arXiv:2302.01217v1 [stat.ML] 2 Feb 2023

使用去噪扩散概率模型进行图像修复的理论论证

Litu Rout*、Advait Parulekar†、Constantine Caramanis‡ 和 Sanjay Shakkott ai§

德克萨斯大学奥斯汀分校

摘要

我们为在线性模型设定下使用基于扩散的图像修复进行样本恢复提供了理论论证。尽管大多数图像修复算法都需要针对每个新掩码重新训练，但我们证明，基于扩散的图像修复无需重新训练即可很好地泛化到未见过的掩码。我们分析了一种近期提出且广受欢迎的基于扩散的图像修复算法 RePaint [LDRYTV22], ，并表明由于错位，即使在双状态扩散过程中，该算法也存在妨碍样本恢复的偏差。受分析结果启发，我们提出了一种改进的 RePaint 算法，将其称为 RePaint$^+$ 。我们证明，该算法能够恢复底层真实样本，并具有线性收敛率。它通过校正反向过程中漂移和弥散中的错位误差实现这一点。据我们所知，这是基于扩散的图像修复算法首次获得线性收敛结果。

1 引言

我们研究使用去噪扩散概率模型（DDPM）进行图像修复取得实证成功背后的数学原理 [SWMG15; HJA20; SSKKEP20], 这些模型是大型生成模型的基础，包括 DALL-E [RPGGVRCS21; RDNCC 22], Imagen [SC-SLWDGAML+22]和 Stable Diffusion [RBLEO22]。图像修复旨在重建缺失部分，使其在语义上与图像中已知的部分一致。这项任务的一大挑战是生成与图像现有部分相符的内容；生成建模是实现这一修复目标的关键工具。

近年来，变分自编码器 (VAEs) [KW13]和生成对抗网络 (GANs) [GP- MXWOCB14] 一直是许多成功图像修复技术的基础。然而，随着基于分数的生成模型 (SGMs) [SE19] 或去噪扩散概率模型 (DDPMs) [SWMG15; HJA20],的出现，研究重点已转向另一种图像修复范式。在这一范式中，大多数方法可分为两类。第一类方法学习针对特定下游任务（如图像修复或超分辨率）的扩散过程 [WDTSDM22; SCCLHSFN22]。第二类方法则学习通用的基于扩散的生成模型，并在推理阶段引导扩散过程以适应下游任务 [JADPDT21; CKJGY21; SSXE21; DDDD22; KEES22]。本文分析了一种 [LDRYTV22] 属于第二类的图像修复方法。我们证明，DDPM 的生成先验足以完整重建图像的缺失部分。

*litu.rout@utexas.edu
†advaitp@utexas.edu
‡constantine@utexas.edu§sanjay.shakkottai@utexas.edu

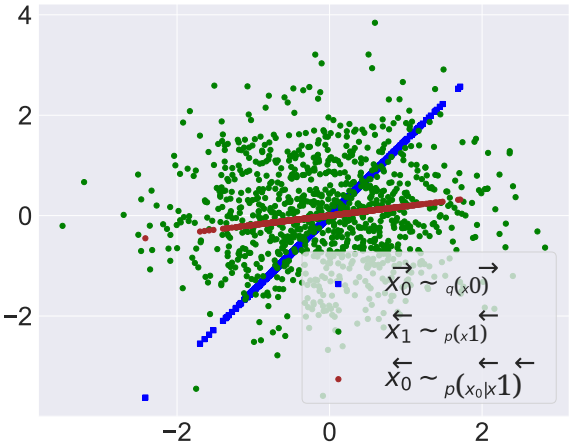


图1：展示 RePaint 偏差的示例 [LDRYTV22]。从高斯先验$\overleftarrow{x_1}$ （绿色圆圈）出发，RePaint 提出的反向 SDE 会生成 $\overleftarrow{x_0}$ （棕色圆圈），该样本在已知坐标上与真实数据 $\overrightarrow{x_0}$ （蓝色方块）相符，但在图像修复区域有所不同。图中，沿着 $\sim 56°$ 线排列的蓝色方块表示真实样本，而沿着$\sim 10°$ 线排列的棕色圆圈表示 RePaint 恢复的样本。请注意，恢复的样本在 x 坐标（已知数据）上与真实样本相符，但在 y 坐标（缺失数据）上存在偏差。

DDPM 是一类生成模型，它们学习将干净图像扩散为易处理噪声，然后通过反向马尔可夫过程逐步对纯噪声进行去噪，从而生成干净图像 [SWMG15; HJA20; SSKKEP20] (见 §2)。凭借强大的表达能力，这些模型在条件/无条件图像生成 [HJA20; SSKKEP20; DN21; KAAL22], 文本到图像合成 [RBLEO22], 时间序列建模 [TSSE21], 音频合成 [KPHZC21], 图像到图像转换 [SCCLHSFN22], 可控文本生成 [LTGLH22],和图像复原 [KEES22]方面均取得了令人瞩目的成果。对于图像修复等复原任务，DDPM 很有吸引力，因为它们能够轻松适应未知任务（例如新的掩码），而无需经历整个重新训练过程。因此，DDPM 相较于基于生成对抗网络 [WYWGLDQC18; YLYSLH18]和变分自编码器 [RVV19; PLXL21]的先前复原技术更具优势。

然而，限制其广泛应用的一个难点在于反向马尔可夫过程中的超参数选择。尽管正向扩散过程可以预先显式计算，但反向过程的计算开销极大，因为它要求在每个中间状态进行采样。此外，常见的基于扩散的图像修复算法（如 RePaint） [LDRYTV22] 还要求在每个状态额外进行重采样，以协调图像修复部分与图像其余部分。这些重采样步骤增加了计算负担，却牺牲了语义上有意义的重建。此外，图1 显示，RePaint 生成的有偏样本妨碍了完美恢复。因此，解决这些瓶颈以促进其成功实际部署，正变得愈发重要。在这方面，一个长期存在的挑战是，尽管基于扩散的图像修复取得了显著进展，我们对其理论的理解仍处于早期阶段。

为了解决这些问题，一项新兴的理论研究致力于为构成去噪扩散概率模型（DDPMs）的基于分数的生成模型提供收敛保证 [CCLLSZ22；LLT22a；LLT22b；CLL22]。这些方法的一个关键假设是，目标分布 $q(\overrightarrow{x_0})$相对于勒贝格测度具有密度。另一类研究 [Bor22；Pid22] 则在著名的流形假设下研究得分生成模型的收敛性 [TSL00；FMN16]。从先验$p(\overleftarrow{x_T})$出发，这些研究分析反向过程 $p_{\boldsymbol{\theta}}(\overleftarrow{x_0})$ 收敛到目标分布的情况

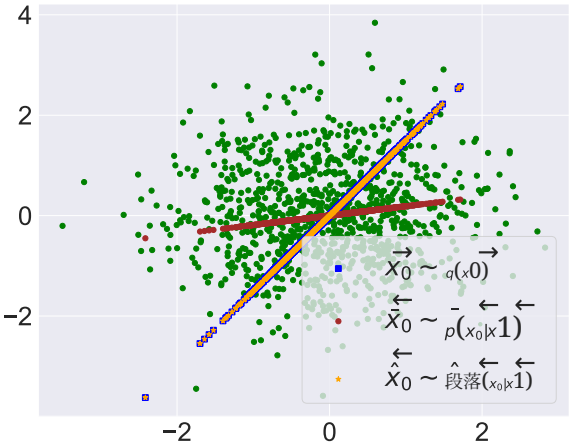


图2：RePaint ($\overleftarrow{\boldsymbol{x}_0}$) 与 RePaint$^+$ ($\overleftarrow{\hat{\ }}\boldsymbol{x}_0$) 的比较。RePaint$^+$ 能找到 ϵ-精确解，而 RePaint 则存在错位偏差。绿色圆圈表示反向马尔可夫过程的高斯先验$\overleftarrow{\boldsymbol{x}_1}$。图中，沿着 $\sim 56°$ 线排列的蓝色方块代表真实样本。与之前一样，沿着 $\sim 10°$ 线排列的棕色圆圈代表 RePaint 恢复的样本（这些样本存在偏差），沿着 $\sim 56°$ 线排列的橙色星形代表 Repaint$^+$恢复的样本。请注意，RePaint$^+$ 恢复的样本没有明显偏差（恢复的样本与真实样本重叠）。

$q(\vec{\boldsymbol{x}_0})$。与这些研究不同，我们的论文关注用于协调修复图像的重采样步骤的收敛性（§3）。

假设目标分布相对于勒贝格测度存在密度，这意味着 $q(\vec{\boldsymbol{x}_0}) > 0$ 对于所有 $\vec{\boldsymbol{x}_0} \in \mathbb{R}^d$。换言之，每个 $\vec{\boldsymbol{x}_0}$ 都是从 $q(\vec{\boldsymbol{x}_0})$中抽取的一个可能样本，这与大多数自然图像都位于具有紧支撑的低维流形上这一事实相矛盾 [TSL00; FMN16]。这一假设要求数字的目标分布为极不可能出现的样本分配非零概率质量，例如人脸、动物和卧室场景。尽管先前的分析提供了一些见解，但在这种流形情形下几乎没有实际意义 [KSSKM22]。为迈出规避这一问题的一步，Pidstrigach [Pid22] 和 Bortoli [Bor22]证明，得分生成模型学习的是由低维子结构支撑、满足流形假设的分布。

然而，对于图像修复，我们需要在数据分布上引入额外的结构，这与相关领域的先前研究类似 [BJPD17; DGE18; JLDC20; JADPDT21]。这是因为我们的目标不仅是证明 $q(\vec{\boldsymbol{x}_0})$ 与 $p_{\boldsymbol{\theta}}(\overleftarrow{\boldsymbol{x}_0})$ 之间相对于某种散度 D 的差异最小，即 $\mathrm{D}\left(p_{\boldsymbol{\theta}}(\overleftarrow{\boldsymbol{x}_0}) \; q(\vec{\boldsymbol{x}_0})\right) \leq \epsilon$，还要证明逐样本收敛，即 $\|\overleftarrow{\boldsymbol{x}_0} - \vec{\boldsymbol{x}_0}\| \leq \epsilon$。值得一提的是，样本恢复问题已在优化文献中得到了充分研究 [Tib96; FNW07; BHSN10; FN03; BJPD17]。相较之下，我们的论文旨在利用这一广为人知的设定，为使用基于DDPM的非优化技术进行样本恢复提供理论依据 [LDRYTV22]。在该技术中，我们借助基于扩散的生成模型的表达能力 [HJA20]，而不试图直接求解约束优化问题。遵循传统观点 [BJPD17; JLDC20; JADPDT21], ，我们在数据生成分布上植入一个满足流形假设的特定模型，并证明学习算法能够通过检测底层子结构来恢复该模型：见图2。

1.1 贡献

我们的主要贡献是为线性模型设定下基于扩散的图像修复提供理论依据。通过分析两个状态之间的扩散，我们的分析解释了此前尚未得到理解的现象。重要的是，我们推导出了算法层面的启示；正如我们稍后将展示的，这些启示带来的改进超越了两个状态的过程。

首先，与以往的图像修复方法不同，我们证明基于扩散的图像修复无需重新训练即可轻松适应每个新掩码（**定理 4** 见 §3）。

其次，受近期提出的一种基于扩散的图像修复方法 [LDRYTV22], 的经验成功启发，我们在双状态扩散过程系统中分析了其理论性质。我们发现，错位会造成偏差，从而妨碍完美恢复：见图1。我们的分析促使我们纠正这种错位，以消除偏差。我们将此方法称为 RePaint$^+$。我们在$^+$ 中给出了 RePaint**算法 2**的简化版本，并在**算法 3**中给出了通用版本。我们在附录B中进行了玩具实验。

在线性模型设定中，我们利用去噪扩散概率模型（DDPM）的转移核，推导出生成模型的闭式解（**定理 3**）。利用该解，我们证明 RePaint$^+$ 在重采样阶段具有线性收敛率（**定理 4**）。这使我们能够在实践中恰当地选择重采样轮次数。我们将这一收敛概念推广到生成模型为近似模型的情形（**定理 6** ）。此外，我们阐明了在图像修复场景中，相较于减慢扩散过程，重采样的优势（ §3.2.4 ）。

符号说明：我们用粗体大写字母 $\boldsymbol{A}$ 表示矩阵，用粗体小写字母 $\boldsymbol{x}$ 表示向量，用普通小写字母 x 表示标量。 $\boldsymbol{I}_k$ 表示 $k \times k$ 单位矩阵。整数集合$\{1, \dots, N\}$记为 $[N]$。逐元素乘积记为 $\odot$。算子 $\boldsymbol{D}(\boldsymbol{x})$ 将向量 $\boldsymbol{x}$对角化。$\|\boldsymbol{A}\|$ 表示谱范数，而$\|\boldsymbol{A}\|_F$表示矩阵 $\boldsymbol{A}$的弗罗贝尼乌斯范数。对于向量，$\|\boldsymbol{x}\|$ 表示其欧几里得范数。$\mathcal{U}(\cdot\cdot)$表示均匀分布。

2 扩散模型背景

与其他生成模型类似，例如生成对抗网络 [GPMXWOCB14], 变分自编码器 [KW13], 和流模型 [DSB17],基于分数的生成模型（SGM） [SE19] 通过给定从未知分布中抽取的一组样本，学习从该分布中采样。本节首先简要介绍 SGM 的关键要素，然后详细说明 SGM 的 DDPM 解释 §2。2，这是 RePaint 的基础$^+$。

2.1 基于分数的生成模型

分数生成模型的核心由两个随机微分方程（SDE）构成。最简单形式的正向随机微分方程(1)是一个奥恩斯坦-乌伦贝克（OU）过程，它将数据分布 $q(\vec{\boldsymbol{x}_0})$转化为参考分布，在大多数情况下，该分布为 正态分布 $(\boldsymbol{0}, \boldsymbol{I})$。时刻 $\vec{\boldsymbol{x}_t}$ 的样本 t 满足：

$$d\vec{\boldsymbol{x}_t} = -\frac{1}{2}\vec{\boldsymbol{x}_t}dt + dW_t;\ \vec{\boldsymbol{x}_0} \sim q(\vec{\boldsymbol{x}_0}). \tag{1}$$

给定时间重缩放参数 α_t和 β_t，其中 $\alpha_t = \int_0^t \beta_s ds$，以及归一化时间 $t = \{0, \cdots, 1\}$， （1）的转移核为：

$$q\left(\vec{\boldsymbol{x}_t} \mid \vec{\boldsymbol{x}_0}\right) = \mathcal{N}\left(\vec{\boldsymbol{x}_t}; \exp\left(-\frac{\alpha_t}{2}\right)\vec{\boldsymbol{x}_0}, \left(1 - \exp\left(-\alpha_t\right)\right)\boldsymbol{I}\right),$$

这将导出广为人知的正向随机微分方程形式：

$$d\vec{\boldsymbol{x}_t} = -\frac{1}{2}\beta_t\vec{\boldsymbol{x}_t} + \sqrt{\beta_t}d\mathbf{B}_t;\ \vec{\boldsymbol{x}_0} \sim q(\vec{\boldsymbol{x}_0}).$$

其中，$\{d\mathbf{B}_t\}_{t\geq 0}$ 表示 $\mathbb{R}^d$中的标准布朗运动。反向随机微分方程 的形式与正向随机微分方程类似，但时间发生了反转：

$$d\overleftarrow{\boldsymbol{x}_t} = \frac{1}{2}\beta_{1-t}\overleftarrow{\boldsymbol{x}_t} + \beta_{1-t}\nabla \log q(\overrightarrow{\boldsymbol{x}_t}|\overrightarrow{\boldsymbol{x}_0}) + \sqrt{\beta_{1-t}}d\mathbf{B}_t,$$

其中 $\overleftarrow{\boldsymbol{x}_T} \sim p\left(\overleftarrow{\boldsymbol{x}_T}\right) \coloneqq \mathcal{N}(\mathbf{0}, \boldsymbol{I})$。实际应用这些随机微分方程时，通常采用欧拉-丸山离散化方案。由于我们无法访问 $q(\overrightarrow{\boldsymbol{x}_0})$，因此使用神经网络 $s_{\boldsymbol{\theta}}\left(\overrightarrow{\boldsymbol{x}_t}, t\right)$ 来近似得分函数 $\nabla \log q(\overrightarrow{\boldsymbol{x}_t}|\overrightarrow{\boldsymbol{x}_0})$。训练时通过最小化得分匹配目标 [HD05; Vin11; SE19]:

$$\min_{\theta} \mathop{\mathbb{E}}_{\substack{\overrightarrow{\boldsymbol{x}_0}\sim q(\overrightarrow{\boldsymbol{x}_0}) \\ \overrightarrow{\boldsymbol{x}_t}\sim q(\overrightarrow{\boldsymbol{x}_t}|\overrightarrow{\boldsymbol{x}_0})}} \left[\left\|\nabla \log q(\overrightarrow{\boldsymbol{x}_t}|\overrightarrow{\boldsymbol{x}_0}) - s_{\boldsymbol{\theta}}\left(\overrightarrow{\boldsymbol{x}_t}, t\right)\right\|^2\right].$$

