

压缩扩散模型

Jyotirmai Singh

Stanford University

joesingh@stanford.edu

Samar Khanna

Stanford University

samarkhanna@cs.stanford.edu

James Burgess

Stanford University

jmhb@stanford.edu

Abstract

扩散模型通常注入各向同性高斯噪声，忽视数据中的结构。受海森堡不确定性原理下量子挤压态重新分配不确定性的启发，我们引入了**挤压扩散模型 (SDM)**，其沿训练分布的主要成分以各向异性方式缩放噪声。挤压在物理学中增强了信噪比，因此我们认为以数据依赖的方式缩放噪声可以更好地帮助扩散模型学习重要的数据特征。我们研究了两种配置：(i) 一种海森伯格扩散模型，它通过在正交方向上进行逆缩放来补偿主轴上的缩放，并且(ii) 一种仅缩放主轴的标准 SDM 变体。出人意料的是，在 CIFAR-10/100 和 CelebA-64 上，温和的抗挤压-即增加主轴上的方差-一致地将 FID 提高了高达 15%，并将精确度-召回率前沿移向更高的召回率。我们的结果表明，简单的、数据感知的噪声整形可以在不改变架构的情况下提供稳健的生成收益。

1 介绍

基于分数的扩散模型已成为高质量图像生成的标准工具，在从 CIFAR-10 到 ImageNet[1, 2] 的数据集上达到了最先进的性能。虽然扩散模型通常依赖于在训练过程中添加标准各向同性高斯噪声，但有证据表明这种“一刀切”的方法可能会在某些情况下（如自然图像，其在频域中具有各向异性功率分布）降低生成质量 [3]。这促使人们研究以数据相关的方式对扩散过程中的噪声进行工程化处理的方法，以克服这些缺点 [4, 5]。

量子计量学领域 [6] 已经发展出了一系列技术，用于在精密测量的背景下操纵噪声。其中一种技术是量子压缩 [7]，它重新分配了噪声，使得沿着“测量”轴的方差减小，从而提高信噪比 (SNR) 并实现弱信号提取。受此启发，我们引入了挤压扩散模型 (SDM)，该模型沿数据集的主要成分各向异性地缩放噪声。我们认为这鼓励了模型在训练过程中学习更具语义意义的模式，从而提高了生成样本的质量。

为了测试这一点，我们研究了两种模型变体：(1) 一个海森伯格扩散模型，它放大主方向并缩小正交方向；(2) 只放大主方向的“标准”SDM。令人惊讶的是，我们在多个数据集上观察到沿主轴轻微抗挤压的最佳性能——即注入噪声和降低信噪比。

背景 量子挤压利用海森堡不确定性原理：

$$\Delta A \cdot \Delta B \geq \frac{1}{2} |\langle [A, B] \rangle| \quad (1)$$

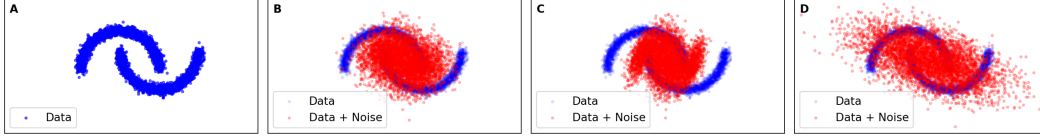


图 1: 扩散与挤压噪声。(a) 原始数据分布。(b) 标准各向同性加噪过程。(c) – (d) 沿主方向进行 (c) 挤压和 (d) 反挤压的各向异性加噪过程。

通过减少一个物理变量 ΔA 的方差同时增加其共轭变量 ΔB 的方差，但保持乘积不变。挤压状态提高了沿选定轴的信噪比，并支持诸如引力波 [8] 和暗物质 [9] 检测器等精密测量。

Figure 1 说明了这一效果如何应用于扩散模型。在 fig. 1(c) 中，噪声（红色）沿第一主成分被压缩，由于相对较高的信噪比，原始分布（蓝色）仍然保持相当完整，相对于各向同性噪声 (fig. 1(b))。在 fig. 1(d) 中，噪声反而变为反挤压或沿相同方向增强，与各向同性情况相比降低了信噪比。

