根据式（2.14）有 $s = \nabla \log \rho$，这自动蕴含式（A.6）。为推导式（A.7），使用下文式（B.12）定义的函数 m：

$$m(t,k) = \sum_{i=1}^{N_0} \sum_{j=1}^{N_1} p_i^0 p_j^1 (\dot{m}_{ij}(t) + \tfrac{1}{2} i \dot{C}_{ij}(t) k) e^{ik \cdot m_{ij}(t) - \frac{1}{2} k^\mathsf{T} C_{ij}^\gamma(t) k}. \tag{A.15}$$

S369 · Original / 原文 · PDF p.56 · A1.p3.3

From (B.13), we know that the inverse Fourier transform of this function is $b\rho$, so that we obtain

由式（B.13）可知，该函数的逆 Fourier 变换为 $b\rho$，因此得到

$$b(t,x)\rho(t,x) = \sum_{i=1}^{N_0} \sum_{j=1}^{N_1} p_i^0 p_j^1 \left(\dot{m}_{ij}(t) + \tfrac{1}{2} \dot{C}_{ij}(t) C_{ij}^{-1}(t)(x - m_{ij}(t)) \right) \mathsf{N}(x | m_{ij}(t), C_{ij}^\epsilon(t)). \tag{A.16}$$

S370 · Original / 原文 · PDF p.56 · A1.p3.4

This gives (A.5). ■

由此得到式（A.5）。■

Appendix B Proofs
附录 B 证明

S371 · Original / 原文 · PDF p.56 · A2.p1.1

In this appendix, we provide the details for proofs omitted from the main text. For ease of reading, a copy of the original theorem statement is provided with the proof.

本附录给出正文中略去的证明细节。为便于阅读，我们在证明之前附上原定理陈述的副本。

B.1 Proof of Theorems 6, 7, and 8, and Corollary 10.
B.1 定理 6、7、8 与推论 10 的证明

S372 · Original / 原文 · PDF p.56 · A2.SS1.p1.1

See 6

见定理 6。

Proof.
证明

S373 · Original / 原文 · PDF p.56 · A2.SS1.p2.3

Let $g(t,k)=\mathbb{E}e^{ik\cdot x_t}$, $k\in\mathbb{R}^d$, be the characteristic function of $\rho(t,x)$. From the definition of x_t in (2.1),

设 $g(t,k)=\mathbb{E}e^{ik\cdot x_t}$（$k\in\mathbb{R}^d$）是 $\rho(t,x)$ 的特征函数。根据式（2.1）中 x_t 的定义，

$$g(t,k)=\mathbb{E}e^{ik\cdot(I(t,x_0,x_1)+\gamma(t)z)}. \tag{B.1}$$

S374 · Original / 原文 · PDF p.56 · A2.SS1.p2.4

Using the independence between (x_0,x_1) and z , we have

利用 (x_0,x_1) 与 z 的独立性，有

$$g(t,k)=\mathbb{E}\left(e^{ik\cdot I(t,x_0,x_1)}\right)\mathbb{E}\left(e^{i\gamma(t)k\cdot z}\right)\equiv g_0(t,k)e^{-\frac{1}{2}\gamma^2(t)|k|^2} \tag{B.2}$$

S375 · Original / 原文 · PDF p.56 · A2.SS1.p2.5

where we defined

其中定义了

$$g_0(t,k)=\mathbb{E}\left(e^{ik\cdot I(t,x_0,x_1)}\right) \tag{B.3}$$

S376 · Original / 原文 · PDF p.56 · A2.SS1.p2.6

The function $g_0(t,k)$ is the characteristic function of $I(t,x_0,x_1)$ with $(x_0,x_1)\sim\nu$. From (B.2), we have

函数 $g_0(t,k)$ 是 $I(t,x_0,x_1)$ 的特征函数，其中 $(x_0,x_1)\sim\nu$。由式（B.2）可得

$$|g(t,k)|=|g_0(t,k)|e^{-\frac{1}{2}\gamma^2(t)|k|^2}\leq e^{-\frac{1}{2}\gamma^2(t)|k|^2} \tag{B.4}$$

S377 · Original / 原文 · PDF p.57 · A2.SS1.p2.7

Since $\gamma(t) > 0$ for all $t \in (0,1)$ by assumption, this shows that

根据假设，对所有 $t \in (0,1)$ 都有 $\gamma(t) > 0$，因此可知

$$\forall p \in \mathbb{N} \text{ and } t \in (0,1) \quad : \quad \int_{\mathbb{R}^d} |k|^p |g(t,k)| dk < \infty, \tag{B.5}$$

S378 · Original / 原文 · PDF p.57 · A2.SS1.p2.8

implying that $\rho(t,\cdot)$ is in $C^p(\mathbb{R}^d)$ for any $p \in \mathbb{N}$ and all $t \in (0,1)$. From (B.2), we also have

这意味着，对任意 $p \in \mathbb{N}$ 和所有 $t \in (0,1)$，$\rho(t,\cdot)$ 都属于 $C^p(\mathbb{R}^d)$。由式（B.2）还可得

$$\begin{aligned} |\partial_t g(t,k)|^2 &= \left| \mathbb{E}[(ik \cdot \partial_t I_t(x_0,x_1) - \gamma(t)\dot{\gamma}(t)|k|^2) e^{ik \cdot I_t(x_0,x_1)}] \right|^2 e^{-\gamma^2(t)|k|^2} \\ &\leq 2\left(|k|^2 \mathbb{E}\left[|\partial_t I_t(x_0,x_1)|^2 \right] + |\gamma(t)\dot{\gamma}(t)|^2 |k|^4 \right) e^{-\gamma^2(t)|k|^2} \\ &\leq 2\left(|k|^2 M_1 + 4|\gamma(t)\dot{\gamma}(t)|^2 |k|^4 \right) e^{-\gamma^2(t)|k|^2} \end{aligned} \tag{B.6}$$

S379 · Original / 原文 · PDF p.57 · A2.SS1.p2.9

and

以及

$$\begin{aligned} |\partial_t^2 g(t,k)|^2 \leq\ & 4\left(|k|^2 \mathbb{E}\left[|\partial_t^2 I_t(x_0,x_1)|^2 \right] + (|\dot{\gamma}(t)|^2 + \gamma(t)\ddot{\gamma}(t))^2 |k|^4 \right) e^{-\gamma^2(t)|k|^2} \\ & + 8\left(|k|^2 \mathbb{E}\left[|\partial_t I_t(x_0,x_1)|^4 \right] + (\gamma(t)\dot{\gamma}(t))^4 |k|^8 \right) e^{-\gamma^2(t)|k|^2} \\ \leq\ & 4\left(|k|^2 M_2 + (|\dot{\gamma}(t)|^2 + \gamma(t)\ddot{\gamma}(t))^2 |k|^4 \right) e^{-\gamma^2(t)|k|^2} \\ & + 8\left(|k|^2 M_1 + (\gamma(t)\dot{\gamma}(t))^4 |k|^8 \right) e^{-\gamma^2(t)|k|^2} \end{aligned} \tag{B.7}$$

S380 · Original / 原文 · PDF p.57 · A2.SS1.p2.10

where in both cases we used (2.8) in Assumption 5 to get the last inequalities. These imply that

在这两种情形中，我们都利用假设 5 中的式（2.8）得到最后的不等式。由此可知

$$\forall p \in \mathbb{N} \text{ and } t \in (0,1) \quad : \quad \int_{\mathbb{R}^d} |k|^p |\partial_t g(t,k)| dk < \infty; \quad \int_{\mathbb{R}^d} |k|^p |\partial_t^2 g(t,k)| dk < \infty \tag{B.8}$$

S381 · Original / 原文 · PDF p.57 · A2.SS1.p2.11

indicating that $\partial_t \rho(t,\cdot)$ and $\partial_t^2 \rho(t,\cdot)$ are in $C^p(\mathbb{R}^d)$ for any $p \in \mathbb{N}$, i.e. $\rho \in C^1((0,1); C^p(\mathbb{R}^d))$ as claimed. To show that ρ is also positive, denote by $\mu_0(t,dx)$ the unique (by the Fourier inversion theorem) probability measure associated with $g_0(t,k)$, i.e. the measure such that

这表明，对任意 $p \in \mathbb{N}$，$\partial_t \rho(t,\cdot)$ 与 $\partial_t^2 \rho(t,\cdot)$ 都属于 $C^p(\mathbb{R}^d)$，即如所声称的那样，$\rho \in C^1((0,1); C^p(\mathbb{R}^d))$。为进一步证明 ρ 为正，令 $\mu_0(t,dx)$ 表示与 $g_0(t,k)$ 对应的概率测度；根据 Fourier 反演定理，该测度唯一，即满足下式的测度：

$$g_0(t,k) = \int_{\mathbb{R}^d} e^{ik \cdot x} \mu_0(t,dx). \tag{B.9}$$

S382 · Original / 原文 · PDF p.57 · A2.SS1.p2.12

From (B.2) and the convolution theorem it follows that we can express ρ as

由式（B.2）和卷积定理可知，ρ 可写为

$$\rho(t,x) = \int_{\mathbb{R}^d} \frac{e^{-|x-y|^2/(2\gamma^2(t))}}{(2\pi\gamma^2(t))^{d/2}} \mu_0(t,dy), \tag{B.10}$$

S383 · Original / 原文 · PDF p.57 · A2.SS1.p2.13

This shows that $\rho > 0$ for all $(t,x) \in (0,1)\times\mathbb{R}^d$. Since $x_{t=0} = x_0$ and $x_{t=1} = x_1$ by definition of the interpolant, we also have $\rho(0) = \rho_0$ and $\rho(1) = \rho_1$, which shows that ρ is also positive and in $C^p(\mathbb{R}^d)$ at $t = 0, 1$ by Assumption 5. Note that since $\rho \in C^1((0,1); C^p(\mathbb{R}^d))$ and is positive, we also immediately deduce that $s = \nabla \log \rho = \nabla\rho/\rho \in C^1((0,1); (C^p(\mathbb{R}^d))^d)$.

这表明，对所有 $(t,x) \in (0,1)\times\mathbb{R}^d$，都有 $\rho > 0$。根据插值的定义，$x_{t=0} = x_0$、$x_{t=1} = x_1$，因此也有 $\rho(0) = \rho_0$、$\rho(1) = \rho_1$；再由假设 5 可知，ρ 在 $t = 0, 1$ 时同样为正，并属于 $C^p(\mathbb{R}^d)$。注意，由于 $\rho \in C^1((0,1); C^p(\mathbb{R}^d))$ 且为正，我们还立即得到 $s = \nabla \log \rho = \nabla\rho/\rho \in C^1((0,1); (C^p(\mathbb{R}^d))^d)$。

S384 · Original / 原文 · PDF p.57 · A2.SS1.p3.2

To show that ρ satisfies the TE (2.9), we take the time derivative of (B.1) to deduce that

为证明 ρ 满足输运方程 TE（2.9），我们对式（B.1）求时间导数，得到

$$\partial_t g(t,k) = ik \cdot m(t,k) \tag{B.11}$$

S385 · Original / 原文 · PDF p.57 · A2.SS1.p3.3

where $m : [0,1]\times\mathbb{R}^d \to \mathbb{C}^d$ is the vector-valued function defined as

其中，$m : [0,1]\times\mathbb{R}^d \to \mathbb{C}^d$ 是如下定义的向量值函数：

$$m(t,k) = \mathbb{E}\left((\partial_t I(t,x_0,x_1) + \dot\gamma(t) z) e^{ik\cdot x_t}\right). \tag{B.12}$$

S386 · Original / 原文 · PDF p.58 · A2.SS1.p3.4

By definition of the conditional expectation, $m(t,k)$ can be expressed as

根据条件期望的定义，$m(t,k)$ 可以写成

$$\begin{aligned} m(t,k) &= \int_{\mathbb{R}^d} \mathbb{E}\left((\partial_t I(t,x_0,x_1) + \dot\gamma(t) z) e^{ik\cdot x_t} | x_t = x\right) \rho(t,x) dx \\ &= \int_{\mathbb{R}^d} e^{ik\cdot x} \mathbb{E}\left((\partial_t I(t,x_0,x_1) + \dot\gamma(t) z) | x_t = x\right) \rho(t,x) dx \\ &= \int_{\mathbb{R}^d} e^{ik\cdot x} b(t,x) \rho(t,x) dx \end{aligned} \tag{B.13}$$

S387 · Original / 原文 · PDF p.58 · A2.SS1.p3.5

where the last equality follows from the definition of b in (2.10). Inserting (B.13) in (B.11), we deduce that this equation can be written in real space as the TE (2.9).