接下来，我们讨论去噪扩散概率模型（DDPM）及其与分数生成模型（SGM）的联系。

2.2 去噪扩散概率模型

扩散模型 [SWMG15; HJA20] 是一类新兴的生成模型，与基于分数的生成模型具有显著相似的特性 [SE19]。它由两个随机过程组成。首先，前向过程（扩散过程）根据方差调度逐渐向图像添加高斯噪声。这是一个马尔可夫链，其平稳分布通常设为易于处理且便于采样的分布，例如标准高斯分布 $\mathcal{N}(\mathbf{0}, \boldsymbol{I})$。其次，反向过程（去噪过程）学习逐步对从易处理分布中抽取的样本进行去噪。去噪过程持续进行，直至从原始数据分布中生成高质量图像。

2.2.1 前向过程

令 $\overrightarrow{\boldsymbol{x}_0} \sim q(\overrightarrow{\boldsymbol{x}_0})$ 表示从数据分布 $q(\overrightarrow{\boldsymbol{x}_0})$ 中抽取的 $\mathbb{R}^d$ 个样本。对于离散时间步 $1 \leq t \leq T$， $q(\boldsymbol{x}_t)$ 表示时间 t处的分布。在前向过程中，扩散按照固定 高斯转移核进行，即

$$q\left(\overrightarrow{\boldsymbol{x}_{1:T}}|\overrightarrow{\boldsymbol{x}_0}\right) \coloneqq \prod_{t=1}^{T} q(\overrightarrow{\boldsymbol{x}_t}|\overrightarrow{\boldsymbol{x}_{t-1}});\ q(\overrightarrow{\boldsymbol{x}_t}|\overrightarrow{\boldsymbol{x}_{t-1}}) \coloneqq \mathcal{N}\left(\overrightarrow{\boldsymbol{x}_t}; \sqrt{1-\beta_t}\overrightarrow{\boldsymbol{x}_{t-1}}, \beta_t \boldsymbol{I}_d\right), \tag{2}$$

其中 $\beta_1, \ldots, \beta_T$表示确定性方差调度。因此，从 $q(\overrightarrow{\boldsymbol{x}_t}|\overrightarrow{\boldsymbol{x}_{t-1}})$ 中抽取的样本表示为 $\overrightarrow{\boldsymbol{x}_t} = \sqrt{1-\beta_t}\overrightarrow{\boldsymbol{x}_{t-1}} + \sqrt{\beta_t}\epsilon$，其中 $\epsilon \sim$ 正态分布 $(\mathbf{0}, \boldsymbol{I}_d)$。第一项称为漂移，第二项称为弥散。高斯扩散过程的一个重要性质是，在任意中间时间步进行采样时，其形式都很简单。对于 $\alpha_t \coloneqq 1-\beta_t$ 和 $\bar{\alpha}_t \coloneqq \prod_{s=1}^{t} \alpha_s$，时间 t 处的条件概率变为：

$$q\left(\overrightarrow{\boldsymbol{x}_t}|\overrightarrow{\boldsymbol{x}_{t-1}}\right) = \mathcal{N}\left(\overrightarrow{\boldsymbol{x}_t}; \sqrt{\bar{\alpha}_t}\overrightarrow{\boldsymbol{x}_0}, (1-\bar{\alpha}_t)\boldsymbol{I}_d\right).$$

2.2.2 反向过程

反向过程 是一个马尔可夫链，初始化于前向过程的平稳分布，即 $p(\overleftarrow{\boldsymbol{x}_T})$ = 正态分布 $(\mathbf{0}, \boldsymbol{I}_d)$。与前向过程不同，反向过程 $p_\theta\left(\overleftarrow{\boldsymbol{x}_{0:T}}\right)$ 由学习得到的 高斯核生成，即，

$$p_\theta\left(\overleftarrow{\boldsymbol{x}_{0:T}}\right) = p(\overleftarrow{\boldsymbol{x}_T}) \prod_{t=1}^{T} p_\theta\left(\overleftarrow{\boldsymbol{x}_{t-1}}|\overleftarrow{\boldsymbol{x}_t}\right);\ p_\theta\left(\overleftarrow{\boldsymbol{x}_{t-1}}|\overleftarrow{\boldsymbol{x}_t}\right) \coloneqq \mathcal{N}\left(\overleftarrow{\boldsymbol{x}_{t-1}}; \mu_\theta\left(\overleftarrow{\boldsymbol{x}_t}, t\right), \Sigma_\theta\left(\overleftarrow{\boldsymbol{x}_t}, t\right)\right), \tag{3}$$

其中 $\mu_\theta\left(\overleftarrow{\boldsymbol{x}_t}, t\right)$ 和 $\Sigma_\theta\left(\overleftarrow{\boldsymbol{x}_t}, t\right)$ 是由 θ 参数化的神经网络，分别用于预测 $p_\theta\left(\overleftarrow{\boldsymbol{x}_{t-1}}|\overleftarrow{\boldsymbol{x}_t}\right)$的均值和方差。

2.2.3 训练与推理

DDPM 旨在最大化由反向过程生成的样本的似然。根据方程 (3)，通过对其余随机变量进行边缘化，可以得到赋予该样本的概率，即 $\overleftarrow{\boldsymbol{x}}_1, \overleftarrow{\boldsymbol{x}}_2, \cdots \overleftarrow{\boldsymbol{x}}_T$ 如下式所示：

$$p(\overleftarrow{\boldsymbol{x}}_0) = \int p(\overleftarrow{\boldsymbol{x}}_{0:T}) d\overleftarrow{\boldsymbol{x}}_{1:T},$$

这不易计算。通过考虑前向过程与反向过程之间的相对概率 [Jar97；SWMG15]，从而得到变分下界，

$$\mathbb{E}\left[-\log p_\theta\left(\overleftarrow{\boldsymbol{x}}_0\right)\right] \leq \mathbb{E}_q\left[-\log \frac{p_\theta\left(\overleftarrow{\boldsymbol{x}}_{0:T}\right)}{q\left(\overrightarrow{\boldsymbol{x}}_{1:T} | \overrightarrow{\boldsymbol{x}}_0\right)}\right] = \mathbb{E}_q\left[-\log p\left(\overleftarrow{\boldsymbol{x}}_T\right) - \sum_{t \geq 1} \log \frac{p_\theta\left(\overleftarrow{\boldsymbol{x}}_{t-1} | \overleftarrow{\boldsymbol{x}}_t\right)}{q\left(\overrightarrow{\boldsymbol{x}}_t | \overleftarrow{\boldsymbol{x}}_{t-1}\right)}\right]. \tag{4}$$

该目标函数可进一步简化为

$$\mathbb{E}_q\Big[\mathcal{D}_{KL}\left(q\left(\overrightarrow{\boldsymbol{x}}_T | \overrightarrow{\boldsymbol{x}}_0\right) \| p\left(\overleftarrow{\boldsymbol{x}}_T\right)\right) + \sum_{t>1} \mathcal{D}_{KL}\left(q\left(\overrightarrow{\boldsymbol{x}}_{t-1} | \overrightarrow{\boldsymbol{x}}_t, \overrightarrow{\boldsymbol{x}}_0\right) \| p_\theta\left(\overleftarrow{\boldsymbol{x}}_{t-1} | \overleftarrow{\boldsymbol{x}}_t\right)\right) - \log p_\theta\left(\overleftarrow{\boldsymbol{x}}_0 | \overleftarrow{\boldsymbol{x}}_1\right)\Big], \tag{5}$$

其中 $q\left(\overrightarrow{\boldsymbol{x}}_{t-1} | \overrightarrow{\boldsymbol{x}}_t, \overrightarrow{\boldsymbol{x}}_0\right)$ = 正态分布 $\left(\overrightarrow{\boldsymbol{x}}_{t-1}; \tilde{\mu}_t\left(\overrightarrow{\boldsymbol{x}}_t, \overrightarrow{\boldsymbol{x}}_0\right), \tilde{\beta}_t \boldsymbol{I}_d\right)$，$\tilde{\mu}_t\left(\overrightarrow{\boldsymbol{x}}_t, \overrightarrow{\boldsymbol{x}}_0\right)$: $= \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t} \overrightarrow{\boldsymbol{x}}_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} \overrightarrow{\boldsymbol{x}}_t$ 以及 $\tilde{\beta}_t \boldsymbol{I}_d$: $= \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$。方程（5）包含三个关键项。第一项衡量的是：在给定干净样本 $\overrightarrow{\boldsymbol{x}}_0 \sim q\left(\overrightarrow{\boldsymbol{x}}_0\right)$ 的条件下，前向过程的平稳分布与反向过程初始分布之间的散度。实际上，由于 OU 过程呈指数收敛，该散度可以忽略不计。因为没有可训练参数，所以忽略这一项，$\boldsymbol{\theta}$。

第二 项衡量了中间时间步上正向过程与反向过程之间的散度，$1 < t \leq T$。由于条件概率为高斯分布，因此可以显式计算为：

$$\mathbb{E}_q\left[\frac{1}{2\beta_t^2}\left\|\tilde{\mu}_t\left(\overrightarrow{\boldsymbol{x}}_t, \overrightarrow{\boldsymbol{x}}_0\right) - \mu_\theta\left(\overleftarrow{\boldsymbol{x}}_t, t\right)\right\|^2\right] + Constant, \tag{6}$$

其中 $p_\theta\left(\overleftarrow{\boldsymbol{x}}_{t-1} | \overleftarrow{\boldsymbol{x}}_t\right)$ = 正态分布 $\left(\overleftarrow{\boldsymbol{x}}_{t-1}; \mu_\theta\left(\overleftarrow{\boldsymbol{x}}_t, t\right), \beta_t \boldsymbol{I}\right)$。在训练阶段，$\overleftarrow{\boldsymbol{x}}_t$ 替换为 $\overrightarrow{\boldsymbol{x}}_t$，因为第一项（5）可以忽略。此外，将后验均值重参数化为 $\mu_\theta\left(\overrightarrow{\boldsymbol{x}}_t, t\right) = \frac{1}{\sqrt{\alpha_t}}\left(\overrightarrow{\boldsymbol{x}}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} \epsilon_\theta\left(\overrightarrow{\boldsymbol{x}}_t, t\right)\right)$ 可得到更好的结果以及简化后的损失函数 [HJA20]：

$$\mathbb{E}_{\boldsymbol{x}_0, \epsilon}\left[\frac{1}{2\alpha_t\left(1-\bar{\alpha}_t\right)}\left\|\epsilon - \epsilon_\theta\left(\sqrt{\bar{\alpha}_t}\, \overrightarrow{\boldsymbol{x}}_0 + \sqrt{1-\bar{\alpha}_t}\, \epsilon, t\right)\right\|^2\right]. \tag{7}$$

在推理过程中，对于 $\epsilon \sim$ 正态分布 $(\mathbf{0}, \boldsymbol{I})$，通过 $\overleftarrow{\boldsymbol{x}}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\overleftarrow{\boldsymbol{x}}_t - \frac{\beta_t}{\sqrt{1}}$ 生成一个样本 $\frac{}{-\bar{\alpha}_t} \epsilon_\theta\left(\overleftarrow{\boldsymbol{x}}_t, t\right)\right) + \sqrt{\beta_t}\, \epsilon$。

第三个项是标准的最大似然估计量。通过这种重参数化，概率学习问题转化为一个实际可行的最小均方误差（MMSE）问题。训练去噪扩散概率模型的目标是获得一个最优的 θ^*，它通过后验均值（6）或噪声（7）求解最小均方误差问题。

推理期间，反向过程从前向过程的平稳分布初始化，通常为 $\mathcal{N}(\mathbf{0}, \boldsymbol{I})$。接着，按照反向高斯转移核进行采样，使用 $\mu_{\boldsymbol{\theta}^*}\left(\overleftarrow{\boldsymbol{x}}_t, t\right)$。最后一步中，不含噪声的 $\mu_{\boldsymbol{\theta}^*}\left(\overleftarrow{\boldsymbol{x}}_1, 1\right)$ 会被显示，它是来自原始数据分布的一个样本。忽略一些常数因子，如果使用高斯转移核，SGM 的前向和反向随机微分方程 与 DDPM 的前向和反向过程 是相同的。本文中，我们交替使用这些术语。

算法 1： DDPM 训练

输入： 初始化权重 $\boldsymbol{\theta}$，步长 η

重复

采样 $\overrightarrow{\boldsymbol{x}_0} \sim q(\overrightarrow{\boldsymbol{x}_0})$，$\epsilon \sim$ 正态分布 $(\mathbf{0}, \boldsymbol{I}_d)$，

$t \sim Uniform\ (\{1, \ldots, T\})$

损失函数 (6)：$\mathcal{L}(\boldsymbol{\theta}) = \left\| \tilde{\mu}_t\left(\overrightarrow{\boldsymbol{x}_t}, \overrightarrow{\boldsymbol{x}_0}\right) - \mu_\theta\left(\overrightarrow{\boldsymbol{x}_t}, t\right) \right\|^2$ 或

(7)：$\mathcal{L}(\boldsymbol{\theta}) = \left\| \epsilon - \epsilon_\theta\left(\sqrt{\bar{\alpha}_t}\ \overrightarrow{\boldsymbol{x}_0} + \sqrt{1-\bar{\alpha}_t}\ \epsilon, t\right) \right\|^2$

运行梯度下降 $\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \eta \nabla \mathcal{L}(\boldsymbol{\theta})$

直至 收敛

输出： 训练后的权重 $\boldsymbol{\theta}$

算法 2： RePaint$^+$ 推理特例

输入：给定 $\overrightarrow{\boldsymbol{x}_0} \sim q\left(\overrightarrow{\boldsymbol{x}_0}\right)$、图像修复掩码 $\boldsymbol{m}$、DDPM 训练后的权重 $\boldsymbol{\theta}$、固定方差 $\beta > 0$ 绘制 $\overleftarrow{\boldsymbol{x}_1} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{I}_d)$，对于 $r = 1, \ldots\ldots, R$ 执行反向马尔可夫过程 (3) 的漂移：$\mu_{\boldsymbol{\theta}}\left(\overleftarrow{\boldsymbol{x}_1}\right)$ 漂移对齐：$\overleftarrow{\boldsymbol{x}_0} = \mu_{\boldsymbol{\theta}}\left(\overleftarrow{\boldsymbol{x}_1}\right) \times \frac{1}{\sqrt{1-\beta}}$ 图像修复：$\overleftarrow{\boldsymbol{x}_0} = \boldsymbol{m} \odot \overleftarrow{\boldsymbol{x}_0} + (1-\boldsymbol{m}) \odot \overrightarrow{\boldsymbol{x}_0}$ 更新 $\overleftarrow{\boldsymbol{x}_1} \leftarrow \overleftarrow{\boldsymbol{x}_0}$ 循环结束反向马尔可夫过程 (3) 的漂移：$\mu_{\boldsymbol{\theta}}\left(\overleftarrow{\boldsymbol{x}_0}\right)$ 漂移对齐：$\overleftarrow{\hat{\boldsymbol{x}}_0} = \mu_{\boldsymbol{\theta}}\left(\overleftarrow{\boldsymbol{x}_0}\right) \times \frac{1}{\sqrt{1-\beta}}$ 输出：$\overleftarrow{\hat{\boldsymbol{x}}_0}$

3 理论结果

本节将介绍我们的主要生成与图像修复结果：首先针对双状态模型（§3.2.1、§3.2.2、 §3.2.3），随后将其推广至多状态扩散模型（§3.2.4）。

3.1 问题设定

目标是最小化散度 D $(p_{\boldsymbol{\theta}}(\overleftarrow{\boldsymbol{x}_0})\ q(\overrightarrow{\boldsymbol{x}_0}))$，其中 D(., .) 可以是全变差（TV）距离 [CCLLSZ22], 、KL 散度 [HJA20], ，或 Wasserstein 距离 [Bor22]。对于生成模型，只要 $\overleftarrow{\boldsymbol{x}_0}$ 的分布与 $q(\overrightarrow{\boldsymbol{x}_0})$ 相匹配，就达成了目标。然而，对于基于扩散的图像修复，有必要生成一个与 ϵ 接近的图像，即 $\|\overleftarrow{\boldsymbol{x}_0} - \overrightarrow{\boldsymbol{x}_0}\| \leq \epsilon$，并且 $\overleftarrow{\boldsymbol{x}_0}$ 位于 $q(\overrightarrow{\boldsymbol{x}_0})$ 的支撑集内。假设数据分布 $q\left(\overrightarrow{\boldsymbol{x}_0}\right)$ 的支撑集是 $\mathbb{R}^d$ 中的一个 k 维子空间。对于 $\boldsymbol{A} \in \mathbb{R}^{d \times k}$，我们有原始样本 $\overrightarrow{\boldsymbol{x}_0} = \boldsymbol{A}\boldsymbol{z}_0$，其中 $\boldsymbol{z}_0 \in \mathbb{R}^k$ 服从 N（0，$\boldsymbol{I}_k$），且 $\boldsymbol{A}$ 为满秩，即 $rank(\boldsymbol{A}) = k \leq d$。我们要探究的是，为生成建模任务训练的模型是否有助于在图像修复任务中恢复图像缺失的部分。如前文 §1 所述，使用优化技术在该模型下进行恢复已有充分研究。我们的目标是利用这一成熟框架，分析并为使用去噪扩散概率模型进行样本恢复提供新的见解。