相关工作 近期的研究表明，替代的噪声调度可以影响生成性能 [10]。例如使用余弦调度 [2] 和变分优化调度 [11] 已经证明了改进。其他工作则专注于通过例如等化或偏置某些频率的方式改变频域中的噪声结构 [3, 4, 12]。更接近我们的目标，其他研究已经表明学习一个多变量数据相关的噪声过程 [5] 可以带来好处。我们的方法不同之处在于使用一种简单的、受物理学启发的 PCA 对齐来应用单参数各向异性缩放。

2 方法: 压缩扩散

如标准的 DDPM 一样，SDM 会在一个时间表 $\{\beta_t\}_{t=1}^T \subset (0, 1)$ 上通过加性高斯噪声来破坏一个干净的数据点 x_0 。然而，在 SDM 中，每一步我们都会用一个强度由单一超参数控制的挤压矩阵 $S_t(s)$ 对噪声进行各向异性地缩放，该超参数可能是 s ，也可能随 t 变化：

$$x_t^{\text{sq}} = \sqrt{\alpha_t} x_{t-1}^{\text{sq}} + \sqrt{1 - \alpha_t} S_t(s) \varepsilon_t, \quad \varepsilon_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, I), \quad \alpha_t = 1 - \beta_t. \quad (2)$$

边缘分布 通过展开 eq. (2)，我们得到

$$x_t^{\text{sq}} = \sqrt{\bar{\alpha}_t} x_0 + \sum_{i=1}^t \left(\sqrt{1 - \alpha_i} \prod_{j=i+1}^t \sqrt{\alpha_j} \right) S_i(s) \varepsilon_i, \quad \bar{\alpha}_t = \prod_{i=1}^t \alpha_i, \quad (3)$$

它仍然是具有均值 $\sqrt{\bar{\alpha}_t} x_0$ 和协方差 $\Sigma_t = \sum_{i=1}^t ((1 - \alpha_i) \prod_{j=i+1}^t \alpha_j) S_i S_i^\top$ 的高斯分布。

训练目标 网络被训练来预测压缩噪声 $\varepsilon_t^{\text{sq}} = S_t(s) \varepsilon_t$ 通过标准均方误差：

$$\mathcal{L}_{\text{SDM}} = \mathbb{E}_{x_0, t, \varepsilon_t} \left\| \varepsilon_t^{\text{sq}} - \hat{\varepsilon}_\theta^{\text{sq}}(x_t^{\text{sq}}, t) \right\|_2^2. \quad (4)$$

反向步骤（白化-去噪-再挤压） 对于反向步骤，我们首先漂白每个状态：

$$\tilde{x}_t = S_t^{-1} x_t^{\text{sq}}, \quad \tilde{\varepsilon}_\theta = S_t^{-1} \varepsilon_\theta^{\text{sq}}. \quad (5)$$

通过这个变量变换，eq. (2) 变为大约等于一个标准的 DDPM 在 \tilde{x}_t 中，这使我们能够重用后验概率 Ho et al. [1]：

$$\tilde{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\tilde{x}_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \tilde{\varepsilon}_\theta(\tilde{x}_t, t) \right) + \sqrt{\tilde{\beta}_t} z, \quad z \sim \mathcal{N}(0, I), \quad (6)$$

其中 $\tilde{\beta}_t = (1 - \bar{\alpha}_{t-1}) / (1 - \bar{\alpha}_t) \beta_t$ 。最后我们使用重新挤压和矩阵下一个将 S_{t-1} 保持在压缩坐标中：

$$x_{t-1}^{\text{sq}} = S_{t-1}(s) \tilde{x}_{t-1}. \quad (7)$$

请注意，通过白化处理，eq. (2) 变为

$$\tilde{x}_t = \sqrt{\alpha_t} S_t^{-1} S_{t-1} \tilde{x}_{t-1} + \sqrt{1 - \alpha_t} \varepsilon_t \quad (8)$$