最后一个等号来自式（2.10）中 b 的定义。将式（B.13）代入式（B.11），可知该方程在实空间中可写为输运方程 TE（2.9）。

S388 · Original / 原文 · PDF p.58 · A2.SS1.p4.3

Let us now investigate the regularity of b. To that end, we go back to m and use the independence between x_0, x_1, and z, as well as Gaussian integration by parts to deduce that

现在考察 b 的正则性。为此，我们回到 m，利用 x_0、x_1 与 z 之间的独立性，以及高斯分部积分，得到

$$m(t,k) = \mathbb{E}\left((\partial_t I(t,x_0,x_1) - i\gamma(t)\dot\gamma(t) k) e^{ik\cdot I(t,x_1,x_0)}\right) e^{-\frac{1}{2}\gamma^2(t)|k|^2}, \tag{B.14}$$

S389 · Original / 原文 · PDF p.58 · A2.SS1.p4.4

As a result

因此

$$
\begin{aligned}
|m(t,k)|^2 &= \left|\mathbb{E}\left((\partial_t I(t,x_0,x_1) - i\gamma(t)\dot\gamma(t)k)e^{ik\cdot I(t,x_1,x_0)}\right)\right|^2 e^{-\gamma^2(t)|k|^2} \\
&\le 2\left(\mathbb{E}\left[|\partial_t I(t,x_0,x_1)|^2\right] + |\gamma(t)\dot\gamma(t)|^2|k|^2\right) e^{-\gamma^2(t)|k|^2} \\
&\le 2M_1 e^{-\gamma^2(t)|k|^2},
\end{aligned}
\tag{B.15}
$$

S390 · Original / 原文 · PDF p.58 · A2.SS1.p4.5

and

以及

$$
\begin{aligned}
|\partial_t m(t,k)|^2 \le\ & 4\left(\mathbb{E}\left[|\partial_t^2 I(t,x_0,x_1)|^2 + (\gamma(t)\ddot\gamma(t) + \dot\gamma^2(t))^2\right) e^{-\gamma^2(t)|k|^2}\right. \\
& \left. +8|k|^2\left(\mathbb{E}\left|\partial_t I(t,x_0,x_1)\right|^4\right] + (\gamma(t)\dot\gamma(t))^4|k|^4\right) e^{-\gamma^2(t)|k|^2} \\
\le\ & 4\left(M_1 + (\gamma(t)\ddot\gamma(t) + \dot\gamma^2(t))^2\right) e^{-\gamma^2(t)|k|^2} \\
& +8|k|^2\left(M_2 + (\gamma(t)\dot\gamma(t))^4|k|^4\right) e^{-\gamma^2(t)|k|^2},
\end{aligned}
\tag{B.16}
$$

S391 · Original / 原文 · PDF p.58 · A2.SS1.p4.6

where in both cases the last inequalities follow from (2.8). Therefore

这两种情形的最后一个不等式均由式（2.8）得到。因此，

$$
\forall p \in \mathbb{N} \text{ and } t \in (0,1) \quad : \quad \int_{\mathbb{R}^d} |k|^p |m(t,k)| dk < \infty, \quad \int_{\mathbb{R}^d} |k|^p |\partial_t m(t,k)| dk < \infty,
\tag{B.17}
$$

S392 · Original / 原文 · PDF p.58 · A2.SS1.p4.7

which implies that the inverse Fourier transform of m is a function $j : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ that satisfies $j(t,\cdot) \in (C^p(\mathbb{R}^d))^d$ for any $p \in \mathbb{N}$ and any $t \in (0,1)$ and can be expressed as

这意味着，m 的逆 Fourier 变换是一个函数 $j : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$；对于任意 $p \in \mathbb{N}$ 和任意 $t \in (0,1)$，它满足 $j(t,\cdot) \in (C^p(\mathbb{R}^d))^d$，并可表示为

$$
j(t,x) = (2\pi)^{-d} \int_{\mathbb{R}^d} e^{-ik\cdot x} m(t,k) dk = \mathbb{E}\left(\partial_t I(t,x_0,x_1) + \dot\gamma(t) z | x_t = x\right) \rho(t,x) \equiv b(t,x)\rho_t(x)
\tag{B.18}
$$

S393 · Original / 原文 · PDF p.58 · A2.SS1.p4.8

where the last equality follows from the definition of b in (2.10). We deduce that $b \in C^0([0,1]; (C^p(\mathbb{R}^d))^d)$ for any $p \in \mathbb{N}$ since $j \in C^0([0,1]; (C^p(\mathbb{R}^d))^d)$ and $\rho \in C^1([0,1]; C^p(\mathbb{R}^d))$, and $\rho > 0$.

最后一个等号来自式（2.10）中 b 的定义。由于 $j \in C^0([0,1]; (C^p(\mathbb{R}^d))^d)$、$\rho \in C^1([0,1]; C^p(\mathbb{R}^d))$，且 $\rho > 0$，所以可推出：对任意 $p \in \mathbb{N}$，$b \in C^0([0,1]; (C^p(\mathbb{R}^d))^d)$。

S394 · Original / 原文 · PDF p.58 · A2.SS1.p5.3

Finally, let us establish (2.11). By (2.8) we have

最后证明式（2.11）。由式（2.8）可得

$$\begin{aligned}\int_{\mathbb{R}^d}|b(t,x)|^2\rho(t,x)dx&=\int_{\mathbb{R}^d}|\mathbb{E}\left(\partial_t I(t,x_0,x_1)+\dot\gamma(t)z|x_t=x\right)|^2\rho(t,x)dx\\&\le 2\int_{\mathbb{R}^d}\mathbb{E}\left(|\partial_t I(t,x_0,x_1)|^2+|\dot\gamma(t)|^2|z|^2|x_t=x\right)\rho(t,x)dx\\&\le 2\mathbb{E}\left[|\partial_t I(t,x_0,x_1)|^2+|\dot\gamma(t)|^2|z|^2\right]\\&<2M_1^{1/2}+2d|\dot\gamma(t)|^2,\end{aligned}\tag{B.19}$$

S395 · Original / 原文 · PDF p.59 · A2.SS1.p5.4

so that this integral is bounded for all $t\in(0,1)$. To analyze its behavior at the end points, notice that the decomposition (2.27) implies that

因此，这一积分对所有 $t\in(0,1)$ 都有界。为分析其在端点的行为，注意分解式（2.27）意味着

$$\begin{aligned}b_0(x)&\equiv\lim_{t\to 0}b(t,x)=\mathbb{E}_1[\partial_t I(0,x,x_1)]-\lim_{t\to 0}\dot\gamma(t)\gamma(t)s_0(x),\\b_1(x)&\equiv\lim_{t\to 1}b(t,x)=\mathbb{E}_0[\partial_t I(0,x_0,x)]-\lim_{t\to 1}\dot\gamma(t)\gamma(t)s_1(x),\end{aligned}\tag{B.20}$$

S396 · Original / 原文 · PDF p.59 · A2.SS1.p5.5

where $s_0=\nabla\log\rho_0$, $s_1=\nabla\log\rho_1$, $\mathbb{E}_0$ and $\mathbb{E}_1$ denote expectations over $x_0\sim\rho_0$ and $x_1\sim\rho_1$, respectively, and we used the property that $x_{t=0}=x_0$ and $x_{t=1}=x_1$. Since $\lim_{t\to 0,1}\dot\gamma(t)\gamma(t)$ exists by our assumption that $\gamma^2\in C^1([0,1])$, b_0 and b_1 are well defined, and

其中，$s_0=\nabla\log\rho_0$，$s_1=\nabla\log\rho_1$；$\mathbb{E}_0$ 和 $\mathbb{E}_1$ 分别表示对 $x_0\sim\rho_0$ 和 $x_1\sim\rho_1$ 取期望。这里还使用了 $x_{t=0}=x_0$、$x_{t=1}=x_1$ 这一性质。由假设 $\gamma^2\in C^1([0,1])$ 可知，$\lim_{t\to 0,1}\dot\gamma(t)\gamma(t)$ 存在，因此 b_0 和 b_1 都有良好定义，并且

$$\int_{\mathbb{R}^d}|b_0(x)|^2\rho_0(x)dx<\infty,\qquad\int_{\mathbb{R}^d}|b_1(x)|^2\rho_1(x)dx<\infty.\tag{B.21}$$

S397 · Original / 原文 · PDF p.59 · A2.SS1.p5.6

by Assumption 5. As a result, the integral in (B.19) is continuous at $t=0$ and $t=1$, so it must be integrable on $[0,1]$, and (2.11) holds. ■

这是由假设 5 得到的。因此，式 (B.19) 中的积分在 $t=0$ 和 $t=1$ 处连续，故它在 $[0,1]$ 上必可积，于是式 (2.11) 成立。■

S398 · Original / 原文 · PDF p.59 · A2.SS1.p6.1

See 7

见 7。

Proof.

证明

S399 · Original / 原文 · PDF p.59 · A2.SS1.p7.3

By definition of ρ, the objective $\mathcal{L}_b$ defined in (2.13) can also be written as

根据 ρ 的定义，式 (2.13) 定义的目标泛函 $\mathcal{L}_b$ 也可以写成

$$\begin{aligned}\mathcal{L}_b[\hat b]&=\int_0^1\int_{\mathbb{R}^d}\left(\tfrac{1}{2}|\hat b(t,x)|^2-\mathbb{E}\left((\partial_t I(t,x_0,x_1)+\dot\gamma(t)z|x_t=x)\cdot\hat b(t,x)\right)\right)\rho(t,x)dxdt\\&=\int_0^1\int_{\mathbb{R}^d}\left(\tfrac{1}{2}|\hat b(t,x)|^2-b(t,x)\cdot\hat b(t,x)\right)\rho(t,x)dxdt\end{aligned}\tag{B.22}$$

S400 · Original / 原文 · PDF p.59 · A2.SS1.p7.4

where we used the definition of b in (2.10). This quadratic objective is bounded from below since

其中使用了式 (2.10) 中 b 的定义。这个二次目标泛函有下界，因为

$$\begin{aligned}\mathcal{L}_b[\hat{b}] &= \tfrac{1}{2}\int_0^1\int_{\mathbb{R}^d}\left|\hat{b}(t,x)-b(t,x)\right|^2\rho(t,x)dxdt - \tfrac{1}{2}\int_0^1\int_{\mathbb{R}^d}|b(t,x)|^2\rho(t,x)dxdt \\ &\geq -\tfrac{1}{2}\int_0^1\int_{\mathbb{R}^d}|b(t,x)|^2\rho(t,x)dxdt > -\infty\end{aligned} \tag{B.23}$$

S401 · Original / 原文 · PDF p.60 · A2.SS1.p7.5

where the last inequality follows from (2.11). Since ρ_t is positive the minimizer of (B.22) is unique and given by $\hat{b} = b$.

其中最后一个不等式由式 (2.11) 得到。由于 ρ_t 为正，式 (B.22) 的极小化解唯一，且为 $\hat{b} = b$。

S402 · Original / 原文 · PDF p.60 · A2.SS1.p8.1

■

■

S403 · Original / 原文 · PDF p.60 · A2.SS1.p9.1

See 8

见 8。

Proof.