3.2 分析得出的主要见解

我们的主要结果是：给定一个为生成建模任务训练的去噪扩散概率模型的权重，我们无需针对图像修复任务重新训练，就能将其目标函数用于图像修复，并展示出完美的

算法 3： RePaint$^+$ 推理一般情形

输入： 给定 $\overrightarrow{\boldsymbol{x}_0} \sim q\left(\overrightarrow{\boldsymbol{x}_0}\right)$、DDPM 训练后的权重 $\boldsymbol{\theta}$、初始方差 $\beta_0 = 0$、方差调度 $\{\beta_t\}_{t\geq 0}$、漂移的对齐系数（$\{\omega_t\}_{t\geq 0}$）和弥散的对齐系数（$\{\xi_t\}_{t\geq 0}$）、初始 $\overleftarrow{\boldsymbol{x}_T} \sim \mathcal{N}\left(\mathbf{0},\ \boldsymbol{I}_d\right)$

对于 $t = T, \ldots, 1$ **do**

  对于 $r = 1, \ldots, R$ **do**

    如果 $t > 1$，**则**

      $\overrightarrow{\epsilon} \sim$ 正态分布 $(\mathbf{0},\ \boldsymbol{I}_d)$ 且 $\overleftarrow{\epsilon} \sim$ 正态分布 $(\mathbf{0},\ \boldsymbol{I}_d)$

    else

      $\overrightarrow{\epsilon} = 0$ 且 $\overleftarrow{\epsilon} = 0$

    结束如果

    已知部分：$\overrightarrow{\boldsymbol{x}_{t-1}} = \sqrt{\bar{\alpha}_{t-1}}\overrightarrow{\boldsymbol{x}_0} + \sqrt{1-\bar{\alpha}_{t-1}}\overrightarrow{\epsilon}$

    漂移 $\mu_{\boldsymbol{\theta}}\left(\overleftarrow{\boldsymbol{x}_t},\ t\right) = \frac{1}{\sqrt{\alpha_t}}\left(\overleftarrow{\boldsymbol{x}_t} - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_{\boldsymbol{\theta}}\left(\overleftarrow{\boldsymbol{x}_t},\ t\right)\right)$ 以及弥散 $\sqrt{\beta_t}\overleftarrow{\epsilon}$ 来自反向过程（3）

    漂移对齐：$\mu_{\boldsymbol{\theta}}\left(\overleftarrow{\boldsymbol{x}_t},\ t\right) \times \omega_t$

    离散项对齐：$\sqrt{\beta_t}\overleftarrow{\epsilon} \times \xi_t$

    计算 $\overleftarrow{\boldsymbol{x}_{t-1}} = \mu_{\boldsymbol{\theta}}\left(\overleftarrow{\boldsymbol{x}_t},\ t\right) \times \omega_t + \sqrt{\beta_t}\overleftarrow{\epsilon} \times \xi_t$

    图像修复：$\overleftarrow{\boldsymbol{x}_{t-1}} = \boldsymbol{m} \odot \overleftarrow{\boldsymbol{x}_{t-1}} + (1-\boldsymbol{m}) \odot \overrightarrow{\boldsymbol{x}_{t-1}}$

    如果 $r < R$ 且 $t > 1$ **则**

      抽取 $\epsilon \sim$ 正态分布 $(\mathbf{0}, \boldsymbol{I}_d)$

      推前：$\overleftarrow{\boldsymbol{x}_t} = \sqrt{1-\beta_t}\overleftarrow{\boldsymbol{x}_{t-1}} + \sqrt{\beta_t}\epsilon$

    else

      更新 $\overleftarrow{\boldsymbol{x}_1} \leftarrow \overleftarrow{\boldsymbol{x}_0}$

    结束 if

  结束 for

结束 for

漂移 $\mu_{\boldsymbol{\theta}}\left(\overleftarrow{\boldsymbol{x}_1},\ 1\right) = \frac{1}{\sqrt{\alpha_1}}\left(\overleftarrow{\boldsymbol{x}_1} - \frac{\beta_1}{\sqrt{1-\bar{\alpha}_1}}\epsilon_{\boldsymbol{\theta}}\left(\overleftarrow{\boldsymbol{x}_1}, 1\right)\right)$ 来自反向马尔可夫过程（3）

漂移对齐：$\mu_{\boldsymbol{\theta}}\left(\overleftarrow{\boldsymbol{x}_1},\ 1\right) \times \omega_1$

计算 $\overleftarrow{\hat{\boldsymbol{x}}_0} = \mu_{\boldsymbol{\theta}}\left(\overleftarrow{\boldsymbol{x}_1},\ 1\right) \times \omega_1$

输出： $\overleftarrow{\hat{\boldsymbol{x}}_0}$

以线性收敛率恢复。在假设 1 下，我们首先证明 DDPM训练算法1 学到的权重能够成功捕捉底层数据生成分布，见定理 3。然后，我们在假设 1 和 2 下证明，DDPM训练算法1 的最优解能够很好地泛化到具有未知掩码的图像修复任务，见算法 2 和算法 3。与方差调度 $\{\beta_t\}_{t\geq 0}$ 类似，漂移的对齐系数 $\{\omega_t\}_{t\geq 0}$ 和弥散的对齐系数 $\{\xi_t\}_{t\geq 0}$ 可以预先计算。例如，在算法 2 中，我们选择 $\beta_t = \beta$ 和 $\omega_t = 1/\sqrt{1-\beta}$[1]。

假设 1。 T 数据生成模型的列向量 $\boldsymbol{A}$ 是正交归一的，即 $\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{I}_k$.

假设 2。 Gi 给定图像修复掩码 $\boldsymbol{m} \in \{0,1\}^d$，以下条件成立：$\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m}) \prec \boldsymbol{I}_d$.

假设 1 是用于简化表达式的一个温和假设。对于**假设 2**，回顾一下，$\boldsymbol{D}(\boldsymbol{m})$ 是一个 $d \times d$ 对角矩阵（一个掩码），其对角线元素在数据缺失（即数据被掩码）时设为 1，否则（数据可用时）设为0 。符号 $\prec$ 表示 $\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})$的谱范数（记作 $\lambda_{\max}$）严格小于 1。

假设2具有符合直觉的物理解释。由于对图像进行掩码处理会降低其自由能，因此掩码 $\boldsymbol{D}(\boldsymbol{m})$ 的行为必须良好，才能确保保留足够的

[1]对 RePaint 的改进 [LDRYTV22]已在算法 2 和**算法 3中突出显示**。

能量，使掩码图像得以忠实重建。如果 $\boldsymbol{D}(\boldsymbol{m}) = \boldsymbol{I}$，那么原始信号的全部能量都会丢失。在这种情况下，原始信号无法恢复。当$\boldsymbol{D}(\boldsymbol{m}) = \mathbf{0}$时，这是一个平凡情形，因为算法可以直接访问原始信号。我们的可恢复性结果适用于所有其余情形，即 $\boldsymbol{m} \sim \{0,1\}^d$ 且满足 $\lambda_{\max} < 1$。

一个值得进一步研究的方向，是精确刻画为了实现完美恢复，我们需要多少形式为 $\boldsymbol{D}(\boldsymbol{m})$ 的信息。压缩感知领域的文献已对这一问题进行了广泛研究 [BJPD17; WRL19]，但这并非本文的重点。我们将此类问题留待未来研究。

3.2.1 使用去噪扩散概率模型进行生成建模

这里，我们提出**定理3**，用于计算双状态模型下去噪扩散概率模型的解析解。一般而言，计算解析解徒劳无益，因为我们事先并不知道 $q(\overrightarrow{\boldsymbol{x}_0})$。然而，数据生成分布所具有的额外结构（如下游任务，例如图像修复中通常所假设的结构），使我们能够推导出解的显式形式 [BJPD17; WRL19]。我们考虑目标函数（6），该目标函数用于估计后验分布的均值。也可以使用重参数化目标函数（7）来估计噪声。这两种选择在实验中都能产生相似的结果 [HJA20]。

定理 3（生成建模）。假设 假设条件 ***1*** 成立。记 $\boldsymbol{\theta}^* = \operatorname{argmin} \mathcal{L}(\boldsymbol{\theta})$，其中 $\mathcal{L}(\boldsymbol{\theta})$ 定义如下：

$$\mathbb{E}_{\overrightarrow{\boldsymbol{x}_0}, \overrightarrow{\epsilon}} \left[\left\| \tilde{\mu}_1 \left(\overrightarrow{\boldsymbol{x}_1}(\overrightarrow{\boldsymbol{x}_0}, \overrightarrow{\epsilon}), \overrightarrow{\boldsymbol{x}_0} \right) - \mu_\theta \left(\overrightarrow{\boldsymbol{x}_1} \left(\overrightarrow{\boldsymbol{x}_0}, \overrightarrow{\epsilon} \right) \right) \right\|^2 \right].$$

对于固定方差 $\beta > 0$，若考虑函数近似器 $\mu_{\boldsymbol{\theta}} \left(\overrightarrow{\boldsymbol{x}_1} \left(\overrightarrow{\boldsymbol{x}_0}, \overrightarrow{\epsilon} \right) \right) = \boldsymbol{\theta} \overrightarrow{\boldsymbol{x}_1} \left(\overrightarrow{\boldsymbol{x}_0}, \overrightarrow{\epsilon} \right)$，则闭式解 $\boldsymbol{\theta}^* = \sqrt{1-\beta}\boldsymbol{A}\boldsymbol{A}^T$经由 $(1/\sqrt{1-\beta})$ 重新归一化后，可恢复 $q\left(\overrightarrow{\boldsymbol{x}_0}\right)$ 的真实子空间。 ——

证明。证明见附录A.1。 □

对于任意 $\overleftarrow{\boldsymbol{x}_1} \sim$ 正态分布 $(\mathbf{0}, \boldsymbol{I}_d)$，采用最优解 $\boldsymbol{\theta}^*$ 的反向过程会生成 $\overleftarrow{\boldsymbol{x}_0} = \sqrt{1-\beta}\boldsymbol{A}\boldsymbol{A}^T\overleftarrow{\boldsymbol{x}_1}$。经过重新归一化后，得到 $\overleftarrow{\boldsymbol{x}_0} = \boldsymbol{A}\left(\boldsymbol{A}^T\overleftarrow{\boldsymbol{x}_1}\right)$ 其 **无偏** 且具有 **单位协方差**，即$\mathbb{E}\left[\boldsymbol{A}^T\overleftarrow{\boldsymbol{x}_0}\right] = \mathbf{0}$并且 $\mathbb{E}\left[\left(\boldsymbol{A}^T\overleftarrow{\boldsymbol{x}_1}\right)\left(\boldsymbol{A}^T\overleftarrow{\boldsymbol{x}_1}\right)^T\right] = \mathbb{E}\left[\boldsymbol{A}^T\left(\overleftarrow{\boldsymbol{x}_1}\overleftarrow{\boldsymbol{x}_1}^T\right)\boldsymbol{A}\right] = \boldsymbol{I}_k$。这验证了 **算法 1**能够恢复底层数据分布的子空间，该分布支撑于具有高斯边缘分布的线性流形上。接下来，我们将详细分析图像修复 **算法 2**。

3.2.2 使用 RePaint 进行图像修复+

继续采用双状态模型，基于扩散的图像修复主要包含三个关键要素。首先，我们将反向过程初始化为 $\overleftarrow{\boldsymbol{x}_1} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_d)$。利用 **定理 3**中得到的反向高斯转移核，我们生成 $\overleftarrow{\boldsymbol{x}_0} = (1/\sqrt{1-\beta})\,\mu_{\boldsymbol{\theta}^*}\left(\overleftarrow{\boldsymbol{x}_1}\right)$。根据 **定理 3**，$\overleftarrow{\boldsymbol{x}_0}$ $\overleftarrow{\boldsymbol{x}_0}$ 是一个样本，它位于流形上。不过，我们注意到，它并不是我们要找的样本，因为它可能距离 $\overrightarrow{\boldsymbol{x}_0}$很远。

其次，根据给定的掩码 $\boldsymbol{m}$，我们按照以下规则：$\overrightarrow{\boldsymbol{x}_0^1} = \boldsymbol{m} \odot \overleftarrow{\boldsymbol{x}_0} + (\mathbf{1} - \boldsymbol{m}) \odot \overrightarrow{\boldsymbol{x}_0}$，用来自 $\overrightarrow{\boldsymbol{x}_0}$ 的已知信息替换 $\overleftarrow{\boldsymbol{x}_0}$ 的某些部分。这会生成一个任意样本 $\overrightarrow{\boldsymbol{x}_0^1} \in \mathbb{R}^d$。由于反向随机微分方程会将 $\mathbb{R}^d$ 中的任意样本推至流形上，因此在第一轮重采样轮次中，我们将 $\overrightarrow{\boldsymbol{x}_0^1}$ 作为 输入 提供给反向随机微分方程。或者，也可以在不添加任何噪声的情况下，让 $\overrightarrow{\boldsymbol{x}_0^1}$ 通过正向随机微分方程。在这种情况下，生成的样本 $\overrightarrow{\boldsymbol{x}_1^1}$ 将作为反向随机微分方程的输入。这只会带来一个额外的缩放因子，在我们的分析中很容易处理。

最后，反向随机微分方程生成 $\overleftarrow{\boldsymbol{x}}_0^1$，经过重新对齐后，预计与原始图像 $\left\|\overleftarrow{\boldsymbol{x}}_0^1 - \overrightarrow{\boldsymbol{x}}_0\right\| \leq \epsilon$ 相差 ϵ 以内。这就完成了一次重采样步骤。

通常，一次重采样步骤不足以获得 ϵ精度的解。我们希望了解**算法 2**需要多少个重采样步骤才能生成令人满意的修复图像。此外，它收敛到原始图像 ϵ邻域的速率是多少？定理 **定理 4**对这些问题给出了肯定的回答。

定理 4 （图像修复）。 设 假设 *1* 和 假设 *2* 成立。假设 $\lambda_{\max} := \left\|\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\right\|$ 且 $\boldsymbol{\theta}^* = \operatorname{argmin} \mathcal{L}(\boldsymbol{\theta})$, 其中 $\mathcal{L}(\boldsymbol{\theta})$ 定义如下:

$$\mathbb{E}_{\overrightarrow{\boldsymbol{x}}_0, \overrightarrow{\epsilon}}\left[\left\|\tilde{\mu}_1\left(\overrightarrow{\boldsymbol{x}}_1(\overrightarrow{\boldsymbol{x}}_0, \overrightarrow{\epsilon}), \overrightarrow{\boldsymbol{x}}_0\right) - \mu_\theta\left(\overrightarrow{\boldsymbol{x}}_1\left(\overrightarrow{\boldsymbol{x}}_0, \overrightarrow{\epsilon}\right)\right)\right\|^2\right].$$

对于任意掩码 $\boldsymbol{m} \sim \{0,\ 1\}^d$、固定方差 $\beta > 0$、部分已知图像 $\overrightarrow{\boldsymbol{x}}_0 \sim q\left(\overrightarrow{\boldsymbol{x}}_0\right)$以及初始化于 $\overleftarrow{\boldsymbol{x}}_1 \sim$正态分布 $(\mathbf{0},\ \boldsymbol{I}_d)$的反向随机微分方程，恢复 ϵ精度修复图像所需的算法算法 *2*所需的最大重采样步骤数 (r) ，即:

$$\left\|\overleftarrow{\boldsymbol{x}}_0^r - \overrightarrow{\boldsymbol{x}}_0\right\| \leq \lambda_{\max}^r\left(\frac{\|\boldsymbol{\theta}^*\|\left\|\overleftarrow{\boldsymbol{x}}_1 - \overrightarrow{\boldsymbol{x}}_0\right\|}{\sqrt{1-\beta}}\right) \leq \epsilon$$

上界为

$$\mathcal{O}\left(\frac{\log\left(\frac{\|\theta^*\|\left\|\overleftarrow{x}_1 - \overrightarrow{\boldsymbol{x}}_0\right\|}{\epsilon\sqrt{1-\beta}}\right)}{\log\left(\frac{1}{\lambda_{\max}}\right)}\right).$$

证明。证明见附录A.2。 □

我们从**定理4**中得出若干关键见解。

通用掩码原则：一个重要的观察是，只要掩码 $\boldsymbol{m}$ 是有效的图像修复掩码，收敛速率就与其无关。因此，**算法2** 无论原始样本是否曾使用此类掩码图像进行训练，都能将其恢复。这一点很重要，因为它使我们能够重新利用基于扩散的生成模型的目标函数来解决图像修复问题。

线性收敛率：定理**4 的一个重要推论**是，**算法2** 具有线性收敛率。因此，我们只需略微增加重采样步骤的数量，就能在 ϵ准确度方面取得显著提升。

信息瓶颈：影响收敛速率的一个有趣控制参数是$\lambda_{\max}$。根据**定理4**，显然， $\lambda_{\max}$ 的值越大，所需的重采样步骤就越多。由于$\lambda_{\max}$ 与缺失信息相关，因此可以理解，**算法2** 需要进行更多迭代，才能实现合理的协调。

低范数解： 在其他控制参数中，**算法 2** 似乎倾向于生成建模中的低范数解。这表明，在 DDPM 训练中具有归纳偏置 的**算法 1** 若倾向于低范数解，可能有助于基于扩散的图像修复。

与初始化的距离： 最后，与原始样本的距离，即 $\|\overleftarrow{x}_1 - \overrightarrow{\boldsymbol{x}}_0\|$，仅以对数级增长迭代复杂度。

3.2.3 使用含噪生成器进行图像修复

定理 **4 的一个直接推论是：当** 算法 **1 返回一个**近似解$\hat{\boldsymbol{\theta}} := \boldsymbol{\theta}^* + \delta\sqrt{1-\beta}$，与精确的解析解 $\boldsymbol{\theta}^*$相比。这是合理的设定，因为当数据生成分布 $q\left(\overrightarrow{\boldsymbol{x}}_0\right)$ 事先未知时，闭式解无法被精确计算，而大多数有意义的实际应用都属于这种情况。我们在考虑生成建模中的近似误差后，给出**定理 6**，其基于**假设 5** 。

近似解的扰动 δ 满足 $\left(\boldsymbol{A}\boldsymbol{A}^T+\delta\right)\boldsymbol{D}(\boldsymbol{m}) \prec \boldsymbol{I}_d$

假设 5。定理 .