当 S_t 不随时间变化时，方程 eq. (8) 是一个精确的 DDPM，因为 $S_t = S_{t-1}$ 和 eq. (6) 是精确的后验分布。对于时间依赖的挤压，这只是近似的 DDPM，并且具有一些漂移 $R_t = S_t^{-1} S_{t-1}$ 。然而，对于我们定义的挤压矩阵形式和线性噪声调度，这种漂移最终变得可以忽略不计，因此后验估计不会影响生成性能。更多细节请参见 section B。

压缩矩阵 $S_t(s)$ 为了选择挤压的方向，我们在单个像素级别对 RGB 空间中的数据进行 PCA 分解。对于大型自然图像集，这与对立色轴亮度、红-绿和蓝-黄对齐 [13, 14]。令 \hat{v} 为第一主成分方向上的单位向量，在 RGB 图像中即为亮度。我们可以定义海森伯扩散模型的压缩矩阵 S_t^{HDM} 和标准压缩扩散模型的压缩矩阵 S_t^{SDM} ：

$$S_t^{\text{HDM}}(s) = e^{-s} \hat{v} \hat{v}^\top + e^{\frac{s}{n-1}} (I - \hat{v} \hat{v}^\top), \quad S_t^{\text{SDM}}(s) = I + \hat{v} \hat{v}^\top (e^{-s} - 1). \quad (9)$$

这里 s 是对应于压缩强度的超参数，而 n 是进行压缩操作的向量空间的维度。对于 RGB 像素的情况， $n = 3$ 。注意， S_t^{SDM} 是通过仅仅丢弃在与 $\hat{v} \hat{v}^\top$ 正交的子空间中的反挤压从 S_t^{HDM} 导出的。这些矩阵反映了量子力学中电磁挤压算子的指数形式 [15]：

$$S(z) = \exp \left(\frac{z^* \hat{a}^2 + z (\hat{a}^\dagger)^2}{2} \right) \quad (10)$$

对于 $z = r e^{i\theta}$ ， $S(z)$ 在由电磁创造和湮灭算子 \hat{a}^\dagger, \hat{a} 决定的相空间中的 θ 方向将噪声缩放为 e^{-r} 倍，正如 S_t 在由 \hat{v} 确定的方向上通过 e^{-s} 进行的那样。

最后，为了使挤压过程更好地与扩散模型的噪声时间表对齐，在我们的实验中该时间表是线性的，我们将挤压强度 s 作为时间 $s(t) = s_0 \frac{\beta_t}{\beta_{\max}}$ 的线性函数变化，这使得挤压矩阵 S_t 成为时间依赖的。

3 结果

我们的实验包括消融研究，以评估在各种标准图像数据集上的压缩扩散。结果如 fig. 2 所示。在 CIFAR-10 数据集 [16] 上，FID-10K 显示出与压缩强度 (fig. 2a) 一致的 U 型依赖关系，跨多个随机种子。基线 DDPM 的平均 FID 为 39.8，而 SDM 达到了 $s \approx -0.4$ 的 FID 33.6 - 改善了 ~15%。值得注意的是，在抗挤压条件下信噪比下降。

接下来，使用固定的种子，我们将标准压缩扩散模型与保持不确定性的海森堡变体在 CIFAR-10 上进行比较。FID-10K (fig. 2b) 和 Inception Score (fig. 2c) 在 s 中都遵循相同的 U 形趋势，最优值分别接近 $s \approx -0.4$ (FID) 和 $s \approx -0.3$ (IS)。SDM 在轻微的反压缩区域优于 HDM。

为了检查压缩的鲁棒性，我们在 CIFAR-100 [16] 上评估了 SDM 生成性能 fig. 2d，在 CelebA-64 [17] 上评估了 fig. 2e。抛物线形状在数据集之间得到了重现，最佳效果再次出现在温和的反压缩区域。最优位置的确切位置因数据集而异，CIFAR-100 的最佳性能位于 $s \sim -0.3$ ，CelebA-64 的最佳性能位于 $s \sim -0.2$ 。