证明

S404 · Original / 原文 · PDF p.60 · A2.SS1.p10.1

Since $\rho \in C^1((0,1);C^p(\mathbb{R}^d))$ and is positive by Theorem 6, we already know that $s = \nabla\log\rho = \nabla\rho/\rho \in C^1((0,1);(C^p(\mathbb{R}^d))^d)$. To establish (2.14), note that, for $t \in (0,1)$ where $\gamma(t) > 0$, we have

由定理 6 可知，$\rho \in C^1((0,1);C^p(\mathbb{R}^d))$ 且为正，因此我们已经知道 $s = \nabla\log\rho = \nabla\rho/\rho \in C^1((0,1);(C^p(\mathbb{R}^d))^d)$。为了证明式 (2.14)，注意到对于满足 $\gamma(t) > 0$ 的 $t \in (0,1)$，有

$$\mathbb{E}\left(ze^{i\gamma(t)k\cdot z}\right) = -\gamma^{-1}(t)(i\partial_k)\mathbb{E}e^{i\gamma(t)k\cdot z} = -\gamma^{-1}(t)(i\partial_k)e^{-\frac{1}{2}\gamma^2(t)|k|^2} = i\gamma(t)ke^{-\frac{1}{2}\gamma^2(t)|k|^2}. \tag{B.24}$$

S405 · Original / 原文 · PDF p.60 · A2.SS1.p10.2

As a result, using the independence between x_0, x_1, and z, we have

因此，利用 x_0、x_1 与 z 的独立性，可得

$$\mathbb{E}\left(ze^{ik\cdot x_t}\right) = i\gamma(t)kg(t,k) \tag{B.25}$$

S406 · Original / 原文 · PDF p.60 · A2.SS1.p10.3

where g is the characteristic function of x_t defined in (B.1). Using the properties of the conditional expectation, the left-hand side of this equation can be written as

其中，g 是式 (B.1) 定义的 x_t 的特征函数。利用条件期望的性质，这个等式的左侧可以写成

$$\mathbb{E}\left(ze^{ik\cdot x_t}\right) = \int_{\mathbb{R}^d}\mathbb{E}\left(ze^{ik\cdot x_t}|x_t = x\right)\rho(t,x)dx = \int_{\mathbb{R}^d}\mathbb{E}\left(z|x_t = x\right)e^{ikx_t}\rho(t,x)dx \tag{B.26}$$

S407 · Original / 原文 · PDF p.60 · A2.SS1.p10.4

Since the left hand side of (B.25) is the Fourier transform of $-\gamma(t)\nabla\rho(t,x)$, we deduce that

由于式 (B.25) 左侧是 $-\gamma(t)\nabla\rho(t,x)$ 的傅里叶变换，我们推得

$$\mathbb{E}\left(z|x_t=x\right)\rho(t,x)=-\gamma(t)\nabla\rho(t,x)=-\gamma(t)s(t,x)\rho(t,x). \tag{B.27}$$

S408 · Original / 原文 · PDF p.60 · A2.SS1.p10.5

Since $\rho(t,x)>0$, this implies (2.14) for $t\in(0,1)$ where $\gamma(t)>0$.

由于 $\rho(t,x)>0$，这就说明，对于满足 $\gamma(t)>0$ 的 $t\in(0,1)$，式 (2.14) 成立。

S409 · Original / 原文 · PDF p.60 · A2.SS1.p11.2

To establish (2.15), notice that

为了证明式 (2.15)，注意到

$$\begin{aligned}\int_{\mathbb{R}^d}|s(t,x)|^2\rho(t,x)dx&=\int_{\mathbb{R}^d}|\mathbb{E}\left(\left(\gamma^{-1}(t)z|x_t=x\right)\right|^2\rho(t,x)dx\\&\leq\int_{\mathbb{R}^d}\gamma^{-2}(t)\mathbb{E}\left(|z|^2|x_t=x\right)\rho(t,x)dx\\&\leq d\gamma^{-2}(t).\end{aligned} \tag{B.28}$$

S410 · Original / 原文 · PDF p.61 · A2.SS1.p11.3

This means that this integral is bounded for all $t\in(0,1)$. Since the integral is also continuous at $t=0$ and $t=1$, with values given by (2.7), it must be integrable on $[0,1]$ and (2.15) holds.

这意味着该积分对于所有 $t\in(0,1)$ 都有界。由于该积分在 $t=0$ 和 $t=1$ 处也连续，其值由式 (2.7) 给出，因此它在 $[0,1]$ 上必可积，于是式 (2.15) 成立。

S411 · Original / 原文 · PDF p.61 · A2.SS1.p12.3

The objective $\mathcal{L}_s$ defined in (2.16) can also be written as

式 (2.16) 定义的目标泛函 $\mathcal{L}_s$ 也可以写成

$$\begin{aligned}\mathcal{L}_s[\hat{s}]&=\int_0^1\int_{\mathbb{R}^d}\left(\tfrac{1}{2}|\hat{s}(t,x)|^2+\gamma^{-1}(t)\mathbb{E}\left(z|x_t=x\right)\cdot\hat{s}(t,x)\right)\rho(t,x)dxdt\\&=\int_0^1\int_{\mathbb{R}^d}\left(\tfrac{1}{2}|\hat{s}(t,x)|^2-s(t,x)\cdot\hat{s}(t,x)\right)\rho(t,x)dxdt,\end{aligned} \tag{B.29}$$

S412 · Original / 原文 · PDF p.61 · A2.SS1.p12.4

where we used the definition of s in (2.14). This quadratic objective is bounded from below since

其中使用了式 (2.14) 中 s 的定义。这个二次目标泛函有下界，因为

$$\begin{aligned}\mathcal{L}_s[\hat{s}]&=\tfrac{1}{2}\int_0^1\int_{\mathbb{R}^d}\left|\hat{s}(t,x)-s(t,x)\right|^2\rho(t,x)dxdt-\tfrac{1}{2}\int_0^1\int_{\mathbb{R}^d}\left|s(t,x)\right|^2\rho(t,x)dxdt\\&\geq-\tfrac{1}{2}\int_0^1\int_{\mathbb{R}^d}\left|s(t,x)\right|^2\rho(t,x)dxdt>-\infty\end{aligned} \tag{B.30}$$

S413 · Original / 原文 · PDF p.61 · A2.SS1.p12.5

where the last inequality follows from (2.15). Since ρ is positive the minimizer of (B.29) is unique and given by $\hat{s}=s$. ■

其中最后一个不等式由式 (2.15) 得到。由于 ρ 为正，式 (B.29) 的极小化解唯一，且为 $\hat{s}=s$。■

S414 · Original / 原文 · PDF p.61 · A2.SS1.p13.1

See 10

见 10。

Proof.

证明

S415 · Original / 原文 · PDF p.61 · A2.SS1.p14.1

The forward FPE (2.20) and the backward FPE (2.22) are direct consequences of the TE (2.9) and (2.14), since the equality

正向 Fokker–Planck 方程 (2.20) 与反向 Fokker–Planck 方程 (2.22) 都直接由输运方程 (2.9) 和式 (2.14) 得到，因为等式

$$\epsilon(t)\Delta\rho = \epsilon(t)\nabla\cdot(\rho\nabla\log\rho) = \epsilon(t)\nabla\cdot(s\rho) \tag{B.31}$$

S416 · Original / 原文 · PDF p.61 · A2.SS1.p14.2

can be used to convert between these equations.

可以用来在这些方程之间进行转换。

S417 · Original / 原文 · PDF p.61 · A2.SS1.p15.1

■

■

B.2 Proof of Lemma 19

B.2 引理 19 的证明

S418 · Original / 原文 · PDF p.61 · A2.SS2.p1.1

See 19

见 19。

Proof.

证明

S419 · Original / 原文 · PDF p.62 · A2.SS2.p2.5

The SDE (2.33) and the ODE (2.32) are the evolution equations for the processes whose densities solve (2.20) and (2.9), respectively. The equation that requires some explanation is the backwards SDE (2.34), which can be solved backwards in time from $t = 1$ to $t = 0$. As discussed in the main text, by definition, its solution is $X_t^{\mathsf{B}} = Z_{1-t}^{\mathsf{F}}$ where Z_t^{F} solves the forward SDE

随机微分方程 (2.33) 和常微分方程 (2.32) 分别是其概率密度满足式 (2.20) 和式 (2.9) 的过程所遵循的演化方程。需要稍作解释的是反向随机微分方程 (2.34)，它可以从 $t = 1$ 到 $t = 0$ 沿时间反向求解。如正文所述，按照定义，其解为 $X_t^{\mathsf{B}} = Z_{1-t}^{\mathsf{F}}$，其中 Z_t^{F} 满足正向随机微分方程

$$dZ_t^{\mathsf{F}} = -b_{\mathsf{B}}(1-t, Z_t^{\mathsf{F}})dt + \sqrt{2\epsilon}dW_t. \tag{B.32}$$

S420 · Original / 原文 · PDF p.62 · A2.SS2.p2.6

To see how to write the backward Itô formula (2.36) note that given any $f \in C^1([0,1]; C_0^2(\mathbb{R}^d))$, we have

为了说明如何写出反向 Itô 公式 (2.36)，注意到，对任意 $f \in C^1([0,1]; C_0^2(\mathbb{R}^d))$，有

$$\begin{aligned} df(1-t, Z_t^{\mathsf{F}}) &= -\partial_t f(1-t, Z_t^{\mathsf{F}})dt + \nabla f(1-t, Z_t^{\mathsf{F}})\cdot dZ_t^{\mathsf{F}} + \epsilon\Delta f(1-t, Z_t^{\mathsf{F}})dt \\ &= -\partial_t f(1-t, Z_t^{\mathsf{F}})dt + \left(-b_{\mathsf{B}}(1-t, Z_t^{\mathsf{F}})\cdot\nabla f(1-t, Z_t^{\mathsf{F}}) + \epsilon\Delta f(1-t, Z_t^{\mathsf{F}})\right)dt \\ &\quad + \sqrt{2\epsilon}\nabla f(1-t, Z_t^{\mathsf{F}})\cdot dW_t \end{aligned} \tag{B.33}$$

S421 · Original / 原文 · PDF p.62 · A2.SS2.p2.7

In integral form, this equation can be written as

以积分形式，这个方程可以写成

$$\begin{aligned} f(1, Z_1^{\mathsf{F}}) = & f(1-t, Z_{1-t}^{\mathsf{F}}) \\ & - \int_{1-t}^{1} \left(\partial_t f(1-s, Z_s^{\mathsf{F}}) + b_{\mathsf{B}}(1-s, Z_s^{\mathsf{F}}) \cdot \nabla f(1-s, Z_s^{\mathsf{F}}) - \epsilon \Delta f(1-s, Z_s^{\mathsf{F}}) \right) ds \\ & - \sqrt{2\epsilon} \int_{1-t}^{1} \nabla f(1-s, Z_s^{\mathsf{F}}) \cdot dW_s. \end{aligned} \tag{B.34}$$

S422 · Original / 原文 · PDF p.62 · A2.SS2.p2.8

Using $X_t^{\mathsf{B}} = Z_{1-t}^{\mathsf{F}}$ and $W_t^{\mathsf{B}} = -W_{1-t}$ and changing integration variable from s to $1-s$, this is

利用 $X_t^{\mathsf{B}} = Z_{1-t}^{\mathsf{F}}$ 和 $W_t^{\mathsf{B}} = -W_{1-t}$，并把积分变量由 s 换成 $1-s$，上式成为

$$\begin{aligned} f(1, X_0^{\mathsf{B}}) = & f(1-t, X_t^{\mathsf{B}}) - \int_{t}^{1} \left(\partial_t f(s, X_s^{\mathsf{B}}) + b_{\mathsf{B}}(s, X_s^{\mathsf{B}}) \cdot \nabla f(s, X_s^{\mathsf{B}}) - \epsilon \Delta f(s, X_s^{\mathsf{B}}) \right) ds \\ & - \sqrt{2\epsilon} \int_{t}^{1} \nabla f(s, X_s^{\mathsf{B}}) \cdot dW_s^{\mathsf{B}}, \end{aligned} \tag{B.35}$$

S423 · Original / 原文 · PDF p.63 · A2.SS2.p2.9

In differential form, this is equivalent to saying that

以微分形式，这等价于说

$$\begin{aligned} df(t, X_t^{\mathsf{B}}) = & \partial_t f(t, X_t^{\mathsf{B}}) dt + \nabla f(t, X_t^{\mathsf{B}}) \cdot dX_t^{\mathsf{B}} - \epsilon \Delta f(X_t^{\mathsf{B}}) dt \\ = & \left(\partial_t f(t, X_t^{\mathsf{B}}) + b_{\mathsf{B}}(t, X_t^{\mathsf{B}}) \cdot \nabla f(t, X_t^{\mathsf{B}}) - \epsilon \Delta f(t, X_t^{\mathsf{B}}) \right) dt + \sqrt{2\epsilon} \nabla f(t, X_t^{\mathsf{B}}) \cdot dW_t \end{aligned} \tag{B.36}$$

S424 · Original / 原文 · PDF p.63 · A2.SS2.p2.10

which is the backward Itô formula (2.36). Similarly, by the Itô isometries we have: for any $g \in C^0([0,1]; (C_0^0(\mathbb{R}_d))^d)$ and $t \in [0,1]$

这就是反向 Itô 公式 (2.36)。类似地，由 Itô 等距性质可得：对任意 $g \in C^0([0,1]; (C_0^0(\mathbb{R}_d))^d)$ 和 $t \in [0,1]$，有

$$\mathbb{E}^x \int_{1-t}^{1} g(s, Z_s^{\mathsf{F}}) \cdot dW_s = 0, \qquad \mathbb{E}^x \left| \int_{1-t}^{1} g(s, Z_s^{\mathsf{F}}) \cdot dW_s \right|^2 = \int_{1-t}^{1} \mathbb{E}^x |g(s, Z_s^{\mathsf{F}})|^2 ds. \tag{B.37}$$

S425 · Original / 原文 · PDF p.63 · A2.SS2.p2.11

Written in terms of X_t^{B}, these are (2.37). ■

用 X_t^{B} 表示时，这些关系就是式 (2.37)。■

Derivation of (2.30).