定理 6。 假设 假设 *5* 成立，且 $\hat{\lambda}_{\max} \coloneqq \left\|\left(\boldsymbol{A}\boldsymbol{A}^T+\delta\right)\boldsymbol{D}(\boldsymbol{m})\right\|$。然后，使用 算法 *2* 进行 r 次重采样步骤，使用 δ-近似模型 $\boldsymbol{\theta}$ 可得：

$$\left\|\overleftarrow{\boldsymbol{x}_0^r}-\overrightarrow{\boldsymbol{x}_0}\right\| \leq \hat{\lambda}_{\max}^r\left(\frac{\|\boldsymbol{\theta}^*\|\left\|\overleftarrow{\boldsymbol{x}_1}-\overrightarrow{\boldsymbol{x}_0}\right\|}{\sqrt{1-\beta}}+\frac{\left\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\right\|\|\overrightarrow{\boldsymbol{x}_0}\|}{\left(1-\hat{\lambda}_{\max}\right)\sqrt{1-\beta}}\right)+\frac{\left\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\right\|\|\overrightarrow{\boldsymbol{x}_0}\|}{\left(1-\hat{\lambda}_{\max}\right)\sqrt{1-\beta}}.$$

证明。 证明见附录A.3。 □

定理 6 指出，由于 $\lambda_{\max}^r<1$，我们最终可恢复出一个 ζ-近似解，其中$\zeta=\frac{\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\|\|\overrightarrow{\boldsymbol{x}_0}\|\left(1-\hat{\lambda}_{\max}^r\right)}{\sqrt{1-\beta}}$。重建误差与去噪扩散概率模型在生成建模中的近似误差成正比。在没有噪声的情况下，我们有 $\left\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\right\| \to 0$，从而恢复出真实底层样本 $\overrightarrow{\boldsymbol{x}_0}$。

推论 7。 假设 $q\left(\overrightarrow{\boldsymbol{x}_0}\right)$ 具有紧支撑，且满足 $\|\overrightarrow{\boldsymbol{x}_0}\| \leq \kappa$。如果 $\delta=O\left(\epsilon\left(1-\hat{\lambda}_{\max}\right)/\kappa\right)$，那么

$$\sup_{\overrightarrow{\boldsymbol{x}_0}\sim q(\overrightarrow{\boldsymbol{x}_0})}\left\|\overleftarrow{\boldsymbol{x}_0^r}-\overrightarrow{\boldsymbol{x}_0}\right\| \leq \epsilon.$$

证明。 证明见附录A.4。 □

实际上，大多数数据生成分布都具有紧支撑。对于这些分布，假设 $\|\overrightarrow{\boldsymbol{x}_0}\| \leq \kappa$ 是合理的。这样，只要 DDPM 生成器的误差为 $\mathcal{O}(\epsilon)$ 量级，我们就能得到一个 ϵ-精确解。

3.2.4 重采样与减缓扩散

现在，我们将其推广到多状态模型。在 §3.2.2 和 §3.2.3中，我们讨论了重采样在协调图像缺失部分方面起着至关重要的作用。我们想知道，能否通过减缓扩散过程，在不进行重采样的情况下使用 $(T+1)$ 个状态，实现类似的结果，而不是使用 2 个状态并进行多次重采样。为此，我们首先在$(T+1)$ 个扩散状态下，推导生成建模的闭式解，见**定理 8**。接着，在假设正向随机微分方程和反向随机微分方程中的噪声独立同分布（IID）的情况下，我们在**推论 10** 中给出一个简化的闭式解。利用该解，我们在**定理 12** 中证明，与使用较少状态但进行多次重采样相比，在中间状态不进行重采样的图像修复效果较差。

定理 8。 设假设 *1*成立。对于确定性方差调度 $\{\beta_t\}_{t\geq 0}$，令
$\alpha_t=1-\beta_t$ $\bar{\alpha}_t=\prod_{s=0}^t\alpha_s$ $\gamma=\mathbb{E}_t\left(\frac{1}{1-\bar{\alpha}_t}\right)\sqrt{\alpha_t}\ 1-\bar{\alpha}_{t-1}$ $\nu=\mathbb{E}_t\left(\frac{1}{1-\bar{\alpha}_t}\right)\mathbb{E}_t(\bar{\alpha}_{t-1}\sqrt{\alpha_t})-$，，$\coloneqq()$
$\mathbb{E}_t\left(\frac{\bar{\alpha}_t}{1-\bar{\alpha}_t}\right)\mathbb{E}_t\ \sqrt{\alpha_t}-$ $t\sim Uniform\ 1,$ T $\boldsymbol{\theta}^*=\min$ $\boldsymbol{\theta}$ $\boldsymbol{\theta}$
$(())$ 以及：$()()()$，其中 $\{\cdots\}$。记arg$L()$，其中 $L()$ 定义如下：

$$\mathop{\mathbb{E}}_{\overrightarrow{\boldsymbol{x}_0},\epsilon,t}\left[\left\|\tilde{\mu}_t\left(\overrightarrow{\boldsymbol{x}_t},\overrightarrow{\boldsymbol{x}_0}\right)-\mu_\theta\left(\overrightarrow{\boldsymbol{x}_t},t\right)\right\|^2\right].$$

对于 $(T+1)$ 状态的扩散，若 $\mu_\theta\left(\overleftarrow{\boldsymbol{x}_t},\ t\right)=\boldsymbol{\theta}\begin{bmatrix}\overleftarrow{\boldsymbol{x}_t}\\ t\end{bmatrix}$，其中 $\boldsymbol{\theta}\in\mathbb{R}^{d\times(d+1)}$，则去噪扩散概率模型 算法 *1* 的闭式解为 $\boldsymbol{\theta}^*=\left[\nu\boldsymbol{A}\boldsymbol{A}^T+\gamma\boldsymbol{I}_d,\ \mathbf{0}\right]$。

证明。证明见附录A.5。 □

一般而言，$\{\beta_t\}_{t\geq 0}$ 表示一个方差列表，其取值范围从 $\beta_{\min}$ 到 $\beta_{\max}$，例如余弦调度或线性调度 [HJA20]。为涵盖一般框架，我们将 ν 和 γ 表示为 β_t 的显式函数。根据方差调度的选择，人们可能希望精确计算这些值。例如，对于具有 $\beta_0 = 0$和 $\beta_1 = \beta$的两状态扩散，我们有 $()()()$和 $()(())$。这与**定理 3**中推导出的闭式解一致。

为便于分析，我们在 **推论 10** 的**假设 9**下给出一个简化的闭式解。

假设 9。 设 $\overrightarrow{\epsilon}$ 表示正向随机微分方程弥散过程中使用的噪声，即$\overrightarrow{\boldsymbol{x}_{t-1}} = \sqrt{\bar{\alpha}_{t-1}}\overrightarrow{\boldsymbol{x}_0} + \sqrt{1-\bar{\alpha}_{t-1}}\overrightarrow{\epsilon}$； $\overleftarrow{\epsilon}$表示反向随机微分方程中的噪声，即 $\overleftarrow{\boldsymbol{x}_{t-1}} = \mu_{\boldsymbol{\theta}}(\overleftarrow{\boldsymbol{x}_t}, t) + \sqrt{\beta_t}\overleftarrow{\epsilon}$，其中 $\overrightarrow{\boldsymbol{x}_t} = \sqrt{\bar{\alpha}_t}\overrightarrow{\boldsymbol{x}_0} + \sqrt{1-\bar{\alpha}_t}\overleftarrow{\epsilon}$。设 $\overrightarrow{\epsilon}$和 $\overleftarrow{\epsilon}$是从 $\boldsymbol{N}(\mathbf{0}, \boldsymbol{I}_d)$中抽取的独立同分布高斯随机变量。

推论 10. 假设 假设 ***9***成立。对于 $\bar{\nu} := \mathbb{E}_t\left(\frac{1}{(1-\bar{\alpha}_t)}\right)\mathbb{E}_t\left(\bar{\alpha}_{t-1}\sqrt{\alpha_t}\right) - \mathbb{E}_t\left(\frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)}\right)\mathbb{E}_t\left(\bar{\alpha}_{t-1}\sqrt{\alpha_t}\right)$在与定理 ***8***相同的设定下，去噪扩散概率模型的闭式解 算法 ***1*** 变为$\boldsymbol{\theta}^* = \left[\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T,\ \mathbf{0}\right]$。

证明。 证明见附录A.6。 □

为简单起见，我们使用这个闭式解来修复中间状态。我们将在**定理 11** 中证明，对于在 $(T+1)$ 个状态上的扩散，使用去噪扩散概率模型的生成模型能够学会从数据流形中采样。

定理 11。 对于漂移系数 $\omega_t = \omega = \frac{1}{\bar{\nu}\sqrt{2}^{1/T}}$和弥散 $\xi_t = \sqrt{\frac{2^{(t-1)/T}}{2\beta_t}\bar{\nu}^{T-1}()}$，如果我们使用 $\boldsymbol{\theta}^* = \left[\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T,\ \mathbf{0}\right]$ 作为反向马尔可夫过程（*3*）中的高斯转移核，则

$$TV\left(p_{\boldsymbol{\theta}^*}\left(\overleftarrow{\boldsymbol{x}_0}\right), q\left(\overrightarrow{\boldsymbol{x}_0}\right)\right) = 0.$$

证明。 证明见附录A.7。 □

由于全变差距离为零，这保证了去噪扩散概率模型 DDPM 学会了目标分布$q\left(\overrightarrow{\boldsymbol{x}_0}\right)$。现在，让我们看看在中间状态下对 $\overrightarrow{\boldsymbol{x}_0}$ 的已知部分进行图像修复是否能恢复真实样本。对此，与 RePaint$^+$ 的一个重要区别在于，中间状态下没有重采样。

定理 12。 对于漂移系数 $\omega_t = \omega = \frac{1}{\bar{\nu}\sqrt{2}^{1/T}}$和弥散 $\xi_t = \sqrt{\frac{2^{(t-1)/T}}{2\beta_t}\bar{\nu}^{T-1}()}$，按照反向马尔可夫过程（*3*）并在中间状态进行图像修复，会生成满足以下条件的样本：

$$\begin{aligned}\overleftarrow{\boldsymbol{x}_0} = &\frac{1}{\sqrt{2}}\left(\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\right)^{T-1}\boldsymbol{A}\boldsymbol{A}^T\overleftarrow{\boldsymbol{x}_T} \\ &+\left(\sum_{t=1}^{T-1}\left(\frac{1}{\sqrt{2}}\right)^{t/T}\sqrt{\bar{\alpha}_t}\left(\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\right)^{t-1}\right)\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\mathbf{1}-\boldsymbol{m})\overrightarrow{\boldsymbol{x}_0} + \overrightarrow{\Sigma}\overrightarrow{\epsilon} + \overleftarrow{\Sigma}\overleftarrow{\epsilon},\end{aligned}$$

其中 $\Rightarrow$ $\leftarrow\Sigma$ 和 Σ 分别表示正向过程和反向过程中集体色散的协方差。

证明。证明见附录A.8。 □

我们注意到，$\overleftarrow{\boldsymbol{x}_0}$ 不一定是位于数据生成分布 $q\left(\overrightarrow{\boldsymbol{x}_0}\right)$流形上的图像。这部分是由于在正向和反向过程中注入噪声所产生的残余偏差。另一个令人担忧的问题是，生成部分与图像其余部分之间缺乏协调。值得注意的是，即使考虑在无限多个状态上进行扩散，这种偏差依然存在，这也说明了重采样比减缓扩散过程更为重要。实践中也观察到了这一现象（§5，[LDRYTV22]）。

相反，如 §3.2.2 中所述，经过多个重采样步骤的两状态扩散可以恢复原始样本。在 **算法 3** 中，我们将这一概念推广到 $(T+1)$ 个状态，并在每个中间状态对生成的图像进行协调。因此，我们逐步消除噪声的影响，避免其在反向转移过程中不断累积，从而实现平滑恢复。

4 结论

我们研究了使用去噪扩散概率模型进行图像修复时样本恢复的数学基础。我们提出了通用掩码原理，以证明基于扩散的图像修复无需重新训练即可轻松适应未见过的掩码。我们对线性设定下一个广受欢迎的修复算法进行了理论分析，揭示出一种此前尚未被理解、会妨碍完美恢复的偏差。通过重新对齐反向过程中的漂移和弥散，我们提出了一种名为 Repaint$^+$ 的算法，该算法消除了这种偏差，从而在极限情况下实现完美恢复。此外，我们证明了 RePaint$^+$ 会线性收敛到真实底层样本的 ϵ邻域。未来，研究更一般的非线性流形设定下的收敛性质将会很有意义。

参考文献

[BHSN10] Waheed U. Bajwa、Jarvis Haupt、Akbar M. Sayeed 和 Robert Nowak。“压缩信道感知：一种估计稀疏多径信道的新方法”。载于：IEEE会刊 98.6（2010年），第 1058–1076 页。doi: 10.1109/JPROC.2010.2042415 （第 3 页）。[BJPD17] Ashish Bora、Ajil Jalal、Eric Price 和 Alexandros G Dimakis。“使用生成模型进行压缩感知”。载于：国际机器学习会议。PMLR。2017年，第 537–546 页（第 3、9 页）。[Bor22] Valentin De Bortoli。“流形假设下去噪扩散模型的收敛性”。载于：机器学习研究汇刊（2022）。网址： https://openreview.net/forum?id=MhK5aXo3gB （第 2、3、7 页）。[CLL22] Hongrui Chen、Holden Lee 和 Jianfeng Lu。“分数函数生成建模的改进分析：在最小光滑性假设下给出易于使用的界”。载于：arXiv 预印本 arXiv:2211.01916（2022）（第 2 页）。[CCLLSZ22] Sitan Chen、Sinho Chewi、Jerry Li、Yuanzhi Li、Adil Salim 和 Anru R Zhang。“采样如同学习分数函数一样简单：具有最小数据假设的扩散模型理论”。载于：arXiv 预印本 arXiv:2209.11215（2022）（第 2、7 页）。[CKJGY21] Jooyoung Choi、Sungwon Kim、Yonghyun Jeong、Youngjune Gwon 和 Sungroh Yoon。“ILVR：去噪扩散概率模型的条件化方法”。载于：IEEE/CVF国际计算机视觉会议论文集。2021年，第 14367–14376 页（第 1 页）。[DDDD22] Giannis Daras、Yuval Dagan、Alex Dimakis 和 Constantinos Daskalakis。“分数引导的中间层优化：用于反问题的快速朗之万混合”。载于：国际机器学习会议。PMLR。2022年，第 4722–4753 页（第 1 页）。[DGE18] Manik Dhar、Aditya Grover 和 Stefano Ermon。“使用生成模型对压缩感知中的稀疏偏差进行建模”。载于：国际机器学习会议。PMLR。2018年，第 1214–1223 页（第 3 页）。[DN21] Prafulla Dhariwal 和 Alexander Nichol。“扩散模型在图像合成方面胜过生成对抗网络”。载于：神经信息处理系统进展 34（2021），第 8780–8794 页（第 2 页）。[DSB17] Laurent Dinh、Jascha Sohl-Dickstein 和 Samy Bengio。“使用 Real NVP 进行密度估计”。载于：国际学习表征会议。2017年。网址： https://openreview.net/forum?id=HkpbnH9lx （第 4 页）。[FMN16] Charles Fefferman、Sanjoy Mitter 和 Hariharan Narayanan。“检验流形假设”。载于：美国数学会杂志 29.4（2016），第 983–1049 页（第 2、3 页）。[FN03] M.A.T. Figueiredo 和 R.D. Nowak。“一种用于基于小波的图像复原的 EM算法”。载于：IEEE图像处理汇刊 12.8（2003年），第 906–916 页。doi: 10.1109/TIP.2003.814255 （第 3 页）。