为了进一步了解压缩与反压缩的影响，我们分析了模型在 CIFAR-10 [18–20] 上的精度和召回率指标。Figure 2f 表明压缩 SDM 噪声对应于减少召回率，而精确度略有提升。相比之下，反压缩噪声会产生较大的召回率增益，同时一定程度上以轻微的精确度损失为代价。总体生成质量，通过 F-score 衡量，再次表现出抛物线行为，直到某个点为止。

定性样本 在图中。3 我们比较了用相同噪声种子生成的基线 DDPM 和 SDM 的 36 个样本在最优反挤压情况下的表现 $s = -0.4$ 。SDM 图像显示物体轮廓更加清晰：第 1 行第 3 列变

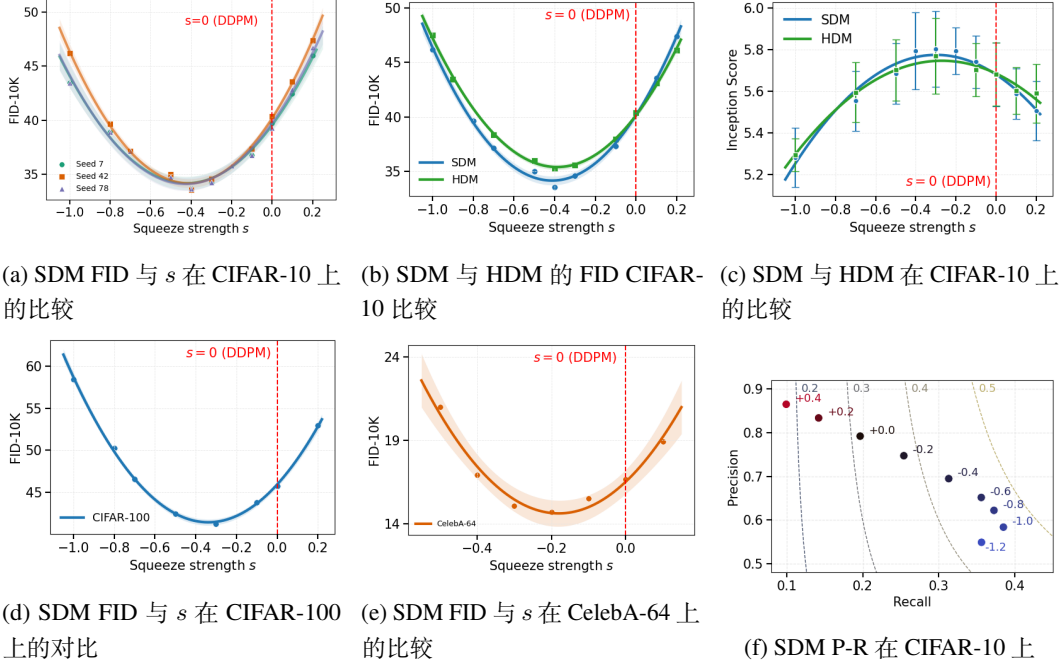


图 2: 结果概述跨数据集和指标。(a) – (c) SDM 与 HDM 在 CIFAR-10 上的 FID 和 IS 比较; (d) – (e) CIFAR-100 和 CelebA-64 上的跨数据集 FID 消融研究; (f) 在 CIFAR-10 上的精确率 – 召回率分析, 点和 F 分数恒定的轮廓上标注了挤压强度。



图 3: 生成的 CIFAR-10 样本用于标准 DDPM 和具有固定种子的 SDM。(a) 标准 DDPM; (b) 在 $s = -0.4$ 处的反挤压 SDM。

成了一辆清晰的小车而不是灰色模糊, 而第 2 行第 6 列则显示出更锐利的船边缘和云朵。这些更锐利的纹理与在 \sim 中看到的 15% FID 下降 $s = -0.4$ 在 2a 中一致。偶尔仍然存在高频振铃伪影 (例如, 第 2 行第 3 列的小车), 这反映了在 PR 分析中随着反挤压增加所看到的精度损失。