式（2.30）的推导

S426 · Original / 原文 · PDF p.63 · A2.SS2.SSS0.Px1.p1.1

For any $t \in [0,1]$ the score is the minimizer of

对任意 $t \in [0,1]$，得分函数是下列目标的极小化解：

$$\begin{aligned} & \int_{\mathbb{R}^d} |\hat{s}(t,x) - \nabla \log \rho(t,x)|^2 \rho(t,x) dx \\ = & \int_{\mathbb{R}^d} \left(|\hat{s}(t,x)|^2 - 2\hat{s}(t,x) \cdot \nabla \log \rho(t,x) + |\nabla \log \rho(t,x)|^2 \right) \rho(t,x) dx \\ = & \int_{\mathbb{R}^d} \left(|\hat{s}(t,x)|^2 + 2\nabla \cdot \hat{s}(t,x) + |\nabla \log \rho(t,x)|^2 \right) \rho(t,x) dx, \end{aligned} \tag{B.38}$$

S427 · Original / 原文 · PDF p.63 · A2.SS2.SSS0.Px1.p1.2

where we used the identity $\hat{s} \cdot \nabla \log \rho \, \rho = \hat{s} \cdot \nabla \rho$ and integration by parts to obtain the second equality. The last term involving $|\nabla \log \rho|^2$ is a constant in $\hat{s}$ that can be neglected for optimization. Expressing the remaining terms as an expectation over x_t and integrating the result in time gives (2.30).

其中，为得到第二个等号，我们使用了恒等式 $\hat{s} \cdot \nabla \log \rho \, \rho = \hat{s} \cdot \nabla \rho$，并进行了分部积分。含 $|\nabla \log \rho|^2$ 的最后一项与 $\hat{s}$ 无关，在优化时可以忽略。将其余各项表示为对 x_t 的期望，再对时间积分，即得到式 (2.30)。

B.3 Proofs of Lemmas 21 and 22, and Theorem 23.

B.3 引理 21、22 与定理 23 的证明

S428 · Original / 原文 · PDF p.63 · A2.SS3.p1.1

See 21

见 21。

Proof.

证明

S429 · Original / 原文 · PDF p.63 · A2.SS3.p2.1

Using (2.38), we compute analytically

利用式 (2.38)，我们可解析地计算出

$$
\begin{aligned}
\frac{d}{dt}\mathrm{KL}(\rho(t) \,\|\, \hat{\rho}(t)) &= \frac{d}{dt} \int_{\mathbb{R}^d} \log\left(\frac{\rho}{\hat{\rho}}\right) \rho dx \\
&= \int_{\mathbb{R}^d} \hat{\rho} \left(\frac{\partial_t \rho}{\hat{\rho}} - \frac{\rho}{(\hat{\rho})^2} \partial_t \hat{\rho} \right) dx + \int \log\left(\frac{\rho}{\hat{\rho}}\right) \partial_t \rho dx \\
&= -\int_{\mathbb{R}^d} \left(\frac{\rho}{\hat{\rho}}\right) \partial_t \hat{\rho} dx + \int_{\mathbb{R}^d} \log\left(\frac{\rho}{\hat{\rho}}\right) \partial_t \rho dx \\
&= \int_{\mathbb{R}^d} \left(\frac{\rho}{\hat{\rho}}\right) \nabla \cdot \left(\hat{b}\hat{\rho}\right) dx - \int \log\left(\frac{\rho}{\hat{\rho}}\right) \nabla \cdot (b\rho) \, dx \\
&= -\int_{\mathbb{R}^d} \nabla\left(\frac{\rho}{\hat{\rho}}\right) \cdot \hat{b}\hat{\rho} dx + \int (\nabla \log \rho - \nabla \log \hat{\rho}) \cdot b\rho dx \\
&= -\int_{\mathbb{R}^d} \left(\frac{\nabla \rho}{\hat{\rho}} - \frac{\rho \nabla \hat{\rho}}{\hat{\rho}^2} \right) \cdot \hat{b}\hat{\rho} dx + \int_{\mathbb{R}^d} (\nabla \log \rho - \nabla \log \hat{\rho}) \cdot b\rho dx \\
&= \int_{\mathbb{R}^d} (\nabla \log \hat{\rho} - \nabla \log \rho) \cdot \hat{b}\rho dx + \int_{\mathbb{R}^d} (\nabla \log \rho - \nabla \log \hat{\rho}) \cdot b\rho dx \\
&= \int_{\mathbb{R}^d} (\nabla \log \hat{\rho} - \nabla \log \rho) \cdot (\hat{b} - b)\rho dx.
\end{aligned}
$$

S430 · Original / 原文 · PDF p.64 · A2.SS3.p2.2

where we omitted the argument (t, x) of all functions for simplicity of notation. Integrating both sides from 0 to 1 completes the proof. ■

其中为简化记号，省略了所有函数的自变量 (t, x)。对等式两侧从 0 到 1 积分，即完成证明。■

S431 · Original / 原文 · PDF p.64 · A2.SS3.p3.1

See 22

见 22。

Proof.

证明

S432 · Original / 原文 · PDF p.64 · A2.SS3.p4.2

Similar to the proof of Lemma 21, we can use the FPE in (2.40) to compute $\frac{d}{dt}\mathsf{KL}(\rho(t) \,\|\, \hat{\rho}(t))$, which leads to the main result. Instead, we take a simpler approach, leveraging the result in Lemma 21. We re-write the Fokker-Planck equations in (2.40) as the (score-dependent) transport equations

与引理 21 的证明类似，我们可以利用式 (2.40) 中的 Fokker–Planck 方程计算 $\frac{d}{dt}\mathsf{KL}(\rho(t) \,\|\, \hat{\rho}(t))$，从而得到主要结论。这里我们采用更简单的办法，直接利用引理 21 的结果。将式 (2.40) 中的 Fokker–Planck 方程改写为下列依赖得分函数的输运方程：

$$\begin{aligned}\partial_t \rho &= -\nabla \cdot ((b_{\mathsf{F}} - \epsilon \nabla \log \rho)\rho),\\ \partial_t \hat{\rho} &= -\nabla \cdot ((\hat{b}_{\mathsf{F}} - \epsilon \nabla \log \hat{\rho})\hat{\rho}).\end{aligned} \tag{B.39}$$

S433 · Original / 原文 · PDF p.64 · A2.SS3.p4.3

Applying Lemma 21 directly, we find that

直接应用引理 21，得到

$$\mathsf{KL}(\rho(1) \,\|\, \hat{\rho}(1)) = \int_0^1 \int_{\mathbb{R}^d} \left[(\nabla \log \hat{\rho} - \nabla \log \rho) \cdot \left(\left[\hat{b}_{\mathsf{F}} - \epsilon \nabla \log \hat{\rho}\right] - \left[b_{\mathsf{F}} - \epsilon \nabla \log \rho\right]\right)\right] \rho dx dt. \tag{B.40}$$

S434 · Original / 原文 · PDF p.64 · A2.SS3.p4.4

Expanding the above,

展开上式，

$$\begin{aligned}\mathsf{KL}(\rho(1) \,\|\, \hat{\rho}(1)) = &\int_0^1 \int_{\mathbb{R}^d} \left[(\nabla \log \hat{\rho} - \nabla \log \rho) \cdot (\hat{b}_{\mathsf{F}} - b_{\mathsf{F}})\right] \rho dx dt\\ &-\epsilon \int_0^1 \int_{\mathbb{R}^d} \left[|\nabla \log \rho - \nabla \log \hat{\rho}|^2\right] \rho dx dt,\end{aligned} \tag{B.41}$$

S435 · Original / 原文 · PDF p.65 · A2.SS3.p4.5

which proves the first part of the lemma. Now, by Young's inequality, it holds for any fixed $\eta > 0$ that

这就证明了引理的第一部分。现在，由 Young 不等式可知，对任意固定的 $\eta > 0$，都有

$$\begin{aligned}\mathsf{KL}(\rho_1 \,\|\, \hat{\rho}(1)) \le &\frac{1}{2\eta} \int_0^1 \int_{\mathbb{R}^d} \left|\hat{b}_{\mathsf{F}} - b_{\mathsf{F}}\right|^2 \rho_t dx dt\\ &+ \left(\tfrac{1}{2}\eta - \epsilon\right) \int_0^1 \int_{\mathbb{R}^d} |\nabla \log \rho_t - \nabla \log \hat{\rho}_t|^2 \rho_t dx dt.\end{aligned} \tag{B.42}$$

S436 · Original / 原文 · PDF p.65 · A2.SS3.p4.6

Hence, for $\eta = 2\epsilon$,

因此，取 $\eta = 2\epsilon$ 时，

$$\mathsf{KL}(\rho_1 \,\|\, \hat{\rho}(1)) \le \frac{1}{4\epsilon} \int_0^1 \int_{\mathbb{R}^d} \left|\hat{b}_{\mathsf{F}} - b_{\mathsf{F}}\right|^2 \rho dx dt, \tag{B.43}$$

S437 · Original / 原文 · PDF p.65 · A2.SS3.p4.7

which proves the second part of the lemma. ■

这就证明了引理的第二部分。■

S438 · Original / 原文 · PDF p.65 · A2.SS3.p5.1

See 23

见 23。

Proof.

证明

S439 · Original / 原文 · PDF p.66 · A2.SS3.p6.1

Observe that by Proposition 18, the target density $\rho_1 = \rho(1,\cdot)$ is the density of the process X_t that evolves according to SDE (2.33). By Lemma 22 and an additional application of Young's inequality, we then have that

注意，由命题 18 可知，目标密度 $\rho_1 = \rho(1,\cdot)$ 是按照随机微分方程 (2.33) 演化的过程 X_t 的概率密度。由引理 22，再应用一次 Young 不等式，便有

$$\mathsf{KL}(\rho_1 \,\|\, \hat{\rho}(1)) \leq \frac{1}{2\epsilon}\int_0^1 \int_{\mathbb{R}^d} \left(|\hat{b}-b|^2 + \epsilon^2\,|s-\hat{s}|^2\right)\rho dx dt. \tag{B.44}$$

S440 · Original / 原文 · PDF p.66 · A2.SS3.p6.2

Using the definition of $\mathcal{L}_b[\hat{b}]$ and $\mathcal{L}_s[\hat{s}]$ in (2.13), and (2.16), we conclude that

利用式 (2.13) 和式 (2.16) 中 $\mathcal{L}_b[\hat{b}]$ 与 $\mathcal{L}_s[\hat{s}]$ 的定义，我们得到

$$\mathsf{KL}(\rho_1 \,\|\, \hat{\rho}(1)) \leq \frac{1}{2\epsilon}\left(\mathcal{L}_b[\hat{b}] - \min_{\hat{b}} \mathcal{L}_b[\hat{b}]\right) + \frac{\epsilon}{2}\left(\mathcal{L}_s[\hat{s}] - \min_{\hat{s}} \mathcal{L}_s[\hat{s}]\right), \tag{B.45}$$

S441 · Original / 原文 · PDF p.66 · A2.SS3.p6.3

which is (2.45). Using that $b = v - \gamma\dot{\gamma}s$ and $\hat{b} = \hat{v} - \gamma\dot{\gamma}\hat{s}$, we can write (B.43) in this case as

这就是式 (2.45)。利用 $b = v - \gamma\dot{\gamma}s$ 和 $\hat{b} = \hat{v} - \gamma\dot{\gamma}\hat{s}$，在这种情况下可将式 (B.43) 写成

$$\begin{aligned}
\mathsf{KL}(\rho_1 \,\|\, \hat{\rho}(1)) &\leq \frac{1}{4\epsilon}\int_0^1 \int_{\mathbb{R}^d} \left|\hat{v} - v - (\gamma(t)\dot{\gamma}(t) - \epsilon)(\hat{s}-s)\right|^2 \rho dx dt, && \text{(B.46)}\\
&\leq \frac{1}{2\epsilon}\int_0^1 \int_{\mathbb{R}^d} \left(|\hat{v}-v|^2 + (\gamma(t)\dot{\gamma}(t)-\epsilon)^2\,|\hat{s}-s|^2\right)\rho dx dt, && \text{(B.47)}\\
&\leq \frac{1}{2\epsilon}\left(\mathcal{L}_v[\hat{v}] - \min_{\hat{v}} \mathcal{L}_v[\hat{v}]\right) + \frac{\max_{t\in[0,1]}(\gamma(t)\dot{\gamma}(t)-\epsilon)^2}{2\epsilon}\left(\mathcal{L}_s[\hat{s}] - \min_{\hat{v}} \mathcal{L}_s[\hat{s}]\right), && \text{(B.48)}
\end{aligned}$$

S442 · Original / 原文 · PDF p.66 · A2.SS3.p6.4

which is the final part of the theorem. ■

这就是该定理的最后一部分。■

B.4 Proofs of Lemma 25 and Theorem 26

B.4 引理 25 与定理 26 的证明

S443 · Original / 原文 · PDF p.66 · A2.SS4.p1.1

See 25

见 25。

Proof.