[FNW07] Mário A. T. Figueiredo、Robert D. Nowak 和 Stephen J. Wright。《稀疏重建的梯度投影：在压缩感知及其他反问题中的应用》。载于：IEEE Journal of Selected Topics in Sig-nal Processing 1.4（2007），第586–597页。doi：10.1109/JSTSP.2007.910281（第3页）。[GPMXWOCB14] Ian Goodfellow、Jean Pouget-Abadie、Mehdi Mirza、Bing Xu、David Warde-Farley、Sherjil Ozair、Aaron Courville 和 Yoshua Bengio。《生成对抗网络》。载于：《神经信息处理系统进展》第27卷。2014年（第1、4页）。[Hag89] William W Hager。《矩阵逆的更新》。载于：SIAM评论 31.2（1989年），第221–239页（第18、24页）。[HJA20] Jonathan Ho、Ajay Jain 和 Pieter Abbeel。《去噪扩散概率模型》。载于：《神经信息处理系统进展》33（2020），第6840–6851页（第1–3、5–7、9、12页）。[HD05] Aapo Hyvärinen 和 Peter Dayan。《通过分数匹配估计非归一化统计模型》。载于：《机器学习研究期刊》6.4（2005年）（第5页）。[JADPDT21] Ajil Jalal、Marius Arvinte、Giannis Daras、Eric Price、Alexandros G Dimakis 和 Jon Tamir。《基于深度生成先验的稳健压缩感知磁共振成像》。载于：《神经信息处理系统进展》34（2021），第14938–14954页（第1、3页）。[JLDC20] Ajil Jalal、Liu Liu、Alexandros G Dimakis 和 Constantine Caramanis。《使用生成模型进行稳健压缩感知》。载于：《神经信息处理系统进展》33（2020），第713–727页（第3页）。[Jar97] Christopher Jarzynski。《通过非平衡测量获得平衡自由能差：主方程方法》。载于：《物理评论E辑》56.5（1997年），第5018页（第6页）。[KAAL22] Tero Karras、Miika Aittala、Timo Aila 和 Samuli Laine。《阐明基于扩散的生成模型设计空间》。载于：《神经信息处理系统进展》。编者：Alice H. Oh、Alekh Agarwal、Danielle Belgrave 和 Kyunghyun Cho。2022年。网址：https://openreview.net/forum?id=k7FuTOWMOc7（第2页）。[KEES22] Bahjat Kawar、Michael Elad、Stefano Ermon 和 Jiaming Song。《去噪扩散复原模型》。载于：《神经信息处理系统进展》。2022年（第1、2页）。[KSSKM22] Dongjun Kim、Seungjae Shin、Kyungwoo Song、Wanmo Kang 和 Il-Chul Moon。《软截断：用于高精度分数估计的基于分数的扩散模型通用训练技术》。载于：《第39届国际机器学习会议论文集》第162卷。《机器学习研究论文集》。PMLR，2022年，第11201–11228页。网址：https://proceedings.mlr.press/v162/kim22i.html（第3页）。[KW13] Diederik P Kingma 和 Max Welling。《自编码变分贝叶斯》。载于：arXiv预印本 arXiv:1312.6114（2013年）（第1、4页）。

[KPHZC21] Zhifeng Kong、Wei Ping、Jiaji Huang、Kexin Zhao 和 Bryan Catanzaro。“DiffWave：一种用于音频合成的多用途扩散模型”。载于：国际学习表征会议。2021 年。网址：https://openreview.net/forum?id=a-xFK8Ymz5J（第 2 页）。

[LLT22a] Holden Lee、Jianfeng Lu 和 Yixin Tan。“基于分数的生成建模的多项式复杂度收敛性”。载于：*arXiv* 预印本 *arXiv:2206.06227*（2022）（第2页）。

[LLT22b] Holden Lee、Jianfeng Lu 和 Yixin Tan。“一般数据分布下基于分数的生成建模的收敛性”。载于：*arXiv* 预印本 *arXiv:2209.12381*（2022）（第2页）。

[LTGLH22] Xiang Lisa Li、John Thickstun、Ishaan Gulrajani、Percy Liang 和 Tatsunori Hashimoto。《Diffusion-LM 改进了可控文本生成》。载于：神经信息处理系统进展。由 Alice H. Oh、Alekh Agarwal、Danielle Belgrave 和 Kyunghyun Cho。2022。网址：https://openreview.net/forum?id=3s9IrEsjLyk（第2页）。

[LDRY全变差22] Andreas Lugmayr、Martin Danelljan、Andres Romero、Fisher Yu、Radu Timofte，以及 Luc Van Gool。“Repaint：使用去噪扩散概率模型”。载于：*IEEE/CVF*计算机视觉与模式识别。2022 年，第 11461–11471 页（页码 1–4，8，13，29）。

[PLXL21] Jialun Peng、Dong Liu、Songcen Xu 和 Houqiang Li。“生成多样化的用于图像修复的结构，采用分层 VQ-VAE”。载于：论文集 *IEEE/CVF*计算机视觉与模式识别会议。2021，第 10775–10784 页（第2页）。

[Pid22] Jakiw Pidstrigach。“基于分数的生成模型检测流形”。载于：神经信息处理系统进展。编者：Alice H. Oh，Alekh Agarwal、Danielle Belgrave 和 Kyunghyun Cho。2022。网址：https://openreview.net/forum?id=AiNrnIrDfD9 (页码2、3)。

[RDNCC22] Aditya Ramesh、Prafulla Dhariwal、Alex Nichol、Casey Chu 和 Mark Chen。“基于 CLIP 潜变量的分层文本条件图像生成”。载于：*arXiv* 预印本 *arXiv:2204.06125*（2022）（第 1页）。

[RPGGVRCS21] Aditya Ramesh、Mikhail Pavlov、Gabriel Goh、Scott Gray、Chelsea Voss、Alec Radford、Mark Chen 和 Ilya Sutskever。《零样本文本到图像生成》。载于：国际机器学习会议。PMLR。2021，第 8821–8831 页（第 1页）。

[RVV19] Ali Razavi、Aaron Van den Oord 和 Oriol Vinyals。“生成多样化的高保真度图像的 vq-vae-2”。载于：神经信息处理进展系统32 (2019)（页码2）。

[RBLEO22] Robin Rombach、Andreas Blattmann、Dominik Lorenz、Patrick Esser 和 Björn Ommer。“基于潜在扩散模型的高分辨率图像合成——型”。载于：*IEEE/CVF*计算机视觉与模式识别会议（*CVPR*）。2022年6月，第10684–10695页（共1、2页）。

[SCCLHSFN22] Chitwan Saharia, William Chan, HuiwenChang, ChrisLee, Jonathan Ho,Tim Salimans, David Fleet, and Mohammad Norouzi. “Palette：图像到图像扩散模型”。载于：*ACM SIGGRAPH 2022* 会议论文集。2022年，第 1–10 页（页码1、2）。

[SCSLWDGAMF+22] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily Denton, Seyed Kamyar Seyed Ghasemipour, Burcu Karagol Ayan, S Sara Mah- davi, Rapha Gontijo Lopes 等。“借助深度语言理解实现逼真文本到图像扩散模型”。载于：arXiv 预印本 arXiv:2205.11487（2022）（第 1 页）。[SWMG15] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Gan- guli。“利用非平衡热力学进行深度无监督学习”。载于：国际机器学习会议。PMLR。2015（第 1、2、5、6 页）。[SE19] Yang Song and Stefano Ermon。“通过估计数据分布的梯度进行生成建模”。载于：神经信息处理系统进展 32（2019）（第 1、4、5 页）。[SSXE21] Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon。“利用基于分数的生成模型解决医学成像中的逆问题”。载于：国际学习表征会议。2021（第 1 页）。[SSKKEP20] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole。“通过随机微分方程进行基于分数的生成建模”。载于：国际学习表征会议。2020（第 1、2 页）。[TSSE21] Yusuke Tashiro, Jiaming Song, Yang Song, and Stefano Ermon。“CSDI：用于概率时间序列插补的条件基于分数扩散模型”。载于：神经信息处理系统进展 34（2021），第 24804–24816 页（第 2 页）。[TSL00] Joshua B Tenenbaum, Vin de Silva, and John C Langford。“非线性降维的全局几何框架”。载于：《科学》290.5500（2000），第 2319–2323 页（第 2、3 页）。[Tib96] Robert Tibshirani。“通过套索回归进行回归收缩与变量选择”。载于：《英国皇家统计学会杂志：B辑（方法论）》58.1（1996），第 267–288 页（第 3 页）。[Vin11] Pascal Vincent。“分数匹配与去噪自编码器之间的联系”。载于：《神经计算》23.7（2011），第 1661–1674 页（第 5 页）。[WYWGLDQC18] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy。“Esrgan：增强型超分辨率生成对抗网络”。载于：欧洲计算机视觉会议（ECCV）研讨会论文集。2018（第 2 页）。[WDTSDM22] Jay Whang, Mauricio Delbracio, Hossein Talebi, Chitwan Saharia, Alexandros G. Dimakis, and Peyman Milanfar。“通过随机细化进行去模糊”。载于：IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集。2022年6月，第 16293–16303 页（第 1 页）。[WRL19] Yan Wu, Mihaela Rosca, and Timothy Lillicrap。“深度压缩感知”。载于：国际机器学习会议。PMLR。2019，第 6850–6860 页（第 9 页）。[YLYSLH18] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang。“结合上下文注意力的生成式图像修复”。载于：IEEE计算机视觉与模式识别会议论文集。2018，第 5505–5514 页（第 2 页）。

A 技术证明

A.1 定理证明3

对于两个状态之间的扩散，后验均值 $\tilde{\mu}_t\left(\overrightarrow{\boldsymbol{x}_t},\ \overrightarrow{\boldsymbol{x}_0}\right) \coloneqq \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\overrightarrow{\boldsymbol{x}_0} + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\overrightarrow{\boldsymbol{x}_t}$ 可简化为 $\tilde{\mu}_1\left(\overrightarrow{\boldsymbol{x}_1},\ \overrightarrow{\boldsymbol{x}_0}\right) = \overrightarrow{\boldsymbol{x}_0}$。因此，训练损失变为：

$$\begin{aligned}
&\min_{\boldsymbol{\theta}} \mathbb{E}_{\overrightarrow{\boldsymbol{x}_0},\overrightarrow{\epsilon}}\left[\left\|\tilde{\mu}_1(\overrightarrow{\boldsymbol{x}_1}\left(\overrightarrow{\boldsymbol{x}_0},\overrightarrow{\epsilon}\right),\overrightarrow{\boldsymbol{x}_0}) - \mu_\theta\left(\overrightarrow{\boldsymbol{x}_1}\left(\overrightarrow{\boldsymbol{x}_0},\overrightarrow{\epsilon}\right)\right)\right\|^2\right] = \\
&\mathbb{E}_{\overrightarrow{\boldsymbol{x}_0},\overrightarrow{\epsilon}}\left[\left\|\overrightarrow{\boldsymbol{x}_0} - \mu_{\boldsymbol{\theta}}\left(\overrightarrow{\boldsymbol{x}_1}\right)\right\|^2\right] = \mathbb{E}_{\overrightarrow{\boldsymbol{x}_0},\overrightarrow{\epsilon}}\left[\left\|\overrightarrow{\boldsymbol{x}_0} - \boldsymbol{\theta}\overrightarrow{\boldsymbol{x}_1}\right\|^2\right] = \\
&\mathop{\mathbb{E}}_{\substack{\overrightarrow{\boldsymbol{x}_0}\sim q \\ \overrightarrow{\epsilon}\sim\mathcal{N}(\mathbf{0},\boldsymbol{I}_d)}}\left[\left\|\overrightarrow{\boldsymbol{x}_0} - \boldsymbol{\theta}\left(\boldsymbol{x}_0\sqrt{1-\beta} + \overrightarrow{\epsilon}\sqrt{\beta}\right)\right\|^2\right] = \\
&\mathop{\mathbb{E}}_{\substack{\overrightarrow{\boldsymbol{x}_0}\sim q \\ \overrightarrow{\epsilon}\sim\mathcal{N}(\mathbf{0},\boldsymbol{I}_d)}}\left[\sum_{i=1}^{d}\left(\overrightarrow{\boldsymbol{x}_{0,i}} - \boldsymbol{\theta}_i^T\left(\overrightarrow{\boldsymbol{x}_0}\sqrt{1-\beta} + \overrightarrow{\epsilon}\sqrt{\beta}\right)\right)^2\right],
\end{aligned} \tag{8}$$

其中 $\boldsymbol{\theta}_i^T$ 表示 i^{th} 行，属于 $\boldsymbol{\theta}$。求解8式中的最小均方误差问题，关于 $\boldsymbol{\theta}_i^*$ 可得

$$\begin{aligned}
&\boldsymbol{\theta}_i^* = \mathop{\mathbb{E}}_{\overrightarrow{\boldsymbol{x}_0},\overrightarrow{\epsilon}}\left[\left(\overrightarrow{\boldsymbol{x}_0}\sqrt{1-\beta} + \overrightarrow{\epsilon}\sqrt{\beta}\right)\left(\overrightarrow{\boldsymbol{x}_0}\sqrt{1-\beta} + \overrightarrow{\epsilon}\sqrt{\beta}\right)^T\right]^{-1}\mathbb{E}_{\overrightarrow{\boldsymbol{x}_0},\overrightarrow{\epsilon}}\left[\overrightarrow{\boldsymbol{x}_{0,i}}\left(\overrightarrow{\boldsymbol{x}_0}\sqrt{1-\beta} + \overrightarrow{\epsilon}\sqrt{\beta}\right)\right] = \\
&\mathop{\mathbb{E}}_{\overrightarrow{\boldsymbol{z}_0},\overrightarrow{\epsilon}}\left[\left(\mathcal{A}\boldsymbol{z}_0\sqrt{1-\beta} + \overrightarrow{\epsilon}\sqrt{\beta}\right)\left(\mathcal{A}\boldsymbol{z}_0\sqrt{1-\beta} + \overrightarrow{\epsilon}\sqrt{\beta}\right)^T\right]^{-1}\mathbb{E}_{\boldsymbol{z}_0,\overrightarrow{\epsilon}}\left[\left(\boldsymbol{a}_i^T\boldsymbol{z}_0\right)\left(\mathcal{A}\boldsymbol{z}_0\sqrt{1-\beta} + \overrightarrow{\epsilon}\sqrt{\beta}\right)\right] = \\
&\left[(1-\beta)\mathcal{A}\mathbb{E}_{\boldsymbol{z}_0}\left[\boldsymbol{z}_0\boldsymbol{z}_0^T\right]\mathcal{A}^T + \beta\mathbb{E}_{\overrightarrow{\epsilon}}\left[\overrightarrow{\epsilon}\overrightarrow{\epsilon}^T\right]\right]^{-1}\left[\left(\sqrt{1-\beta}\right)\mathcal{A}\mathbb{E}_{\overrightarrow{\boldsymbol{x}_0}}\left[\overrightarrow{\boldsymbol{x}_0}\overrightarrow{\boldsymbol{x}_0}^T\right]\boldsymbol{a}_i\right].
\end{aligned} \tag{9}$$

利用以下事实： $\boldsymbol{z}_0$ 并且 ϵ 是具有单位协方差的相互独立的高斯随机变量，式（9）给出

$$\begin{aligned}
&\boldsymbol{\theta}_i^* = \left[(1-\beta)\mathcal{A}\mathcal{A}^T + \beta\boldsymbol{I}_d\right]^{-1}\left[\left(\sqrt{1-\beta}\right)\mathcal{A}\boldsymbol{a}_i\right] = \\
&\frac{\sqrt{1-\beta}}{\beta}\left[\boldsymbol{I}_d + \left(\sqrt{\frac{1-\beta}{\beta}}\mathcal{A}\right)\left(\sqrt{\frac{1-\beta}{\beta}}\mathcal{A}\right)^T\right]^{-1}\mathcal{A}\boldsymbol{a}_i.
\end{aligned}$$