4 结论与未来工作

在三个图像数据集中，我们观察到样本质量对挤压强度的稳定 U 形依赖关系：适度的反挤压会增大亮度方向上的方差，提高召回率而几乎不牺牲精度，而强烈的挤压或反挤压都会损害两者。标准 SDM 变体与海森堡对应物相匹配或优于后者，这表明扩散模型不必严格保持不确定性就能提升生成性能。未来的工作可以包括扩展到更高分辨率的数据集，并探索频率依赖性 or 特定模态的挤压，例如音频。总体而言，SDM 提供了一种操纵噪声的原则方法，能够显著提高生成质量。

参考文献

- [1] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, 2020.
- [2] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. In *Advances in Neural Information Processing Systems*, 2021.
- [3] Fabian Falck, Teodora Pandevara, Kiarash Zahirnia, Rachel Lawrence, Richard Turner, Edward Meeds, Javier Zazo, and Sushrut Karmalkar. A fourier space perspective on diffusion models. *arXiv preprint arXiv:2505.11278*, 2025.
- [4] Xingchang Huang, Corentin Salaun, Cristina Vasconcelos, Christian Theobalt, Cengiz Oztireli, and Gurprit Singh. Blue noise for diffusion models. In *SIGGRAPH: Computer Graphics and Interactive Techniques*, 2024.
- [5] Subham Sekhar Sahoo, Aaron Gokaslan, Chris De Sa, and Volodymyr Kuleshov. Diffusion models with learned adaptive noise. In *Advances in Neural Information Processing Systems*, 2024.
- [6] A. A. Clerk, M. H. Devoret, S. M. Girvin, F. Marquardt, and R. J. Schoelkopf. Introduction to quantum noise, measurement and amplification. *Review of Modern Physics*, 2008.
- [7] D. F. Walls. Squeezed states of light. *Nature*, 1983.
- [8] D. Ganapathy et al. Broadband quantum enhancement of the ligo detectors with frequency-dependent squeezing. *Phys. Rev. X*, 13:041021, Oct 2023. doi: 10.1103/PhysRevX.13.041021. URL <https://link.aps.org/doi/10.1103/PhysRevX.13.041021>.
- [9] K.M. Backes et al. A quantum enhanced search for dark matter axions. *Nature* 590, 238 – 242, 2021.
- [10] Ting Chen. On the importance of noise scheduling for diffusion models, 2023. URL <https://arxiv.org/abs/2301.10972>.
- [11] Diederik P. Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. In *Proceedings of the 35th International Conference on Neural Information Processing Systems*, 2021.
- [12] Anh Tran Hao Phung, Quan Dao. Wavelet diffusion models are fast and scalable image generators. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

- [13] Gershon Buchsbaum and Allan Gottschalk. Trichromacy, opponent colours coding and optimum colour information transmission in the retina. *Proceedings of the Royal society of London. Series B. Biological sciences*, 220(1218):89–113, 1983.
- [14] Daniel L. Ruderman, Thomas W. Cronin, and Chuan-Chin Chiao. Statistics of cone responses to natural images: implications for visual coding. *J. Opt. Soc. Am. A*, 15(8):2036–2045, Aug 1998. doi: 10.1364/JOSAA.15.002036. URL <https://opg.optica.org/josaa/abstract.cfm?URI=josaa-15-8-2036>.
- [15] Christopher Gerry and Peter L. Knight. *Introductory Quantum Optics*. Cambridge University Press, 2005.
- [16] Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009.
- [17] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 3730–3738, 2015.
- [18] Mehdi S. M. Sajjadi, Olivier Bachem, Mario Lucic, Olivier Bousquet, and Sylvain Gelly. Assessing generative models via precision and recall. In *International Conference on Learning Representations (ICLR)*, 2018.
- [19] Teemu Kynkäänniemi, Tero Karras, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Improved precision and recall metric for assessing generative models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.
- [20] Anton Obukhov, Maximilian Seitzer, Po-Wei Wu, Semen Zhydenko, Jonathan Kyl, and Elvis Yu-Jing Lin. High-fidelity performance metrics for generative models in pytorch, 2020. URL <https://github.com/toshas/torch-fidelity>. Version: 0.3.0, DOI: 10.5281/zenodo.4957738.
- [21] Patrick von Platen, Suraj Patil, Anton Lozhkov, Pedro Cuenca, Nathan Lambert, Kashif Rasul, Mishig Davaadorj, Dhruv Nair, Sayak Paul, Steven Liu, William Berman, Yiyi Xu, and Thomas Wolf. Diffusers: State-of-the-art diffusion models. URL <https://github.com/huggingface/diffusers>.