证明

S444 · Original / 原文 · PDF p.67 · A2.SS4.p2.5

If $\hat{\rho}$ solves the TE (2.50) and $X_{s,t}$ solves the ODE (2.51), we have

如果 $\hat{\rho}$ 满足输运方程 (2.50)，且 $X_{s,t}$ 满足常微分方程 (2.51)，则有

$$\begin{aligned}
\frac{d}{dt}\hat{\rho}(t, X_{s,t}(x)) &= \partial_t\hat{\rho}(t, X_{s,t}(x)) + b(t, X_{s,t}(x))\cdot\nabla\hat{\rho}(t, X_{s,t}(x))\\
&= -\nabla\cdot b(t, X_{s,t}(x))\hat{\rho}(t, X_{s,t}(x))
\end{aligned} \tag{B.49}$$

S445 · Original / 原文 · PDF p.67 · A2.SS4.p2.6

This equation implies that

这个等式意味着

$$\frac{d}{dt}\left(\exp\left(\int_s^t \nabla\cdot b(\tau, X_{s,\tau}(x))d\tau\right)\hat{\rho}(t, X_{s,t}(x))\right) = 0. \quad \text{(B.50)}$$

S446 · Original / 原文 · PDF p.67 · A2.SS4.p2.7

Integrating (B.50) over $[0,t]$ and setting $s=t$ in the result gives

在 $[0,t]$ 上对式 (B.50) 积分，并在所得结果中令 $s=t$，得到

$$\hat{\rho}(t,x) = \exp\left(-\int_0^t \nabla\cdot b(\tau, X_{t,\tau}(x))d\tau\right)\hat{\rho}(0, X_{t,0}(x)) \quad \text{(B.51)}$$

S447 · Original / 原文 · PDF p.67 · A2.SS4.p2.8

If we use the initial condition $\hat{\rho}(0)=\rho_0$, this gives (2.52). Similarly, Integrating (B.50) on $[t,1]$ and setting $s=t$ in the result gives

若使用初始条件 $\hat{\rho}(0)=\rho_0$，就得到式 (2.52)。类似地，在 $[t,1]$ 上对式 (B.50) 积分，并在所得结果中令 $s=t$，得到

$$\hat{\rho}(t,x) = \exp\left(\int_t^1 \nabla\cdot b(\tau, X_{t,\tau}(x))d\tau\right)\hat{\rho}(1, X_{t,1}(x)) \quad \text{(B.52)}$$

S448 · Original / 原文 · PDF p.67 · A2.SS4.p2.9

If we use the final condition $\hat{\rho}(1)=\rho_1$, this gives (2.53).

若使用终端条件 $\hat{\rho}(1)=\rho_1$，就得到式 (2.53)。

S449 · Original / 原文 · PDF p.67 · A2.SS4.p3.1

■

■

S450 · Original / 原文 · PDF p.67 · A2.SS4.p4.1

See 26

见 26。

Proof.

证明

S451 · Original / 原文 · PDF p.67 · A2.SS4.p5.4

Evaluating $d\hat{\rho}_{\mathsf{F}}(t, Y_t^{\mathsf{B}})$ via the backward Itô formula (2.36) we obtain

利用反向 Itô 公式 (2.36) 计算 $d\hat{\rho}_{\mathsf{F}}(t, Y_t^{\mathsf{B}})$，得到

$$\begin{aligned} d\hat{\rho}_{\mathsf{F}}(t, Y_t^{\mathsf{B}}) &= \partial_t\hat{\rho}_{\mathsf{F}}(t, Y_t^{\mathsf{B}})dt + \nabla\hat{\rho}_{\mathsf{F}}(t, Y_t^{\mathsf{B}})\cdot dY_t^{\mathsf{B}} - \epsilon\Delta\hat{\rho}_{\mathsf{F}}(t, Y_t^{\mathsf{B}})dt \\ &= \partial_t\hat{\rho}_{\mathsf{F}}(t, Y_t^{\mathsf{B}})dt + \nabla\hat{\rho}_{\mathsf{F}}(t, Y_t^{\mathsf{B}})\cdot \hat{b}_{\mathsf{F}}(t, Y_t^{\mathsf{B}})dt + \sqrt{2\epsilon}\nabla\hat{\rho}_{\mathsf{F}}(t, Y_t^{\mathsf{B}})\cdot dW_t^{\mathsf{B}} - \epsilon\Delta\hat{\rho}_{\mathsf{F}}(t, Y_t^{\mathsf{B}})dt \\ &= -\nabla\cdot\hat{b}_{\mathsf{F}}(t, Y_t^{\mathsf{B}})\hat{\rho}_{\mathsf{F}}(t, Y_t^{\mathsf{B}})dt + \sqrt{2\epsilon}\nabla\hat{\rho}_{\mathsf{F}}(t, Y_t^{\mathsf{B}})\cdot dW_t^{\mathsf{B}}. \end{aligned} \quad \text{(B.53)}$$

S452 · Original / 原文 · PDF p.68 · A2.SS4.p5.5

where we used (2.56) in the second step and (2.57) in the last one. This equation can be written as a total differential in the form

其中，第二步使用了式 (2.56)，最后一步使用了式 (2.57)。这个方程可以写成如下全微分形式：

$$
\begin{aligned}
&d\left(\exp\left(-\int_t^1 \nabla\cdot\hat{b}_{\mathsf{F}}(\tau,Y_\tau^{\mathsf{B}})d\tau\right)\hat{\rho}_{\mathsf{F}}(t,Y_t^{\mathsf{B}})\right)\\
&=\sqrt{2\epsilon}\exp\left(-\int_t^1 \nabla\cdot\hat{b}_{\mathsf{F}}(\tau,Y_\tau^{\mathsf{B}})d\tau\right)\nabla\hat{\rho}_{\mathsf{F}}(t,Y_t^{\mathsf{B}})\cdot dW_t^{\mathsf{B}},
\end{aligned}
\tag{B.54}
$$

S453 · Original / 原文 · PDF p.68 · A2.SS4.p5.6

which after integration over $t\in[0,1]$ becomes

在 $t\in[0,1]$ 上积分后，它成为

$$
\begin{aligned}
&\hat{\rho}_{\mathsf{F}}(1,Y_{t=1}^{\mathsf{B}})-\exp\left(-\int_0^1 \nabla\cdot\hat{b}_{\mathsf{F}}(t,Y_t^{\mathsf{B}})dt\right)\rho_0(Y_{t=0}^{\mathsf{B}})\\
&=\sqrt{2\epsilon}\int_0^1\exp\left(-\int_t^1 \nabla\cdot\hat{b}_{\mathsf{F}}(\tau,Y_\tau^{\mathsf{B}})d\tau\right)\nabla\hat{\rho}(t,Y_t^{\mathsf{B}})\cdot dW_t^{\mathsf{B}}.
\end{aligned}
\tag{B.55}
$$

S454 · Original / 原文 · PDF p.68 · A2.SS4.p5.7

where we used $\hat{\rho}(0)=\rho_0$. Taking an expectation conditioned on the event $Y_{t=1}^{\mathsf{B}}=x$ and using that the term on the right-hand side has mean zero, we find that

其中使用了 $\hat{\rho}(0)=\rho_0$。以事件 $Y_{t=1}^{\mathsf{B}}=x$ 为条件取期望，并利用右侧项的均值为零，可得

$$
\hat{\rho}_{\mathsf{F}}(1,Y_{t=1}^{\mathsf{B}})-\mathbb{E}_{\mathsf{B}}^{x}\exp\left(-\int_0^1 \nabla\cdot\hat{b}_{\mathsf{F}}(t,Y_t^{\mathsf{B}})dt\right)\rho_0(Y_{t=0}^{\mathsf{B}})=0.
\tag{B.56}
$$

S455 · Original / 原文 · PDF p.68 · A2.SS4.p5.8

This gives (2.58).

这就得到式 (2.58)。

S456 · Original / 原文 · PDF p.68 · A2.SS4.p6.4

Similarly, evaluating $d\hat{\rho}_{\mathsf{B}}(t,Y_t^{\mathsf{F}})$ via Itô's formula, we obtain

类似地，利用 Itô 公式计算 $d\hat{\rho}_{\mathsf{B}}(t,Y_t^{\mathsf{F}})$，得到

$$
\begin{aligned}
d\hat{\rho}_{\mathsf{B}}(t,Y_t^{\mathsf{F}})&=\partial_t\hat{\rho}_{\mathsf{B}}(t,Y_t^{\mathsf{F}})dt+\nabla\hat{\rho}_{\mathsf{B}}(t,Y_t^{\mathsf{F}})\cdot dY_t^{\mathsf{F}}+\epsilon\Delta\hat{\rho}_{\mathsf{B}}(t,Y_t^{\mathsf{F}})dt\\
&=\partial_t\hat{\rho}_{\mathsf{B}}(t,Y_t^{\mathsf{F}})dt+\nabla\hat{\rho}_{\mathsf{B}}(t,Y_t^{\mathsf{F}})\cdot\hat{b}_{\mathsf{B}}(t,Y_t^{\mathsf{F}})dt+\sqrt{2\epsilon}\nabla\hat{\rho}_{\mathsf{B}}(t,Y_t^{\mathsf{F}})\cdot dW_t+\epsilon\Delta\hat{\rho}_{\mathsf{B}}(t,Y_t^{\mathsf{F}})dt\\
&=-\nabla\cdot\hat{b}_{\mathsf{B}}(t,Y_t^{\mathsf{F}})\hat{\rho}_{\mathsf{B}}(t,Y_t^{\mathsf{F}})dt+\sqrt{2\epsilon}\nabla\hat{\rho}_{\mathsf{B}}(t,Y_t^{\mathsf{F}})\cdot dW_t.
\end{aligned}
\tag{B.57}
$$

S457 · Original / 原文 · PDF p.69 · A2.SS4.p6.5

where we used (2.55) in the second step and (2.59) in the last one. This equation can be written as a total differential in the form

其中，第二步使用了式 (2.55)，最后一步使用了式 (2.59)。这个方程可以写成如下全微分形式：

$$
\begin{aligned}
&d\left(\exp\left(\int_0^t \nabla\cdot\hat{b}_{\mathsf{B}}(\tau,Y_\tau^{\mathsf{F}})d\tau\right)\hat{\rho}_{\mathsf{B}}(t,Y_t^{\mathsf{F}})\right)\\
&=\sqrt{2\epsilon}\exp\left(\int_0^t \nabla\cdot\hat{b}_{\mathsf{B}}(\tau,Y_\tau^{\mathsf{F}})d\tau\right)\nabla\hat{\rho}_{\mathsf{B}}(t,Y_t^{\mathsf{F}})\cdot dW_t.
\end{aligned}
\tag{B.58}
$$

S458 · Original / 原文 · PDF p.69 · A2.SS4.p6.6

Integrating the above on $t\in[0,1]$, we find that

在 $t\in[0,1]$ 上对上式积分，得到

$$\exp\left(\int_0^1 \nabla\cdot\hat{b}_{\mathsf{B}}(t,Y_t^{\mathsf{F}})dt\right)\rho_1(Y_1^{\mathsf{F}})-\hat{\rho}_{\mathsf{B}}(0,Y_{t=0}^{\mathsf{F}})$$
$$=\sqrt{2\epsilon}\int_0^1 \exp\left(\int_0^t \nabla\cdot\hat{b}_{\mathsf{B}}(\tau,Y_\tau^{\mathsf{F}})d\tau\right)\nabla\hat{\rho}_{\mathsf{B}}(t,Y_t^{\mathsf{F}})\cdot dW_t. \tag{B.59}$$

S459 · Original / 原文 · PDF p.69 · A2.SS4.p6.7

where we used $\hat{\rho}(1)=\rho_1$. Taking an expectation conditioned on the event $Y_{t=0}^{\mathsf{F}}=x$ and applying the Itô isometry, we deduce that

其中使用了 $\hat{\rho}(1)=\rho_1$。以事件 $Y_{t=0}^{\mathsf{F}}=x$ 为条件取期望，并应用 Itô 等距性质，推得

$$\mathbb{E}_{\mathsf{F}}^{x}\left(\exp\left(\int_0^1 \nabla\cdot\hat{b}_{\mathsf{B}}(t,Y_t^{\mathsf{F}})dt\right)\rho_1(Y_{t=1}^{\mathsf{F}})\right)-\hat{\rho}_{\mathsf{B}}(0,y)=0. \tag{B.60}$$

S460 · Original / 原文 · PDF p.69 · A2.SS4.p6.8

This gives (2.60).