接下来，我们使用伍德伯里矩阵恒等式 [Hag89]得到

$$\begin{aligned}
&\boldsymbol{\theta}_i^* = \\
&\frac{\sqrt{1-\beta}}{\beta}\left[\boldsymbol{I}_d - \left(\sqrt{\frac{1-\beta}{\beta}}\mathcal{A}\right)\left(\boldsymbol{I}_k + \left(\sqrt{\frac{1-\beta}{\beta}}\mathcal{A}\right)^T\left(\sqrt{\frac{1-\beta}{\beta}}\mathcal{A}\right)\right)^{-1}\left(\sqrt{\frac{1-\beta}{\beta}}\mathcal{A}\right)^T\right]\mathcal{A}\boldsymbol{a}_i = \\
&\frac{\sqrt{1-\beta}}{\beta}\left[\boldsymbol{I}_d - \left(\sqrt{\frac{1-\beta}{\beta}}\boldsymbol{A}\right)\left(\boldsymbol{I}_k\left(\frac{1-\beta}{\beta}\right)\mathcal{A}^T\mathcal{A}\right)^{-1}\left(\sqrt{\frac{1-\beta}{\beta}}\mathcal{A}\right)^T\right]\mathcal{A}\boldsymbol{a}_i = \\
&\frac{\sqrt{1-\beta}}{\beta}\left[\boldsymbol{I}_d - \left(\sqrt{\frac{1-\beta}{\beta}}\boldsymbol{A}\right)\beta\boldsymbol{I}_k\left(\sqrt{\frac{1-\beta}{\beta}}\mathcal{A}\right)^T\right]\mathcal{A}\boldsymbol{a}_i = \\
&\frac{\sqrt{1-\beta}}{\beta}\left[\boldsymbol{I}_d - (1-\beta)\boldsymbol{A}\boldsymbol{A}^T\right]\mathcal{A}\boldsymbol{a}_i = \\
&\frac{\sqrt{1-\beta}}{\beta}\left[\mathcal{A}\boldsymbol{a}_i - (1-\beta)\boldsymbol{A}\boldsymbol{a}_i\right] = \sqrt{1-\beta}\boldsymbol{A}\boldsymbol{a}_i.
\end{aligned}$$

将所有行堆叠在一起，我们得到最小均方误差问题（8）的最优解，即 $\boldsymbol{\theta}^* = \sqrt{1-\beta}\boldsymbol{A}\boldsymbol{A}^T$。该最优解刻画了反向随机微分方程（3）的高斯转移核。□

A.2 定理证明4

经过 r 次重采样步骤后，我们得到反向过程的以下输出：

$$\begin{aligned}
\overleftarrow{\boldsymbol{x}_0^r} = \\
\boldsymbol{A}\boldsymbol{A}^T\left(\boldsymbol{D}\left(\boldsymbol{m}\right)\overleftarrow{\boldsymbol{x}_0^{r-1}} + \boldsymbol{D}\left(\boldsymbol{1}-\boldsymbol{m}\right)\overrightarrow{\boldsymbol{x}_0}\right) = \\
\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}\left(\boldsymbol{m}\right)\overleftarrow{\boldsymbol{x}_0^{r-1}} + \boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}\left(\boldsymbol{1}-\boldsymbol{m}\right)\overrightarrow{\boldsymbol{x}_0} = \\
\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}\left(\boldsymbol{m}\right)\left(\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}\left(\boldsymbol{m}\right)\overleftarrow{\boldsymbol{x}_0^{r-2}} + \boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}\left(\boldsymbol{1}-\boldsymbol{m}\right)\overrightarrow{\boldsymbol{x}_0}\right) + \boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}\left(\boldsymbol{1}-\boldsymbol{m}\right)\overrightarrow{\boldsymbol{x}_0} = \\
\left(\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}\left(\boldsymbol{m}\right)\right)^r\overleftarrow{\boldsymbol{x}_0} + \left(\boldsymbol{I}+\sum_{i=1}^{r-1}\left(\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\right)^i\right)\left(\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{1}-\boldsymbol{m})\boldsymbol{A}\boldsymbol{z}_0\right). \qquad (10)
\end{aligned}$$

利用 Neumann 级数截断，式（10）可简化为

$$\begin{aligned}
\overleftarrow{\boldsymbol{x}_0^r} = \left(\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}\left(\boldsymbol{m}\right)\right)^r\overleftarrow{\boldsymbol{x}_0} + \left(\boldsymbol{I}-\left(\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\right)^r\right)\left(\boldsymbol{I}-\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\right)^{-1}\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{1}-\boldsymbol{m})\boldsymbol{A}\boldsymbol{z}_0 = \\
\left(\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}\left(\boldsymbol{m}\right)\right)^r\overleftarrow{\boldsymbol{x}_0} + \left(\boldsymbol{I}-\left(\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\right)^r\right)\left(\boldsymbol{I}-\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\right)^{-1}\boldsymbol{A}\boldsymbol{A}^T\left(\boldsymbol{I}-\boldsymbol{D}(\boldsymbol{m})\right)\boldsymbol{A}\boldsymbol{z}_0 = \\
\left(\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}\left(\boldsymbol{m}\right)\right)^r\overleftarrow{\boldsymbol{x}_0} + \left(\boldsymbol{I}-\left(\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\right)^r\right)\left(\boldsymbol{I}-\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\right)^{-1}\left(\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{z}_0-\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\boldsymbol{A}\boldsymbol{z}_0\right) = \\
\left(\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}\left(\boldsymbol{m}\right)\right)^r\overleftarrow{\boldsymbol{x}_0} + \left(\boldsymbol{I}-\left(\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\right)^r\right)\left(\boldsymbol{I}-\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\right)^{-1}\left(\boldsymbol{A}\boldsymbol{z}_0-\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\boldsymbol{A}\boldsymbol{z}_0\right) = \\
\left(\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}\left(\boldsymbol{m}\right)\right)^r\boldsymbol{A}\boldsymbol{A}^T\overleftarrow{\boldsymbol{x}_1} + \left(\boldsymbol{I}-\left(\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\right)^r\right)\left(\boldsymbol{I}-\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\right)^{-1}\left(\boldsymbol{I}-\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\right)\boldsymbol{A}\boldsymbol{z}_0 = \\
\left(\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}\left(\boldsymbol{m}\right)\right)^r\left(\boldsymbol{A}\boldsymbol{A}^T\overleftarrow{\boldsymbol{x}_1}-\boldsymbol{A}\boldsymbol{z}_0\right)+\boldsymbol{A}\boldsymbol{z}_0. \qquad (11)
\end{aligned}$$

接下来，我们对 $\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})$进行奇异值分解（SVD）。记 $\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})$ 的左奇异向量为 $\boldsymbol{U}\in\mathbb{R}^{d\times d}$，特征值构成的对角矩阵为 $\boldsymbol{\Sigma}\in\mathbb{R}^{d\times d}$，右奇异向量为 $\boldsymbol{V}\in\mathbb{R}^{d\times d}$。因此，式（11）可简化为

$$\overleftarrow{\boldsymbol{x}_0^r} = \boldsymbol{U}\boldsymbol{\Sigma}^r\boldsymbol{V}\left(\boldsymbol{A}\boldsymbol{A}^T\overleftarrow{\boldsymbol{x}_1}-\boldsymbol{A}\boldsymbol{z}_0\right)+\boldsymbol{A}\boldsymbol{z}_0.$$

由于 $\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})$ 的最大特征值严格小于1，我们得到以下结果：$\lim_{r\to\infty}\overleftarrow{\boldsymbol{x}_0^r}=\boldsymbol{A}\boldsymbol{z}_0$。这回答了基于扩散的图像修复可恢复性的问题。事实上，**算法2**在重采样步骤趋于无穷时，可恢复精确样本。对于大多数实际应用，得到一个 ϵ-近似解就足够了。现在，我们来推导**算法2** 收敛到 ϵ-邻域的速率 $\boldsymbol{A}\boldsymbol{z}_0$。

从等式（ $\overrightarrow{\boldsymbol{x}_0}=\boldsymbol{A}\boldsymbol{z}_0$ 的两边减去原始样本11），并取欧几里得范数，可得

$$\begin{aligned}
\left\|\overleftarrow{\boldsymbol{x}_0^r}-\overrightarrow{\boldsymbol{x}_0}\right\| = \\
\left\|\left(\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}\left(\boldsymbol{m}\right)\right)^r\left(\boldsymbol{A}\boldsymbol{A}^T\overleftarrow{\boldsymbol{x}_1}-\boldsymbol{A}\boldsymbol{z}_0\right)\right\| \overset{(i)}{\leq} \\
\left\|\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\right\|^r\left\|\boldsymbol{A}\boldsymbol{A}^T\overleftarrow{\boldsymbol{x}_1}-\boldsymbol{A}\boldsymbol{z}_0\right\| = \\
\left\|\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\right\|^r\left\|\boldsymbol{A}\boldsymbol{A}^T\overleftarrow{\boldsymbol{x}_1}-\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{z}_0\right\| \leq \\
\lambda_{\max}^r\frac{\|\boldsymbol{\theta}^*\|}{\sqrt{1-\beta}}\left\|\overleftarrow{\boldsymbol{x}_1}-\overrightarrow{\boldsymbol{x}_0}\right\|,
\end{aligned}$$

其中，(i) 用到了柯西-施瓦茨不等式和谱范数恒等式。令 $\lambda_{\max}^r \| \boldsymbol{\theta}^* \sqrt{1-\beta} \| \overleftarrow{\boldsymbol{x}_1} - \overrightarrow{\boldsymbol{x}_0} \| \leq \epsilon$，即可确保算法收敛到 ϵ-精确解。通过至少运行 $\log\left(\frac{\|\theta^*\| \| \overleftarrow{x_1} - \overrightarrow{x_0} \| \sqrt{1-\beta}}{\epsilon} \right) \| \log\left(\lambda^{-1}_{\max} \right)$ 重采样步骤，我们实现了收敛，其收敛速率为线性。□

A.3 定理6的证明

与**定理4**的证明类似，在含噪生成模型的情形下，我们从以下表达式开始:

$$
\begin{aligned}
\overleftarrow{\boldsymbol{x}_0^r} = \left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\left(\boldsymbol{D}\left(\boldsymbol{m}\right)\overleftarrow{\boldsymbol{x}_0^{r-1}} + \boldsymbol{D}\left(\boldsymbol{1}-\boldsymbol{m}\right)\overrightarrow{\boldsymbol{x}_0}\right) = \\
\left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}\left(\boldsymbol{m}\right)\overleftarrow{\boldsymbol{x}_0^{r-1}} + \left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}\left(\boldsymbol{1}-\boldsymbol{m}\right)\overrightarrow{\boldsymbol{x}_0} = \\
\left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}\left(\boldsymbol{m}\right)\left(\left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}\left(\boldsymbol{m}\right)\overleftarrow{\boldsymbol{x}_0^{r-2}} + \left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}\left(\boldsymbol{1}-\boldsymbol{m}\right)\overrightarrow{\boldsymbol{x}_0}\right) + \left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}\left(\boldsymbol{1}-\boldsymbol{m}\right)\overrightarrow{\boldsymbol{x}_0} = \\
\left(\left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}\left(\boldsymbol{m}\right)\right)^r\overleftarrow{\boldsymbol{x}_0} + \left(\sum_{i=0}^{r-1}\left(\left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}(\boldsymbol{m})\right)^i\right)\left(\left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}(\boldsymbol{1}-\boldsymbol{m})\boldsymbol{A}\boldsymbol{z}_0\right) = \\
\left(\left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}\left(\boldsymbol{m}\right)\right)^r\overleftarrow{\boldsymbol{x}_0} + \left(\boldsymbol{I} - \left(\left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}(\boldsymbol{m})\right)^r\right)\left(\boldsymbol{I} - \left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}(\boldsymbol{m})\right)^{-1} \\
\times\left(\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{z}_0 + \delta\boldsymbol{A}\boldsymbol{z}_0 - \left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}(\boldsymbol{m})\boldsymbol{A}\boldsymbol{z}_0\right) = \\
\left(\left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}\left(\boldsymbol{m}\right)\right)^r\overleftarrow{\boldsymbol{x}_0} + \left(\boldsymbol{I} - \left(\left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}(\boldsymbol{m})\right)^r\right)\left(\boldsymbol{I} - \left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}(\boldsymbol{m})\right)^{-1} \\
\times\left(\boldsymbol{A}\boldsymbol{z}_0 - \left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}(\boldsymbol{m})\boldsymbol{A}\boldsymbol{z}_0 + \delta\boldsymbol{A}\boldsymbol{z}_0\right) = \\
\left(\left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}\left(\boldsymbol{m}\right)\right)^r\overleftarrow{\boldsymbol{x}_0} + \left(\boldsymbol{I} - \left(\left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}(\boldsymbol{m})\right)^r\right)\left(\boldsymbol{I} - \left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}(\boldsymbol{m})\right)^{-1} \\
\times\left(\left(\boldsymbol{I} - \left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}(\boldsymbol{m})\right)\boldsymbol{A}\boldsymbol{z}_0 + \delta\boldsymbol{A}\boldsymbol{z}_0\right) = \\
\left(\left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}\left(\boldsymbol{m}\right)\right)^r\overleftarrow{\boldsymbol{x}_0} + \left(\boldsymbol{I} - \left(\left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}(\boldsymbol{m})\right)^r\right)\left(\boldsymbol{A}\boldsymbol{z}_0 + \left(\boldsymbol{I} - \left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}(\boldsymbol{m})\right)^{-1}\delta\boldsymbol{A}\boldsymbol{z}_0\right) = \\
\left(\left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}\left(\boldsymbol{m}\right)\right)^r\overleftarrow{\boldsymbol{x}_0} + \boldsymbol{A}\boldsymbol{z}_0 + \left(\boldsymbol{I} - \left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}(\boldsymbol{m})\right)^{-1}\delta\boldsymbol{A}\boldsymbol{z}_0 - \left(\left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}(\boldsymbol{m})\right)^r\boldsymbol{A}\boldsymbol{z}_0 \\
- \left(\left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}(\boldsymbol{m})\right)^r\left(\boldsymbol{I} - \left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}(\boldsymbol{m})\right)^{-1}\delta\boldsymbol{A}\boldsymbol{z}_0.
\end{aligned}
$$

从两边减去 $\overrightarrow{\boldsymbol{x}_0} = \boldsymbol{A}\boldsymbol{z}_0$ 对上述表达式两边取欧几里得范数，得到

$$
\begin{aligned}
\left\|\overleftarrow{\boldsymbol{x}_0^r} - \overrightarrow{\boldsymbol{x}_0}\right\| = \Bigg\|\left(\left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}\left(\boldsymbol{m}\right)\right)^r\overleftarrow{\boldsymbol{x}_0} + \left(\boldsymbol{I} - \left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}(\boldsymbol{m})\right)^{-1}\delta\boldsymbol{A}\boldsymbol{z}_0 - \left(\left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}(\boldsymbol{m})\right)^r\boldsymbol{A}\boldsymbol{z}_0 \\
- \left(\left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}(\boldsymbol{m})\right)^r\left(\boldsymbol{I} - \left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}(\boldsymbol{m})\right)^{-1}\delta\boldsymbol{A}\boldsymbol{z}_0\Bigg\|.
\end{aligned}
$$

重新排列各项后，我们先应用三角不等式，再应用柯西-施瓦茨不等式，得到

$$
\begin{aligned}
\left\|\overleftarrow{\boldsymbol{x}_0^r} - \overrightarrow{\boldsymbol{x}_0}\right\| \leq \\
\left\|\left(\left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}\left(\boldsymbol{m}\right)\right)^r\right\|\left\|\left(\overleftarrow{\boldsymbol{x}_0} - \boldsymbol{A}\boldsymbol{z}_0 - \left(\boldsymbol{I} - \left(\boldsymbol{A}\boldsymbol{A}^T + \delta\right)\boldsymbol{D}(\boldsymbol{m})\right)^{-1}\delta\boldsymbol{A}\boldsymbol{z}_0\right)\right\|
\end{aligned}
$$

$$+\left\|\left(\boldsymbol{I}-\left(\boldsymbol{A}\boldsymbol{A}^T+\delta\right)\boldsymbol{D}(\boldsymbol{m})\right)^{-1}\delta\boldsymbol{A}\boldsymbol{z}_0\right\| \leq$$
$$\left\|\left(\boldsymbol{A}\boldsymbol{A}^T+\delta\right)\boldsymbol{D}\left(\boldsymbol{m}\right)\right\|^r\left(\|\overleftarrow{\boldsymbol{x}_0}-\boldsymbol{A}\boldsymbol{z}_0\|+\left\|\left(\boldsymbol{I}-\left(\boldsymbol{A}\boldsymbol{A}^T+\delta\right)\boldsymbol{D}(\boldsymbol{m})\right)^{-1}\delta\boldsymbol{A}\boldsymbol{z}_0\right\|\right)$$
$$+\left\|\left(\boldsymbol{I}-\left(\boldsymbol{A}\boldsymbol{A}^T+\delta\right)\boldsymbol{D}(\boldsymbol{m})\right)^{-1}\delta\boldsymbol{A}\boldsymbol{z}_0\right\|.$$