A 模型训练的细节

A.1 实现

我们使用带有 UNet2D 模型主干的 HuggingFace 扩散模型库 [21]。为了应用挤压操作，我们保持 UNet 架构不变，并子类化 `DDPMScheduler` 类，覆盖添加噪声和步骤方法。所有其他训练和采样代码保持标准。该模型使用线性噪声时间表。训练使用了 1000 个时间步长，而采样则使用 50 个。我们还使用了 EMA，并将 EMA 衰减值设置为 0.9999。所有实验均在一个 NVIDIA A100 (40GB) GPU 上通过加速以混合精度 (fp16) 运行。这些实验的代码可在 <https://github.com/joe-singh/squeezing> 处获得。

B 正向和反向过程的细节

B.1 前向过程

证明. 我们通过归纳法证明 eq. (3)。该命题对于 x_1 是真的，由 eq. (2) 给出。现在假设它对 x_t 为真。那么：

$$\begin{aligned}
 x_{t+1} &= \sqrt{\alpha_{t+1}}x_t + \sqrt{1 - \alpha_{t+1}}S_{t+1}\varepsilon_{t+1} \\
 &= \sqrt{\alpha_{t+1}} \left[\sqrt{\alpha_t \dots \alpha_1}x_0 + \sum_{i=1}^t S_i \varepsilon_i \sqrt{1 - \alpha_i} \prod_{j=i+1}^t \sqrt{\alpha_j} \right] + \sqrt{1 - \alpha_{t+1}}S_{t+1}\varepsilon_{t+1} \\
 &= \sqrt{\alpha_{t+1}\alpha_t \dots \alpha_1}x_0 + \sum_{i=1}^t S_i \varepsilon_i \sqrt{1 - \alpha_i} \prod_{j=i+1}^{t+1} \sqrt{\alpha_j} + \sqrt{1 - \alpha_{t+1}}S_{t+1}\varepsilon_{t+1} \\
 &= \sqrt{\alpha_{t+1}}x_0 + \sum_{i=1}^{t+1} S_i \varepsilon_i \sqrt{1 - \alpha_i} \prod_{j=i+1}^{t+1} \sqrt{\alpha_j}
 \end{aligned}$$

□

B.2 后验近似的有效性

在文中我们注意到，对于时间依赖的压缩，Ho 等人。后验是一种近似。为了说明这一点，我们计算标准 SDM 情况下漂移因子 $S_t^{-1}S_{t-1}$ 。首先使用 Sherman-Morrison 公式计算 S_t^{SDM} 的逆：

$$S_t^{-1} = I + vv^\top \left(e^{s(t)} - 1 \right) \quad (11)$$

在我们的工作中， $s(t)$ 遵循由 $s(t) = s_0 \frac{\beta_t}{\beta_{\max}}$ 给出的线性计划。使用此线性计划计算 $S_t^{-1}S_{t-1}$ 得到

$$S_t^{-1}S_{t-1} = I + vv^\top \left(\exp \left[\frac{s_0 \Delta \beta}{\beta_{\max}} \right] - 1 \right) \quad (12)$$

在哪里 $\Delta \beta = \beta_t - \beta_{t-1}$ 。默认线性计划扩散器从 $\beta = 1 \times 10^{-4}$ 开始并在 $\beta = 0.02$ 结束。使用 1000 个时间步（训练）时， $\Delta \beta \sim 2 \times 10^{-5}$ 而在 50 个时间步（推理）时， $\Delta \beta \sim 4 \times 10^{-4}$ 。两个数字都足够小，因此对于较小的挤压强度 $|s_0|$ ，指数接近于 1，所以漂移项大约是恒等变换，并且方程 eq. (6) 中的后验近似成立。