这就得到式 (2.60)。

S461 · Original / 原文 · PDF p.69 · A2.SS4.p7.1

■

■

B.5 Proof of Theorem 31

B.5 定理 31 的证明

S462 · Original / 原文 · PDF p.69 · A2.SS5.p1.1

See 31

见 31。

Proof.

证明

S463 · Original / 原文 · PDF p.69 · A2.SS5.p2.1

Let us first consider what happens on the interval $t\in[0,\delta]$ where $I(t,x_0,x_1)=x_0$ by assumption and the stochastic interpolant (3.2) with x_0 fixed and $a(t)=a>0$ reduces to

先考虑区间 $t\in[0,\delta]$ 上的情形。按照假设，此时 $I(t,x_0,x_1)=x_0$；固定 x_0 且令 $a(t)=a>0$ 后，随机插值式 (3.2) 简化为

$$x_t=x_0+\sqrt{2at(1-t)}z\quad\text{with}\quad x_0\text{fixed},\ z\sim\mathsf{N}(0,Id). \tag{B.61}$$

S464 · Original / 原文 · PDF p.69 · A2.SS5.p2.2

The law of this process is simply

这个过程的分布就是

$$x_t\sim\mathsf{N}(x_0,2at(1-t)Id), \tag{B.62}$$

S465 · Original / 原文 · PDF p.70 · A2.SS5.p2.3

which means that its density satisfies for all $t > 0$ the FPE

这意味着其概率密度对所有 $t > 0$ 都满足下列 Fokker–Planck 方程：

$$\partial_t \rho + 2at\nabla \cdot (s(t,x)\rho) = a\Delta\rho, \tag{B.63}$$

S466 · Original / 原文 · PDF p.70 · A2.SS5.p2.4

where $s(t,x) = \nabla \log \rho(t,x)$ is the score, which is explicitly given by

其中，$s(t,x) = \nabla \log \rho(t,x)$ 是得分函数，其显式表达为

$$s(t,x) = -\frac{x - x_0}{2at(1-t)} \qquad \text{for } \ t \in (0,\delta]. \tag{B.64}$$

S467 · Original / 原文 · PDF p.70 · A2.SS5.p2.5

This means that the drift term in the FPE (B.63) can be written as

这意味着 Fokker–Planck 方程 (B.63) 中的漂移项可以写成

$$2ats(t,x) = -\frac{x - x_0}{1-t} \qquad \text{for } \ t \in (0,\delta]. \tag{B.65}$$

S468 · Original / 原文 · PDF p.70 · A2.SS5.p2.6

and this equality also holds in the limit as $t \to 0$. It is also easy to check that

并且当 $t \to 0$ 时，这个等式在极限下依然成立。此外，也容易验证

$$-\frac{x - x_0}{2at(1-t)} = -\frac{\sqrt{2at}}{\sqrt{(1-t)}}\mathbb{E}(z|x_t = x) \qquad \text{for } \ t \in [0,\delta] \tag{B.66}$$

S469 · Original / 原文 · PDF p.70 · A2.SS5.p2.7

which is consistent with the drift in (3.10) on $t \in [0,\delta]$ since $\partial_t I(t,x_0,x_1) = 0$ on this interval. Since the right-hand side of (B.65) is nonsingular at $t = 0$ it also means that the SDE (3.11) is well-defined for $t \in [0,\delta]$ and the law of its solutions coincide with that of the process x_t defined (B.61), $X_t^{\mathsf{d}} \sim \mathsf{N}(x_0, 2at(1-t)Id)$ for $t \in [0,\delta]$.

这与区间 $t \in [0,\delta]$ 上式 (3.10) 中的漂移项一致，因为在该区间上 $\partial_t I(t,x_0,x_1) = 0$。由于式 (B.65) 的右侧在 $t = 0$ 处没有奇异性，这还说明随机微分方程 (3.11) 在 $t \in [0,\delta]$ 上有良好定义，而且其解的分布与式 (B.61) 定义的过程 x_t 的分布相同，即对于 $t \in [0,\delta]$，有 $X_t^{\mathsf{d}} \sim \mathsf{N}(x_0, 2at(1-t)Id)$。

S470 · Original / 原文 · PDF p.70 · A2.SS5.p3.1

Considering next what happens on $t \in (\delta,1]$, since $\gamma(t) = \sqrt{2at(1-t)}$ is only zero at the endpoint $t = 1$ where we assume that the density ρ_1 satisfies Assumption 5 we can mimick all the arguments in the proof of of Theorem 6 and Corollaries 10 and 18 to terminate the proof. ■

接下来考虑 $t \in (\delta,1]$ 上的情形。由于 $\gamma(t) = \sqrt{2at(1-t)}$ 仅在端点 $t = 1$ 处为零，而我们假设该端点的密度 ρ_1 满足假设 5，因此可以仿照定理 6 以及推论 10、18 证明中的全部论证，完成证明。■

S471 · Original / 原文 · PDF p.70 · A2.SS5.p4.1

Note that this proof shows that is enough to have $\partial_t I(t=0,x_0,x_1) = 0$, since this implies that $I(t,x_0,x_1) = x_0 + O(t^2)$. It also shows that, while it is key to use an SDE on $t \in [0,\delta]$ so that the generative process can spread the mass away from x_0, the

diffusive noise is no longer necessary and we could switch back to a probability flow ODE on $t \in (\delta, 1]$ (using a time-dependent $a(t)$ to that effect with $a(t) = 0$ for $t \in (\delta, 1]$).

注意，这个证明表明，只要有 $\partial_t I(t = 0, x_0, x_1) = 0$ 就足够了，因为它蕴含 $I(t, x_0, x_1) = x_0 + O(t^2)$。证明还表明，虽然在 $t \in [0, \delta]$ 上使用随机微分方程至关重要，以便生成过程能将质量从 x_0 向外扩散，但此后扩散噪声不再必要，我们可以在 $t \in (\delta, 1]$ 上切换回概率流常微分方程；为此可使用随时间变化的 $a(t)$，并令其在 $t \in (\delta, 1]$ 上满足 $a(t) = 0$。

B.6 Proof of Lemma 40 and Theorem 41.

B.6 引理 40 与定理 41 的证明

S472 · Original / 原文 · PDF p.70 · A2.SS6.p1.1

See 40

见 40。

Proof.

证明

S473 · Original / 原文 · PDF p.70 · A2.SS6.p2.1

By definition of the map T in (3.36), if $x_0 \sim \rho_0$ and $x_1 \sim \rho_1$, then $T^{-1}(0, x_0) \sim \mathsf{N}(0, Id)$ and $T^{-1}(1, x_1) \sim \mathsf{N}(0, Id)$. As a result, since x_0, x_1, and z are independent, and $z \sim \mathsf{N}(0, Id)$, we have

根据式 (3.36) 中映射 T 的定义，若 $x_0 \sim \rho_0$ 且 $x_1 \sim \rho_1$，则 $T^{-1}(0, x_0) \sim \mathsf{N}(0, Id)$，$T^{-1}(1, x_1) \sim \mathsf{N}(0, Id)$。因此，利用 x_0、x_1 与 z 相互独立，且 $z \sim \mathsf{N}(0, Id)$，可得

$$\alpha(t)T^{-1}(0, x_0) + \beta(t)T^{-1}(1, x_1) + \gamma(t)z \sim \mathsf{N}(0, \alpha^2(t)Id + \beta^2(t)Id + \gamma^2(t)Id) = \mathsf{N}(0, Id) \tag{B.67}$$

S474 · Original / 原文 · PDF p.71 · A2.SS6.p2.2

where the second equality follows from the condition $\alpha^2(t) + \beta^2(t) + \gamma^2(t) = 1$. Therfore, using again the definition of the map T

其中第二个等号由条件 $\alpha^2(t) + \beta^2(t) + \gamma^2(t) = 1$ 得到。因此，再次利用映射 T 的定义，

$$T(t, \alpha(t)T^{-1}(0, x_0) + \beta(t)T^{-1}(1, x_1) + \gamma(t)z) \sim \rho(t), \tag{B.68}$$

S475 · Original / 原文 · PDF p.71 · A2.SS6.p2.3

and we are done. ■

证明完毕。■

S476 · Original / 原文 · PDF p.71 · A2.SS6.p3.1

See 41

见 41。

Proof.

证明

S477 · Original / 原文 · PDF p.71 · A2.SS6.p4.4

Denote by $\hat{\rho}(t, x)$ the density of $\hat{x}_t$ and define the current $\hat{\jmath} : [0, 1] \times \mathbb{R}^d \to \mathbb{R}^d$ as

用 $\hat{\rho}(t, x)$ 表示 $\hat{x}_t$ 的密度，并将概率流 $\hat{\jmath} : [0, 1] \times \mathbb{R}^d \to \mathbb{R}^d$ 定义为

$$\hat{\jmath}(t,x) = \mathbb{E}\big(\partial_t \hat{I}_t + \gamma(t) z | x = \hat{x}_t\big)\hat{\rho}(t,x). \tag{B.69}$$

S478 · Original / 原文 · PDF p.71 · A2.SS6.p4.5

The max-min problem (3.38) can then be formulated as the constrained optimization problem:

于是，最大—最小问题 (3.38) 可以表述为下列约束优化问题：

$$\begin{aligned} &\max_{\hat{\rho},\hat{\jmath}} \min_{\hat{u}} \int_0^1 \int_{\mathbb{R}^d} \left(\tfrac{1}{2}|\hat{u}(t,x)|^2 \hat{\rho}(t,x) - \hat{u}(t,x)\cdot\hat{\jmath}(t,x)\right) dxdt \\ &\text{subject to:} \partial_t \hat{\rho} + \nabla\cdot\hat{\jmath} = \epsilon\Delta\hat{\rho}, \quad \hat{\rho}(t=0) = \rho_0, \quad \hat{\rho}(t=1) = \rho_1 \end{aligned} \tag{B.70}$$

S479 · Original / 原文 · PDF p.71 · A2.SS6.p4.6

To solve this problem we can use the extended objective

为求解此问题，可以使用扩展目标泛函

$$\begin{aligned} \max_{\hat{\rho},\hat{\jmath}} \min_{\hat{u}} \Bigg(& \int_0^1 \int_{\mathbb{R}^d} \left(\tfrac{1}{2}|\hat{u}(t,x)|^2 \hat{\rho}(t,x) - \hat{u}(t,x)\cdot\hat{\jmath}(t,x)\right) dxdt \\ & - \int_0^1 \int_{\mathbb{R}^d} \lambda(t,x)\left(\partial_t\hat{\rho}(t,x) + \nabla\cdot\hat{\jmath}(t,x) - \epsilon\Delta\hat{\rho}(t,x)\right) dxdt \\ & + \int_{\mathbb{R}^d} \eta_0(x)\left(\hat{\rho}(0,x) - \rho_0(x)\right) dx - \int_{\mathbb{R}^d} \eta_1(x)\left(\hat{\rho}(1,x) - \rho_1(x)\right) dx \Bigg) \end{aligned} \tag{B.71}$$

S480 · Original / 原文 · PDF p.71 · A2.SS6.p4.7

where $\lambda(t,x)$, $\eta_0(x)$, and $\eta_1(x)$ are Lagrange multipliers used to enforce the constraints. The unique minimizer (ρ, j, λ) of this optimization problem solves the Euler-Lagrange equations:

其中，$\lambda(t,x)$、$\eta_0(x)$ 和 $\eta_1(x)$ 是用于施加约束的拉格朗日乘子。这个优化问题的唯一极小化解 (ρ, j, λ) 满足如下 Euler–Lagrange 方程：

$$\begin{aligned} &\partial_t\rho + \nabla\cdot j = \epsilon\Delta\rho, \quad \rho(t=0) = \rho_0 \quad \rho(t=1) = \rho_1 \\ &\partial_t\lambda + \tfrac{1}{2}|u|^2 = -\epsilon\Delta\lambda, \\ &j = u\rho \\ &u = \nabla\lambda \end{aligned} \tag{B.72}$$

S481 · Original / 原文 · PDF p.72 · A2.SS6.p4.8

We can use the last two equations to write the first two as (3.35), with $u = \nabla\lambda$. Since under Assumption 39 there is an interpolant that realizes the density $\rho(t)$ that solves (3.35), we conclude that an optimizer (I, u) of the the max-min problem (3.38) exists. For any optimizer, I will be such that $\rho(t)$ is the density of $x_t = I(t, x_0, x_1) + \gamma(t)z$, and u will satisfy $u = \nabla\lambda$. ■

利用最后两个方程，可将前两个方程写为式 (3.35)，其中 $u = \nabla\lambda$。在假设 39 下，存在一个插值函数能够实现满足式 (3.35) 的密度 $\rho(t)$，因此我们断定最大—最小问题 (3.38) 存在一个最优解 (I, u)。对于任意最优解，其中的 I 都使得 $\rho(t)$ 是 $x_t = I(t, x_0, x_1) + \gamma(t)z$ 的密度，而 u 满足 $u = \nabla\lambda$。■

B.7 Proof of Theorem 45

B.7 定理 45 的证明

S482 · Original / 原文 · PDF p.72 · A2.SS7.p1.1

See 45

见 45。

Proof.

证明

S483 · Original / 原文 · PDF p.72 · A2.SS7.p2.3

Use

利用

$$\beta(t_{j+1})\beta^{-1}(t_j) = 1 + \dot\beta(t_j)\beta^{-1}(t_i)(t_{j+1}-t_j) + O\big((t_{j+1}-t_j)^2\big) \tag{B.73}$$

S484 · Original / 原文 · PDF p.72 · A2.SS7.p2.4

and

以及

$$\alpha(t_{j+1}) - \alpha(t_j)\beta(t_{j+1})\beta^{-1}(t_j) = \big(\dot\alpha(t_j) - \alpha(t_j)\dot\beta(t_j)\beta^{-1}(t_i)\big)(t_{j+1}-t_j) + O\big((t_{j+1}-t_j)^2\big) \tag{B.74}$$

S485 · Original / 原文 · PDF p.72 · A2.SS7.p2.5

to deduce that (5.14) implies

可推知式 (5.14) 蕴含

$$\begin{aligned} X^{\mathsf{den}}_{j+1} = {} & X^{\mathsf{den}}_j + \dot\beta(t_j)\beta^{-1}(t_j)X^{\mathsf{den}}_j(t_{j+1}-t_j) \\ & + \big(\dot\alpha(t_j) - \alpha(t_j)\dot\beta(t_j)\beta^{-1}(t_i)\big)\eta^{\mathsf{os}}_z(t_j, X^{\mathsf{den}}_j)(t_{j+1}-t_j) + O\big((t_{j+1}-t_j)^2\big). \end{aligned} \tag{B.75}$$

S486 · Original / 原文 · PDF p.72 · A2.SS7.p2.6

or equivalently

或等价地，

$$\frac{X^{\mathsf{den}}_{j+1} - X^{\mathsf{den}}_j}{t_{j+1}-t_j} = \dot\beta(t_j)\beta^{-1}(t_j)X^{\mathsf{den}}_j + \big(\dot\alpha(t_j) - \alpha(t_j)\dot\beta(t_j)\beta^{-1}(t_i)\big)\eta^{\mathsf{os}}_z(t_j, X^{\mathsf{den}}_j) + O(t_{j+1}-t_j). \tag{B.76}$$

S487 · Original / 原文 · PDF p.72 · A2.SS7.p2.7

Taking the limit as $N, j \to \infty$ with $j/N \to t \in [0,1]$, we recover (5.15) and deduce that $X^{\mathsf{den}}_j \to X_t$. ■

令 $N, j \to \infty$，同时满足 $j/N \to t \in [0,1]$，取此极限便恢复式 (5.15)，并推得 $X^{\mathsf{den}}_j \to X_t$。■

S488 · Original / 原文 · PDF p.72 · A2.SS7.p3.1

Note that the proof shows that the result of Theorem 45 also holds if we use a nonuniform grid of times t_j, $j \in \{1, \dots, N\}$.

注意，这个证明表明，如果使用非均匀时间网格 t_j，$j \in \{1, \dots, N\}$，定理 45 的结论仍然成立。

B.8 Proof of Theorem 47

B.8 定理 47 的证明

S489 · Original / 原文 · PDF p.72 · A2.SS8.p1.1

See 47

见 47。

Proof.

证明

S490 · Original / 原文 · PDF p.73 · A2.SS8.p2.1

The first part of the statement can be established by following the same steps as in the proof of Theorem 6 and Corollary 10. For the proof of the second part, use first (5.17) written as $x_t^{\text{rec}} = M(t,z)$ in (5.22) to deduce that

该陈述的第一部分可以按照定理 6 与推论 10 证明中的相同步骤建立。对于第二部分，先将式 (5.17) 写成 $x_t^{\text{rec}} = M(t,z)$，并在式 (5.22) 中使用它，得到

$$b^{\text{rec}}(t, x_t^{\text{rec}}) = \dot{\alpha}(t)N(t, M(t,z)) + \dot{\beta}(t)X_{t-1}(N(t, M(t,z))) = \dot{\alpha}(t)z + \dot{\beta}(t)X_{t-1}(z) = \dot{x}_t^{\text{rec}} \tag{B.77}$$

S491 · Original / 原文 · PDF p.73 · A2.SS8.p2.2

This shows that (5.22) is the unique minimizer of (5.19). Next, use (5.23) written as $X_t^{\text{rec}}(x) = M(t,x)$ in (5.22) to deduce that

这说明式 (5.22) 是式 (5.19) 的唯一极小化解。接着，将式 (5.23) 写成 $X_t^{\text{rec}}(x) = M(t,x)$，并在式 (5.22) 中使用它，得到

$$b^{\text{rec}}(t, X_t^{\text{rec}}(x)) = \dot{\alpha}(t)N(t, M(t,x)) + \dot{\beta}(t)X_{t-1}(N(t, M(t,x))) = \dot{\alpha}(t)x + \dot{\beta}(t)X_{t-1}(x) \tag{B.78}$$

S492 · Original / 原文 · PDF p.73 · A2.SS8.p2.3

This implies that (5.23) solves (5.21), and since the solution of this ODE is unique, we are done. ■

这表明式 (5.23) 满足式 (5.21)。由于该常微分方程的解唯一，证明完毕。■

	ImageNet 32×32	Flowers	Flowers (Mirror)
Dimension	32×32	128×128	128×128
# Training point	1,281,167	315,123	315,123
Batch Size	512	64	64
Training Steps	8×10^5	3.5×10^5	8×10^5
Hidden dim	256	128	128
Attention Resolution	64	64	64
Learning Rate (LR)	0.0002	0.0002	0.0002
LR decay (1k epochs)	0.995	0.995	0.985
U-Net dim mult	[1,2,2,2]	[1,1,2,3,4]	[1,1,2,3,4]
Learned t sinusoidal embedding	Yes	Yes	Yes
t_0, t_f for $t \sim \mathrm{Unif}[t_0, t_f]$, learning η	[0.0, 1.0]	n/a	n/a
t_0, t_f for $t \sim \mathrm{Unif}[t_0, t_f]$, learning s	n/a	[.0002, .9998]	[.0002, .9998]
t_0, t_f when sampling with ODE	[0.0, 1.0]	[0.0001, 0.9999]	[.0001, .9999]
t_0, t_f when sampling with SDE, η	[0.0, 0.97] + denoising	n/a	n/a
t_0, t_f when sampling with SDE, s	n/a	[.0001, .9999]	[.0001, .9999]
$\gamma(t)$ in x_t	$\sqrt{(t(1-t)}$	$\sqrt{(t(1-t)}$	$\sqrt{(10t(1-t)}$
EMA decay rate	0.9999	0.9999	0.9999
EMA start iteration	10000	10000	10000
# GPUs	2	4	2

Table 3: Hyperparameters and architecture for image datasets.

表 3：图像数据集的超参数与网络架构。

Appendix C Experimental Specifications
附录 C 实验配置

S493 · Original / 原文 · PDF p.73 · A3.p1.1

Details for the experiments in Section 7.1 are provided here. Feed forward neural networks of depth 4 and width 512 are used for each model of the velocities b, v, and s. Training was done for 7000 iterations on batches comprised of 25 draws from the base, 400 draws from the target, and 100 time slices. At each iteration, we used a variance reduction technique based on antithetic sampling, in which two samples $\pm z$ are used for each evaluation of the loss. The objectives given in (2.13) and (2.16) were optimized using the Adam optimizer. The learning rate was set to .002 and was dropped by a factor of 2 every 1500 iterations of training. To integrate the ODE/SDE when drawing samples, we used the Heun-based integrator as suggested in *Karras et al. (2022)*.

这里给出第 7.1 节实验的详细设置。对于 b、v 和 s 的每一个模型，均采用深度为 4、宽度为 512 的前馈神经网络。训练共进行 7000 次迭代；每批包含从基分布抽取的 25 个样本、从目标分布抽取的 400 个样本，以及 100 个时间切片。每次迭代都使用基于对偶采样的方差缩减技术，即每次计算损失时都使用两个样本 $\pm z$。式 (2.13) 与式 (2.16) 给出的目标泛函使用 Adam 优化器进行优化。学习率设为 0.002，并在训练中每隔 1500 次迭代减半。在生成样本时，为积分常微分方程或随机微分方程，我们使用了 Karras 等人（2022）建议的基于 Heun 方法的积分器。

C.1 Image Experiments
C.1 图像实验

Network Architectures.
网络架构

S494 · Original / 原文 · PDF p.73 · A3.SS1.SSS0.Px1.p1.1

For all image generation experiments, the U-Net architecture originally proposed in *Ho et al. (2020)* is used. The specification of architecture hyperparameters as well as training hyperparameters are given in Table 3. The same architecture is used regardless of whether learning b, v, s, or η.

所有图像生成实验都采用 Ho 等人（2020）最初提出的 U-Net 架构。架构超参数与训练超参数的具体设置见表 3。不论学习的是 b、v、s 还是 η，均使用相同架构。

S495 · Original / 原文 · PDF p.73 · A3.SS1.SSS0.Px1.p2.1

When using the SDE and learning η_z, we found that integrating to a time slightly before $t_f = 1.0$ and using the denoising formula (5.14) beyond this point provided the best results, as described in the main text.