由于 $\hat{}\,\lambda_{\max}$ 表示 $\left(\boldsymbol{A}\boldsymbol{A}^T+\delta\right)\boldsymbol{D}\left(\boldsymbol{m}\right)$ 的最大特征值，我们进一步简化为：

$$\left\|\overleftarrow{\boldsymbol{x}_0^r}-\overrightarrow{\boldsymbol{x}_0}\right\| \leq$$
$$\hat{\lambda}_{\max}^r\left(\|\overleftarrow{\boldsymbol{x}_0}-\boldsymbol{A}\boldsymbol{z}_0\|+\left\|\left(\boldsymbol{I}-\hat{\lambda}_{\max}\boldsymbol{I}\right)^{-1}\delta\boldsymbol{A}\boldsymbol{z}_0\right\|\right)+\left\|\left(\boldsymbol{I}-\hat{\lambda}_{\max}\boldsymbol{I}\right)^{-1}\delta\boldsymbol{A}\boldsymbol{z}_0\right\| =$$
$$\hat{\lambda}_{\max}^r\left(\|\overleftarrow{\boldsymbol{x}_0}-\boldsymbol{A}\boldsymbol{z}_0\|+\frac{\|\delta\boldsymbol{A}\boldsymbol{z}_0\|}{\left(1-\hat{\lambda}_{\max}\right)}\right)+\frac{\|\delta\boldsymbol{A}\boldsymbol{z}_0\|}{\left(1-\hat{\lambda}_{\max}\right)}.$$

利用 $\overleftarrow{\boldsymbol{x}_0}=\boldsymbol{A}\boldsymbol{A}^T\overleftarrow{\boldsymbol{x}_1}$、$\overrightarrow{\boldsymbol{x}_0}=\boldsymbol{A}\boldsymbol{z}_0$ 和 $\boldsymbol{A}^T\boldsymbol{A}=\boldsymbol{I}$ 这一事实，我们得到

$$\left\|\overleftarrow{\boldsymbol{x}_0^r}-\overrightarrow{\boldsymbol{x}_0}\right\| \leq \hat{\lambda}_{\max}^r\left(\left\|\boldsymbol{A}\boldsymbol{A}^T\overleftarrow{\boldsymbol{x}_1}-\boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})\boldsymbol{z}_0\right\|+\frac{\|\delta\boldsymbol{A}\boldsymbol{z}_0\|}{\left(1-\hat{\lambda}_{\max}\right)}\right)+\frac{\|\delta\boldsymbol{A}\boldsymbol{z}_0\|}{\left(1-\hat{\lambda}_{\max}\right)} \leq$$
$$\hat{\lambda}_{\max}^r\left(\left\|\boldsymbol{A}\boldsymbol{A}^T\right\|\|\overleftarrow{\boldsymbol{x}_1}-\boldsymbol{A}\boldsymbol{z}_0\|+\frac{\|\delta\|\,\|\boldsymbol{A}\boldsymbol{z}_0\|}{\left(1-\hat{\lambda}_{\max}\right)}\right)+\frac{\|\delta\|\,\|\boldsymbol{A}\boldsymbol{z}_0\|}{\left(1-\hat{\lambda}_{\max}\right)} =$$
$$\hat{\lambda}_{\max}^r\left(\frac{\|\boldsymbol{\theta}^*\|\,\|\overleftarrow{\boldsymbol{x}_1}-\overrightarrow{\boldsymbol{x}_0}\|}{\sqrt{1-\beta}}+\frac{\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|\,\|\overrightarrow{\boldsymbol{x}_0}\|}{\left(1-\hat{\lambda}_{\max}\right)\sqrt{1-\beta}}\right)+\frac{\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|\,\|\overrightarrow{\boldsymbol{x}_0}\|}{\left(1-\hat{\lambda}_{\max}\right)\sqrt{1-\beta}},$$

至此，定理的陈述得证。 □

A.4 推论7的证明

我们先陈述**定理6**：

$$\left\|\overleftarrow{\boldsymbol{x}_0^r}-\overrightarrow{\boldsymbol{x}_0}\right\| \leq \hat{\lambda}_{\max}^r\left(\frac{\|\boldsymbol{\theta}^*\|\,\|\overleftarrow{\boldsymbol{x}_1}-\overrightarrow{\boldsymbol{x}_0}\|}{\sqrt{1-\beta}}+\frac{\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|\,\|\overrightarrow{\boldsymbol{x}_0}\|}{\left(1-\hat{\lambda}_{\max}\right)\sqrt{1-\beta}}\right)+\frac{\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|\,\|\overrightarrow{\boldsymbol{x}_0}\|}{\left(1-\hat{\lambda}_{\max}\right)\sqrt{1-\beta}}.$$

由于 $\hat{}\,\lambda_{\max}<1$，第一项在极限中趋于零。利用 $q(\overrightarrow{\boldsymbol{x}_0})$ 具有紧支撑这一事实，我们得到

$$\begin{aligned}\sup_{\overrightarrow{\boldsymbol{x}_0}\sim q(\overrightarrow{\boldsymbol{x}_0})}\left\|\overleftarrow{\boldsymbol{x}_0^r}-\overrightarrow{\boldsymbol{x}_0}\right\| &\leq \sup_{\overrightarrow{\boldsymbol{x}_0}\sim q(\overrightarrow{\boldsymbol{x}_0})}\frac{\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|\,\|\overrightarrow{\boldsymbol{x}_0}\|}{\left(1-\hat{\lambda}_{\max}\right)\sqrt{1-\beta}}\\ &\leq \frac{\kappa\,\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|}{\left(1-\hat{\lambda}_{\max}\right)\sqrt{1-\beta}}\\ &= \frac{\kappa\delta}{\left(1-\hat{\lambda}_{\max}\right)}.\end{aligned}$$

将 $\delta=\mathcal{O}\left(\epsilon\left(1-\hat{}\,\lambda_{\max}\right)/\kappa\right)$ 替换为一个 δ 近似生成器，我们有 $\sup_{\overrightarrow{\boldsymbol{x}_0}\sim q(\overrightarrow{\boldsymbol{x}_0})}\left\|\overleftarrow{\boldsymbol{x}_0^r}-\overrightarrow{\boldsymbol{x}_0}\right\|\leq\epsilon$，从而完成证明。□

A.5 定理8 的证明

对于在 $T+1$ 状态上的扩散，最小化后验均值差异的训练目标为（6）：

$$\min_{\boldsymbol{\theta}} \mathop{\mathbb{E}}_{\substack{\overrightarrow{\boldsymbol{x}_0}\sim q(\overrightarrow{\boldsymbol{x}_0})\\ \epsilon\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ t\sim\mathcal{U}(1,\ldots,T)}} \left[\left\|\tilde{\mu}_t\left(\overrightarrow{\boldsymbol{x}_t},\overrightarrow{\boldsymbol{x}_0}\right)-\mu_\theta\left(\overleftarrow{\boldsymbol{x}_t},t\right)\right\|^2\right] := \left[\left\|\frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{a}_t}\overrightarrow{\boldsymbol{x}_0}+\frac{\sqrt{\alpha_t}\left(1-\bar{\alpha}_{t-1}\right)}{1-\bar{\alpha}_t}\overrightarrow{\boldsymbol{x}_t}-\mu_{\boldsymbol{\theta}}\left(\overleftarrow{\boldsymbol{x}_t},t\right)\right\|^2\right] =$$

$$\mathop{\mathbb{E}}_{\substack{\overrightarrow{\boldsymbol{x}_0}\sim q(\overrightarrow{\boldsymbol{x}_0})\\ \epsilon\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ t\sim\mathcal{U}(1,\ldots,T)}} \left[\left\|\frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{a}_t}\overrightarrow{\boldsymbol{x}_0}+\frac{\sqrt{\alpha_t}\left(1-\bar{\alpha}_{t-1}\right)}{1-\bar{\alpha}_t}\overrightarrow{\boldsymbol{x}_t}-\boldsymbol{\theta}\begin{bmatrix}\overleftarrow{\boldsymbol{x}_t}\\ t\end{bmatrix}\right\|^2\right].$$

根据前面的分析，我们知道上述最小均方误差问题的闭式解表示为:

$$\boldsymbol{\theta}_i^* = \left[\mathop{\mathbb{E}}_{\substack{\overrightarrow{\boldsymbol{x}_0}\sim q(\overrightarrow{\boldsymbol{x}_0})\\ \epsilon\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ t\sim\mathcal{U}(1,\ldots,T)}} \begin{bmatrix}\overleftarrow{\boldsymbol{x}_t}\\ t\end{bmatrix}\begin{bmatrix}\overleftarrow{\boldsymbol{x}_t}\\ t\end{bmatrix}^T\right]^{-1} \times \mathop{\mathbb{E}}_{\substack{\overrightarrow{\boldsymbol{x}_0}\sim q(\overrightarrow{\boldsymbol{x}_0})\\ \epsilon\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ t\sim\mathcal{U}(1,\ldots,T)}} \left[\left(\frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{a}_t}\overrightarrow{\boldsymbol{x}_0}+\frac{\sqrt{\alpha_t}\left(1-\bar{\alpha}_{t-1}\right)}{1-\bar{\alpha}_t}\overrightarrow{\boldsymbol{x}_t}\right)_i\begin{bmatrix}\overleftarrow{\boldsymbol{x}_t}\\ t\end{bmatrix}\right].$$

通过代入 在中间状态生成的样本的表达式，我们得到以下结果:

$$\boldsymbol{\theta}_i^* = \left[\mathop{\mathbb{E}}_{\substack{\overrightarrow{\boldsymbol{x}_0}\sim q(\overrightarrow{\boldsymbol{x}_0})\\ \epsilon\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ t\sim\mathcal{U}(1,\ldots,T)}} \begin{bmatrix}\overrightarrow{\boldsymbol{x}_0}\sqrt{\bar{\alpha}_t}+\epsilon\sqrt{1-\bar{\alpha}_t}\\ t\end{bmatrix}\begin{bmatrix}\overrightarrow{\boldsymbol{x}_0}\sqrt{\bar{\alpha}_t}+\epsilon\sqrt{1-\bar{\alpha}_t}\\ t\end{bmatrix}^T\right]^{-1}$$

$$\times \mathop{\mathbb{E}}_{\substack{\overrightarrow{\boldsymbol{x}_0}\sim q(\overrightarrow{\boldsymbol{x}_0})\\ \epsilon\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ t\sim\mathcal{U}(1,\ldots,T)}} \left[\left(\frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{a}_t}\overrightarrow{\boldsymbol{x}_0}+\frac{\sqrt{\alpha_t}\left(1-\bar{\alpha}_{t-1}\right)}{1-\bar{\alpha}_t}\left(\overrightarrow{\boldsymbol{x}_0}\sqrt{\bar{\alpha}_t}+\epsilon\sqrt{1-\bar{\alpha}_t}\right)\right)_i\begin{bmatrix}\overrightarrow{\boldsymbol{x}_0}\sqrt{\bar{\alpha}_t}+\epsilon\sqrt{1-\bar{\alpha}_t}\\ t\end{bmatrix}\right].$$

取外积并进一步展开，可得

$$\boldsymbol{\theta}_i^* = \left[\mathop{\mathbb{E}}_{\substack{\overrightarrow{\boldsymbol{x}_0}\sim q(\overrightarrow{\boldsymbol{x}_0})\\ \epsilon\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ t\sim\mathcal{U}(1,\ldots,T)}} \begin{bmatrix}\left(\overrightarrow{\boldsymbol{x}_0}\sqrt{\bar{\alpha}_t}+\epsilon\sqrt{1-\bar{\alpha}_t}\right)\left(\overrightarrow{\boldsymbol{x}_0}\sqrt{\bar{\alpha}_t}+\epsilon\sqrt{1-\bar{\alpha}_t}\right)^T & \left(\overrightarrow{\boldsymbol{x}_0}\sqrt{\bar{\alpha}_t}+\epsilon\sqrt{1-\bar{\alpha}_t}\right)t\\ t\left(\overrightarrow{\boldsymbol{x}_0}\sqrt{\bar{\alpha}_t}+\epsilon\sqrt{1-\bar{\alpha}_t}\right) & t^2\end{bmatrix}\right]^{-1}$$

$$\times \mathop{\mathbb{E}}_{\substack{\overrightarrow{\boldsymbol{x}_0}\sim q(\overrightarrow{\boldsymbol{x}_0})\\ \epsilon\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ t\sim\mathcal{U}(1,\ldots,T)}} \left[\left(\overrightarrow{\boldsymbol{x}_0}\left(\frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{a}_t}+\frac{\sqrt{\bar{\alpha}_t}\sqrt{\alpha_t}\left(1-\bar{\alpha}_{t-1}\right)}{1-\bar{\alpha}_t}\right)+\epsilon\left(\frac{\sqrt{\alpha_t}\left(1-\bar{\alpha}_{t-1}\right)\sqrt{1-\bar{\alpha}_t}}{1-\bar{\alpha}_t}\right)\right)_i\begin{bmatrix}\overrightarrow{\boldsymbol{x}_0}\sqrt{\bar{\alpha}_t}+\epsilon\sqrt{1-\bar{\alpha}_t}\\ t\end{bmatrix}\right],$$

重新整理后可得以下表达式:

$$\boldsymbol{\theta}_i^* = \left[\mathop{\mathbb{E}}_{\substack{\overrightarrow{\boldsymbol{x}_0}\sim q(\overrightarrow{\boldsymbol{x}_0})\\ \epsilon\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ t\sim\mathcal{U}(1,\ldots,T)}} \begin{bmatrix}\left(\overrightarrow{\boldsymbol{x}_0}\overrightarrow{\boldsymbol{x}_0}^T\bar{\alpha}_t+\overrightarrow{\boldsymbol{x}_0}\,\epsilon^T\sqrt{\bar{\alpha}_t}\sqrt{1-\bar{\alpha}_t}+\epsilon\,\overrightarrow{\boldsymbol{x}_0}^T\sqrt{\bar{\alpha}_t}\sqrt{1-\bar{\alpha}_t}+\epsilon\,\epsilon^T\left(1-\bar{\alpha}_t\right)\right) & \left(\overrightarrow{\boldsymbol{x}_0}\sqrt{\bar{\alpha}_t}+\epsilon\sqrt{1-\bar{\alpha}_t}\right)t\\ t\left(\overrightarrow{\boldsymbol{x}_0}\sqrt{\bar{\alpha}_t}+\epsilon\sqrt{1-\bar{\alpha}_t}\right) & t^2\end{bmatrix}\right]^{-1}$$

$$\times \underset{\substack{\overrightarrow{\boldsymbol{x}_0}\sim q(\overrightarrow{\boldsymbol{x}_0})\\ \epsilon\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ t\sim\mathcal{U}(1,\ldots,T)}}{\mathbb{E}} \left[\left(\overrightarrow{\boldsymbol{x}_0}\left(\frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{a}_t}+\frac{\sqrt{\bar{\alpha}_t}\sqrt{\alpha}_t\left(1-\bar{\alpha}_{t-1}\right)}{1-\bar{\alpha}_t}\right)+\epsilon\left(\frac{\sqrt{\alpha}_t\left(1-\bar{\alpha}_{t-1}\right)\sqrt{1-\bar{\alpha}_t}}{1-\bar{\alpha}_t}\right)\right)_i\left[\begin{matrix}\overrightarrow{\boldsymbol{x}_0}\sqrt{\bar{\alpha}_t}+\epsilon\sqrt{1-\bar{\alpha}_t}\\ t\end{matrix}\right]\right].$$

回想一下， $\overrightarrow{\boldsymbol{x}_0}$ 和 ϵ 的均值为零。现在，让我们利用这样一个事实： $\overrightarrow{\boldsymbol{x}_0}$、 ϵ和 t 都是相互独立的随机变量。

$$\boldsymbol{\theta}_i^*=\left[\underset{\substack{\boldsymbol{z}_0\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ \epsilon\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ t\sim\mathcal{U}(1,\ldots,T)}}{\mathbb{E}}\left[\begin{matrix}\left(\boldsymbol{A}\boldsymbol{z}_0\left(\boldsymbol{A}\boldsymbol{z}_0\right)^T\bar{\alpha}_t+\epsilon\,\epsilon^T\left(1-\bar{\alpha}_t\right)\right) & \mathbf{0}\\ \mathbf{0} & t^2\end{matrix}\right]\right]^{-1}$$
$$\times \underset{\substack{\overrightarrow{\boldsymbol{x}_0}\sim q(\overrightarrow{\boldsymbol{x}_0})\\ \epsilon\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ t\sim\mathcal{U}(1,\ldots,T)}}{\mathbb{E}} \left[\left(\overrightarrow{\boldsymbol{x}_0}\left(\frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{a}_t}+\frac{\sqrt{\bar{\alpha}_t}\sqrt{\alpha}_t\left(1-\bar{\alpha}_{t-1}\right)}{1-\bar{\alpha}_t}\right)+\epsilon\left(\frac{\sqrt{\alpha}_t\left(1-\bar{\alpha}_{t-1}\right)\sqrt{1-\bar{\alpha}_t}}{1-\bar{\alpha}_t}\right)\right)_i\left[\begin{matrix}\overrightarrow{\boldsymbol{x}_0}\sqrt{\bar{\alpha}_t}+\epsilon\sqrt{1-\bar{\alpha}_t}\\ t\end{matrix}\right]\right].$$