在使用随机微分方程并学习 η_z 时，我们发现，先积分到略早于 $t_f = 1.0$ 的时刻，再在此后使用去噪公式 (5.14)，能够获得最佳结果，具体如正文所述。

References

参考文献

- Albergo and Vanden-Eijnden (2023) Michael S. Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *International Conference on Learning Representations*, 2023.
- Albergo et al. (2023) Michael S. Albergo, Mark Goldstein, Nicholas M. Boffi, Rajesh Ranganath, and Eric Vanden-Eijnden. Stochastic interpolants with data-dependent couplings. *arXiv*, October 2023.
- Anderson (1982) Brian D.O. Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982.
- Bakry and Émery (1985) Dominique Bakry and Michel Émery. Diffusions hypercontractives. In Jacques Azéma and Marc Yor, editors, *Séminaire de Probabilités XIX 1983/84*, pages 177–206, Berlin, Heidelberg, 1985. Springer Berlin Heidelberg. ISBN 978-3-540-39397-9.
- Ben-Hamu et al. (2022) Heli Ben-Hamu, Samuel Cohen, Joey Bose, Brandon Amos, Maximilian Nickel, Aditya Grover, Ricky T. Q. Chen, and Yaron Lipman. Matching normalizing flows and probability paths on manifolds. In *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, volume 162 of *Proceedings of Machine Learning Research*, pages 1749–1763. PMLR, 2022.
- Benamou and Brenier (2000) Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.
- Boffi and Vanden-Eijnden (2023) Nicholas M Boffi and Eric Vanden-Eijnden. Probability flow solution of the fokker–planck equation. *Machine Learning: Science and Technology*, 4(3):035012, jul 2023.
- Bortoli et al. (2021) Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion schrödinger bridge with applications to score-based generative modeling. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, 2021.
- Brenier (1991) Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4):375–417, 1991.
- Bunne et al. (2022) Charlotte Bunne, Ya-Ping Hsieh, Marco Cuturi, and Andreas Krause. The Schr\"odinger Bridge between Gaussian Measures has a Closed Form. *arXiv:2202.05722*, February 2022.
- Chen et al. (2022a) Hongrui Chen, Holden Lee, and Jianfeng Lu. Improved analysis of score-based generative modeling: User-friendly bounds under minimal smoothness assumptions. *ICML*, 2022a.
- Chen et al. (2018) Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.
- Chen and Gopinath (2000) Scott Chen and Ramesh Gopinath. Gaussianization. In T. Leen, T. Dietterich, and V. Tresp, editors, *Advances in Neural Information Processing Systems*, volume 13. MIT Press, 2000.
- Chen et al. (2023) Sitan Chen, Sinho Chewi, Jerry Li, Yuanzhi Li, Adil Salim, and Anru Zhang. Sampling is as easy as learning the score: theory for diffusion models with minimal data assumptions. In *The Eleventh International Conference on Learning Representations*, 2023.
- Chen et al. (2022b) Tianrong Chen, Guan-Horng Liu, and Evangelos Theodorou. Likelihood training of schrödinger bridge using forward-backward SDEs theory. In *International Conference on Learning Representations*, 2022b.
- Chen et al. (2021) Yongxin Chen, Tryphon T. Georgiou, and Michele Pavon. Stochastic control liaisons: Richard sinkhorn meets gaspard monge on a schrödinger bridge. *SIAM Review*, 63(2):249–313, 2021.
- Dinh et al. (2017) Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real NVP. In *International Conference on Learning Representations*, 2017.
- Dockhorn et al. (2022) Tim Dockhorn, Arash Vahdat, and Karsten Kreis. Score-based generative modeling with critically-damped langevin diffusion. In *International Conference on Learning Representations (ICLR)*, 2022.

- Doob (1984) Joseph L. Doob. *Classical potential theory and its probabilistic counterpart*, volume 262 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, 1984. ISBN 0-387-90881-1.
- Durkan et al. (2019) Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Neural spline flows. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.
- El Alaoui and Montanari (2022) Ahmed El Alaoui and Andrea Montanari. An information-theoretic view of stochastic localization. *IEEE Trans. Inf. Theory*, 68(11):7423–7426, 2022.
- Eldan (2013) Ronen Eldan. Thin shell implies spectral gap up to polylog via a stochastic localization scheme. *Geometric and Functional Analysis*, 23(2):532–569, 2013.
- Finlay et al. (2020) Chris Finlay, Joern-Henrik Jacobsen, Levon Nurbekyan, and Adam Oberman. How to train your neural ODE: the world of Jacobian and kinetic regularization. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 3154–3164. PMLR, 13–18 Jul 2020.
- Friedman (1987) Jerome H. Friedman. Exploratory projection pursuit. *Journal of the American Statistical Association*, 82(397):249–266, 1987.
- Grathwohl et al. (2019) Will Grathwohl, Ricky T. Q. Chen, Jesse Bettencourt, and David Duvenaud. Scalable reversible generative models with free-form continuous dynamics. In *International Conference on Learning Representations*, 2019.
- Ho et al. (2020) Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851. Curran Associates, Inc., 2020.
- Hoogeboom et al. (2023) Emiel Hoogeboom, Jonathan Heek, and Tim Salimans. simple diffusion: End-to-end diffusion for high resolution images. In *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, 2023.
- Huang et al. (2018) Chin-Wei Huang, David Krueger, Alexandre Lacoste, and Aaron Courville. Neural autoregressive flows. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 2078–2087. PMLR, 10–15 Jul 2018.
- Huang et al. (2021) Chin-Wei Huang, Ricky T. Q. Chen, Christos Tsirigotis, and Aaron Courville. Convex potential flows: Universal probability distributions with optimal transport and convex optimization. In *International Conference on Learning Representations*, 2021.
- Hyvärinen (2005) Aapo Hyvärinen. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(24):695–709, 2005.
- Hyvärinen (1999) Aapo Hyvärinen. Sparse code shrinkage: Denoising of nongaussian data by maximum likelihood estimation. *Neural Computation*, 11(7):1739–1768, 1999.
- Kadkhodaie and Simoncelli (2021) Zahra Kadkhodaie and Eero P. Simoncelli. Solving linear inverse problems using the prior implicit in a denoiser. *arXiv:2007.13640*, 2021.
- Karras et al. (2022) Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In *Proc. NeurIPS*, 2022.
- Kim and Milman (2010) Young-Heon Kim and Emanuel Milman. A generalization of caffarelli's contraction theorem via (reverse) heat flow. *Mathematische Annalen*, 354:827–862, 2010.
- Kingma et al. (2021) Diederik P Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. On density estimation with diffusion models. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, 2021.
- LeCun et al. (2007) Yann LeCun, Sumit Chopra, and Raia Hadsell. A tutorial on energy-based learning. In Gökhan BakIr, Thomas Hofmann, Alexander J Smola, Bernhard Schölkopf, and Ben Taskar, editors, *Predicting structured data*, chapter 10. MIT press, 2007.

- Lee et al. (2022) Holden Lee, Jianfeng Lu, and Yixin Tan. Convergence for score-based generative modeling with polynomial complexity. *arXiv:2206.06227*, 2022.
- Lee et al. (2023) Holden Lee, Jianfeng Lu, and Yixin Tan. Convergence of score-based generative modeling for general data distributions. In Shipra Agrawal and Francesco Orabona, editors, *Proceedings of The 34th International Conference on Algorithmic Learning Theory*, volume 201 of *Proceedings of Machine Learning Research*, pages 946–985, 2023.
- Léonard (2014) Christian Léonard. A survey of the schrödinger problem and some of its connections with optimal transport. *Discrete and Continuous Dynamical Systems*, 34(4):1533–1574, 2014.
- Lipman et al. (2023) Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023.
- Liu et al. (2023a) Guan-Horng Liu, Arash Vahdat, De-An Huang, Evangelos A Theodorou, Weili Nie, and Anima Anandkumar. I^2sb: Image-to-image schrödinger bridge. *arXiv preprint arXiv:2302.05872*, 2023a.
- Liu (2022) Qiang Liu. Rectified flow: A marginal preserving approach to optimal transport. *arXiv:2209.14577*, 2022.
- Liu et al. (2022) Xingchao Liu, Lemeng Wu, Mao Ye, and Qiang Liu. Let us build bridges: Understanding and extending diffusion generative models. *arXiv:2208.14699*, 2022.
- Liu et al. (2023b) Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh International Conference on Learning Representations*, 2023b.
- Lu et al. (2022) Cheng Lu, Kaiwen Zheng, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Maximum likelihood training for score-based diffusion ODEs by high order denoising score matching. In *Proceedings of the 39th International Conference on Machine Learning*, 2022.
- Maoutsa et al. (2020) Dimitra Maoutsa, Sebastian Reich, and Manfred Opper. Interacting particle solutions of fokker–planck equations through gradient–log–density estimation. *Entropy*, 22(8), 2020.
- Montanari (2023) Andrea Montanari. Sampling, diffusions, and stochastic localization. *arXiv:2305.10690*, 2023.
- Nair and Hinton (2010) Vinod Nair and Geoffrey E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, page 807–814, Madison, WI, USA, 2010. Omnipress. ISBN 9781605589077.
- Nilsback and Zisserman (2006) Maria-Elena Nilsback and Andrew Zisserman. A visual vocabulary for flower classification. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 1447–1454, 2006.
- Onken et al. (2021) Derek Onken, Samy Wu Fung, Xingjian Li, and Lars Ruthotto. Ot-flow: Fast and accurate continuous normalizing flows via optimal transport. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(10):9223–9232, May 2021.
- Otto (2000) Felix Otto. Generalization of an inequality by talagrand and links with the logarithmic sobolev inequality. *Journal of Functional Analysis*, 173(2):361–400, 2000.
- Papamakarios et al. (2017) George Papamakarios, Theo Pavlakou, and Iain Murray. Masked autoregressive flow for density estimation. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 2335–2344, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.
- Peluchetti (2022) Stefano Peluchetti. Non-denoising forward-time diffusions. *https://openreview.net/forum?id=oVfIKuhqfC*, 2022.
- Rezende and Mohamed (2015) Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 1530–1538, Lille, France, 07–09 Jul 2015. PMLR.
- Simoncelli and Adelson (1996) Eero P. Simoncelli and Edward H. Adelson. Noise removal via bayesian wavelet coring. In *Proceedings of 3rd IEEE International Conference on Image Processing*, volume 1, pages 379–382 vol.1, 1996.
- Sohl-Dickstein et al. (2015) Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 2256–2265, Lille, France, 07–09 Jul 2015. PMLR.

- Somnath et al. (2023) Vignesh Ram Somnath, Matteo Pariset, Ya-Ping Hsieh, Maria Rodriguez Martinez, Andreas Krause, and Charlotte Bunne. Aligned diffusion schrödinger bridges. In *The 39th Conference on Uncertainty in Artificial Intelligence*, 2023.
- Song and Ermon (2019) Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.
- Song and Kingma (2021) Yang Song and Diederik P. Kingma. How to train your energy-based models. *arXiv:2101.03288*, 2021.
- Song et al. (2021a) Yang Song, Conor Durkan, Iain Murray, and Stefano Ermon. Maximum likelihood training of score-based diffusion models. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 1415–1428. Curran Associates, Inc., 2021a.
- Song et al. (2021b) Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021b.
- Song et al. (2023) Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency Models. *ICML*, 2023.
- Stein (1981) Charles M. Stein. Estimation of the Mean of a Multivariate Normal Distribution. *The Annals of Statistics*, 9(6):1135 – 1151, 1981.
- Su et al. (2023) Xuan Su, Jiaming Song, Chenlin Meng, and Stefano Ermon. Dual diffusion implicit bridges for image-to-image translation. In *The Eleventh International Conference on Learning Representations*, 2023.
- Tabak and Turner (2013) Esteban G. Tabak and Cristina V. Turner. A family of nonparametric density estimation algorithms. *Communications on Pure and Applied Mathematics*, 66(2):145–164, 2013.
- Tabak and Vanden-Eijnden (2010) Esteban G. Tabak and Eric Vanden-Eijnden. Density estimation by dual ascent of the log-likelihood. *Communications in Mathematical Sciences*, 8(1):217 – 233, 2010.
- Tong et al. (2020) Alexander Tong, Jessie Huang, Guy Wolf, David Van Dijk, and Smita Krishnaswamy. TrajectoryNet: A dynamic optimal transport network for modeling cellular dynamics. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 9526–9536. PMLR, 13–18 Jul 2020.
- Tong et al. (2023) Alexander Tong, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Kilian Fatras, Guy Wolf, and Yoshua Bengio. Conditional flow matching: Simulation-free dynamic optimal transport. *arXiv:2302.00482*, 2023.
- Vempala and Wibisono (2019) Santosh S. Vempala and Andre Wibisono. Rapid convergence of the unadjusted langevin algorithm: Log-sobolev suffices. *arXiv:1903.08568*, 2019.
- Villani (2009) Cédric Villani. *Optimal transport: old and new*, volume 338. Springer, 2009.
- Vincent (2011) Pascal Vincent. A Connection Between Score Matching and Denoising Autoencoders. *Neural Computation*, 23(7):1661–1674, 2011.
- Xiao et al. (2022) Zhisheng Xiao, Karsten Kreis, and Arash Vahdat. Tackling the generative learning trilemma with denoising diffusion GANs. In *International Conference on Learning Representations*, 2022.
- Yang and Karniadakis (2022) Liu Yang and George Em Karniadakis. Potential flow generator with l2 optimal transport regularity for generative models. *IEEE Transactions on Neural Networks and Learning Systems*, 33(2):528–538, 2022.
- Zhang et al. (2018) Linfeng Zhang, Weinan E, and Lei Wang. Monge-Ampère flow for generative modeling. *arXiv.org:1809.10188*, 2018.