由于 $\alpha_t=1-\beta_t$ 并且 $\sqrt{\bar{\alpha}_t}=\sqrt{\bar{\alpha}_{t-1}}\sqrt{\alpha_t}$，我们进一步将闭式解简化如下：

$$\boldsymbol{\theta}_i^*=$$
$$\left[\underset{\substack{\boldsymbol{z}_0\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ \epsilon\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ t\sim\mathcal{U}(1,\ldots,T)}}{\mathbb{E}}\left[\begin{matrix}\left(\boldsymbol{A}\boldsymbol{z}_0\boldsymbol{z}_0^T\boldsymbol{A}^T\bar{\alpha}_t+\epsilon\,\epsilon^T\left(1-\bar{\alpha}_t\right)\right) & \mathbf{0}\\ \mathbf{0} & t^2\end{matrix}\right]\right]^{-1}$$
$$\times \underset{\substack{\overrightarrow{\boldsymbol{x}_0}\sim q(\overrightarrow{\boldsymbol{x}_0})\\ \epsilon\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ t\sim\mathcal{U}(1,\ldots,T)}}{\mathbb{E}} \left[\left(\overrightarrow{\boldsymbol{x}_0}\left(\frac{\sqrt{\bar{\alpha}_{t-1}}\left(\beta_t+\alpha_t\left(1-\bar{\alpha}_{t-1}\right)\right)}{1-\bar{a}_t}\right)+\epsilon\left(\frac{\sqrt{\alpha}_t\left(1-\bar{\alpha}_{t-1}\right)}{\sqrt{1-\bar{\alpha}_t}}\right)\right)_i\left[\begin{matrix}\overrightarrow{\boldsymbol{x}_0}\sqrt{\bar{\alpha}_t}+\epsilon\sqrt{1-\bar{\alpha}_t}\\ t\end{matrix}\right]\right]=$$
$$\left[\underset{\substack{\boldsymbol{z}_0\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ \epsilon\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ t\sim\mathcal{U}(1,\ldots,T)}}{\mathbb{E}}\left[\begin{matrix}\left(\boldsymbol{A}\boldsymbol{A}^T\bar{\alpha}_t+\epsilon\,\epsilon^T\left(1-\bar{\alpha}_t\right)\right) & \mathbf{0}\\ \mathbf{0} & t^2\end{matrix}\right]\right]^{-1}$$
$$\times \underset{\substack{\overrightarrow{\boldsymbol{x}_0}\sim q(\overrightarrow{\boldsymbol{x}_0})\\ \epsilon\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ t\sim\mathcal{U}(1,\ldots,T)}}{\mathbb{E}} \left[\left(\overrightarrow{\boldsymbol{x}_0}\left(\frac{\sqrt{\bar{\alpha}_{t-1}}\left(\beta_t+\alpha_t-\bar{\alpha}_t\right)}{1-\bar{a}_t}\right)+\epsilon\left(\frac{\sqrt{\alpha}_t\left(1-\bar{\alpha}_{t-1}\right)}{\sqrt{1-\bar{\alpha}_t}}\right)\right)_i\left[\begin{matrix}\overrightarrow{\boldsymbol{x}_0}\sqrt{\bar{\alpha}_t}+\epsilon\sqrt{1-\bar{\alpha}_t}\\ t\end{matrix}\right]\right]=$$
$$\underset{t\sim\mathcal{U}(1,\ldots,T)}{\mathbb{E}}\left[\begin{matrix}\left(\boldsymbol{A}\boldsymbol{A}^T\bar{\alpha}_t+\boldsymbol{I}\left(1-\bar{\alpha}_t\right)\right) & \mathbf{0}\\ \mathbf{0} & \frac{T^3-1}{3T}\end{matrix}\right]^{-1}$$
$$\times \underset{\substack{\overrightarrow{\boldsymbol{x}_0}\sim q(\overrightarrow{\boldsymbol{x}_0})\\ \epsilon\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ t\sim\mathcal{U}(1,\ldots,T)}}{\mathbb{E}} \left[\left(\overrightarrow{\boldsymbol{x}_0}\left(\frac{\sqrt{\bar{\alpha}_{t-1}}\left(\beta_t+1-\beta_t-\bar{\alpha}_t\right)}{1-\bar{a}_t}\right)+\epsilon\left(\frac{\sqrt{\alpha}_t\left(1-\bar{\alpha}_{t-1}\right)}{\sqrt{1-\bar{\alpha}_t}}\right)\right)_i\left[\begin{matrix}\overrightarrow{\boldsymbol{x}_0}\sqrt{\bar{\alpha}_t}+\epsilon\sqrt{1-\bar{\alpha}_t}\\ t\end{matrix}\right]\right]=$$
$$\underset{t\sim\mathcal{U}(1,\ldots,T)}{\mathbb{E}}\left[\begin{matrix}\left[\boldsymbol{A}\boldsymbol{A}^T\bar{\alpha}_t+\boldsymbol{I}\left(1-\bar{\alpha}_t\right)\right]^{-1} & \mathbf{0}\\ \mathbf{0} & \frac{3T}{T^3-1}\end{matrix}\right]$$

$$\times \mathop{\mathbb{E}}_{\substack{\overrightarrow{\boldsymbol{x}_0}\sim q(\overrightarrow{\boldsymbol{x}_0})\\ \epsilon\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ t\sim\mathcal{U}(1,\ldots,T)}} \left[\left(\overrightarrow{\boldsymbol{x}_0}\sqrt{\bar{\alpha}_{t-1}}+\epsilon\left(\frac{\sqrt{\alpha_t}\left(1-\bar{\alpha}_{t-1}\right)}{\sqrt{1-\bar{\alpha}_t}}\right)\right)_i \left[\begin{array}{c}\overrightarrow{\boldsymbol{x}_0}\sqrt{\bar{\alpha}_t}+\epsilon\sqrt{1-\bar{\alpha}_t}\\ t\end{array}\right]\right].$$

使用 Woodburry 矩阵求逆 [Hag89]代入上式可得

$$\boldsymbol{\theta}_i^* = \mathop{\mathbb{E}}_{t\sim\mathcal{U}(1,\ldots,T)} \left[\begin{array}{cc}(1-\bar{\alpha}_t)^{-1}\left[\boldsymbol{I}+\left(\boldsymbol{A}\sqrt{\frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)}}\right)\left(\boldsymbol{A}\sqrt{\frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)}}\right)^T\right]^{-1} & \mathbf{0}\\ \mathbf{0} & \frac{3T}{T^3-1}\end{array}\right]$$
$$\times \mathop{\mathbb{E}}_{\substack{\boldsymbol{z}_0\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ \epsilon\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ t\sim\mathcal{U}(1,\ldots,T)}} \left[\left(\boldsymbol{A}\boldsymbol{z}_0\sqrt{\bar{\alpha}_{t-1}}+\epsilon\left(\frac{\sqrt{\alpha_t}\left(1-\bar{\alpha}_{t-1}\right)}{\sqrt{1-\bar{\alpha}_t}}\right)\right)_i \left[\begin{array}{c}\boldsymbol{A}\boldsymbol{z}_0\sqrt{\bar{\alpha}_t}+\epsilon\sqrt{1-\bar{\alpha}_t}\\ t\end{array}\right]\right] =$$

$$\mathop{\mathbb{E}}_{t\sim\mathcal{U}(1,\ldots,T)} \left[\begin{array}{cc}(1-\bar{\alpha}_t)^{-1}\left[\boldsymbol{I}-\left(\boldsymbol{A}\sqrt{\frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)}}\right)\left(\boldsymbol{I}+\left(\boldsymbol{A}\sqrt{\frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)}}\right)^T\left(\boldsymbol{A}\sqrt{\frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)}}\right)\right)^{-1}\left(\boldsymbol{A}\sqrt{\frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)}}\right)^T\right]^{-1} & \mathbf{0}\\ \mathbf{0} & \frac{3T}{T^3-1}\end{array}\right]$$
$$\times \mathop{\mathbb{E}}_{\substack{\boldsymbol{z}_0\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ \epsilon\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ t\sim\mathcal{U}(1,\ldots,T)}} \left[\left(\boldsymbol{A}\boldsymbol{z}_0\sqrt{\bar{\alpha}_{t-1}}+\epsilon\left(\frac{\sqrt{\alpha_t}\left(1-\bar{\alpha}_{t-1}\right)}{\sqrt{1-\bar{\alpha}_t}}\right)\right)_i \left[\begin{array}{c}\boldsymbol{A}\boldsymbol{z}_0\sqrt{\bar{\alpha}_t}+\epsilon\sqrt{1-\bar{\alpha}_t}\\ t\end{array}\right]\right].$$

接下来，我们通过提取 i^{th}分量 $\left(\boldsymbol{A}\boldsymbol{z}_0\sqrt{\bar{\alpha}_{t-1}}+\epsilon\left(\frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{\sqrt{1-\bar{\alpha}_t}}\right)\right)_i$，进一步简化第二项:

$$\boldsymbol{\theta}_i^* =$$
$$\mathop{\mathbb{E}}_{t\sim\mathcal{U}(1,\ldots,T)} \left[\begin{array}{cc}(1-\bar{\alpha}_t)^{-1}\left[\boldsymbol{I}-\left(\boldsymbol{A}\sqrt{\frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)}}\right)\left(\boldsymbol{I}+\left(\boldsymbol{A}\sqrt{\frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)}}\right)^T\left(\boldsymbol{A}\sqrt{\frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)}}\right)\right)^{-1}\left(\boldsymbol{A}\sqrt{\frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)}}\right)^T\right]^{-1} & \mathbf{0}\\ \mathbf{0} & \frac{3T}{T^3-1}\end{array}\right]$$
$$\times \mathop{\mathbb{E}}_{\substack{\boldsymbol{z}_0\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ \epsilon\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ t\sim\mathcal{U}(1,\ldots,T)}} \left[\left(\boldsymbol{a}_i^T\boldsymbol{z}_0\sqrt{\bar{\alpha}_{t-1}}+\epsilon_i\left(\frac{\sqrt{\alpha_t}\left(1-\bar{\alpha}_{t-1}\right)}{\sqrt{1-\bar{\alpha}_t}}\right)\right)\left[\begin{array}{c}\boldsymbol{A}\boldsymbol{z}_0\sqrt{\bar{\alpha}_t}+\epsilon\sqrt{1-\bar{\alpha}_t}\\ t\end{array}\right]\right] =$$
$$\mathop{\mathbb{E}}_{t\sim\mathcal{U}(1,\ldots,T)} \left[\begin{array}{cc}(1-\bar{\alpha}_t)^{-1}\left[\boldsymbol{I}-\left(\boldsymbol{A}\sqrt{\frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)}}\right)\left(\boldsymbol{I}+\boldsymbol{I}\left(\frac{\bar{\alpha}_t}{1-\bar{\alpha}_t}\right)\right)^{-1}\left(\boldsymbol{A}\sqrt{\frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)}}\right)^T\right]^{-1} & \mathbf{0}\\ \mathbf{0} & \frac{3T}{T^3-1}\end{array}\right]$$
$$\times \mathop{\mathbb{E}}_{\substack{\boldsymbol{z}_0\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ \epsilon\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ t\sim\mathcal{U}(1,\ldots,T)}} \left[\begin{array}{c}\left(\boldsymbol{a}_i^T\boldsymbol{z}_0\sqrt{\bar{\alpha}_{t-1}}+\epsilon_i\left(\frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{\sqrt{1-\bar{\alpha}_t}}\right)\right)\left(\boldsymbol{A}\boldsymbol{z}_0\sqrt{\bar{\alpha}_t}+\epsilon\sqrt{1-\bar{\alpha}_t}\right)\\ \left(\boldsymbol{a}_i^T\boldsymbol{z}_0\sqrt{\bar{\alpha}_{t-1}}+\epsilon_i\left(\frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{\sqrt{1-\bar{\alpha}_t}}\right)\right)t\end{array}\right] =$$
$$\mathop{\mathbb{E}}_{t\sim\mathcal{U}(1,\ldots,T)} \left[\begin{array}{cc}(1-\bar{\alpha}_t)^{-1}\left[\boldsymbol{I}-\left(\boldsymbol{A}\sqrt{\frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)}}\right)\boldsymbol{I}\left(1-\bar{\alpha}_t\right)\left(\boldsymbol{A}\sqrt{\frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)}}\right)^T\right] & \mathbf{0}\\ \mathbf{0} & \frac{3T}{T^3-1}\end{array}\right]$$
$$\times \mathop{\mathbb{E}}_{\substack{\boldsymbol{z}_0\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ \epsilon\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ t\sim\mathcal{U}(1,\ldots,T)}} \left[\begin{array}{c}\left(\boldsymbol{A}\boldsymbol{z}_0(\boldsymbol{a}_i^T\boldsymbol{z}_0)\sqrt{\bar{\alpha}_{t-1}}\sqrt{\bar{\alpha}_t}+\epsilon_i\left(\frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{\sqrt{1-\bar{\alpha}_t}}\right)\epsilon\sqrt{1-\bar{\alpha}_t}\right)\\ 0\end{array}\right] =$$

$$\mathop{\mathbb{E}}_{t\sim\mathcal{U}(1,\ldots,T)} \left[\begin{array}{cc}(1-\bar{\alpha}_t)^{-1}\left[\boldsymbol{I}-\boldsymbol{A}\boldsymbol{A}^T\bar{\alpha}_t\right] & \mathbf{0}\\ \mathbf{0} & \frac{3T}{T^3-1}\end{array}\right]$$

$$\times \underset{\substack{\epsilon\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ t\sim\mathcal{U}(1,\ldots,T)}}{\mathbb{E}} \begin{bmatrix} \left(\boldsymbol{A}\boldsymbol{a}_i\bar{\alpha}_{t-1}\sqrt{\alpha_t} + \epsilon_i\left(\sqrt{\alpha}_t\left(1-\bar{\alpha}_{t-1}\right)\right)\epsilon\right) \\ 0 \end{bmatrix} =$$

$$\underset{t\sim\mathcal{U}(1,\ldots,T)}{\mathbb{E}} \begin{bmatrix} \frac{1}{(1-\bar{\alpha}_t)}\left[\boldsymbol{I} - \boldsymbol{A}\boldsymbol{A}^T\bar{\alpha}_t\right] & \mathbf{0} \\ \mathbf{0} & \frac{3T}{T^3-1} \end{bmatrix}$$

$$\times \underset{t\sim\mathcal{U}(1,\ldots,T)}{\mathbb{E}} \begin{bmatrix} \boldsymbol{A}\boldsymbol{a}_i\bar{\alpha}_{t-1}\sqrt{\alpha_t} + \boldsymbol{e}_i\left(\sqrt{\alpha}_t\left(1-\bar{\alpha}_{t-1}\right)\right) \\ 0 \end{bmatrix},$$

其中， $\boldsymbol{e}_i \in \{0,1\}^d$ 是一个独热编码向量，只有 i^{th} 坐标的值为 1。将期望移入并使用矩阵乘积，我们得到

$$\boldsymbol{\theta}_i^* = \begin{bmatrix} \mathbb{E}_t\left(\frac{1}{(1-\bar{\alpha}_t)}\right)\boldsymbol{I} - \mathbb{E}_t\left(\frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)}\right)\boldsymbol{A}\boldsymbol{A}^T & \mathbf{0} \\ \mathbf{0} & \frac{3T}{T^3-1} \end{bmatrix} \begin{bmatrix} \mathbb{E}_t\left(\bar{\alpha}_{t-1}\sqrt{\alpha_t}\right)\boldsymbol{A}\boldsymbol{a}_i + \mathbb{E}_t\left(\sqrt{\alpha}_t\left(1-\bar{\alpha}_{t-1}\right)\right)\boldsymbol{e}_i \\ 0 \end{bmatrix} =$$

$$\begin{bmatrix} \left(\mathbb{E}_t\left(\frac{1}{(1-\bar{\alpha}_t)}\right)\boldsymbol{I} - \mathbb{E}_t\left(\frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)}\right)\boldsymbol{A}\boldsymbol{A}^T\right)\left(\mathbb{E}_t\left(\bar{\alpha}_{t-1}\sqrt{\alpha_t}\right)\boldsymbol{A}\boldsymbol{a}_i + \mathbb{E}_t\left(\sqrt{\alpha}_t\left(1-\bar{\alpha}_{t-1}\right)\right)\boldsymbol{e}_i\right) \\ 0 \end{bmatrix} =$$

$$\begin{bmatrix} \mathbb{E}_t\left(\frac{1}{(1-\bar{\alpha}_t)}\right)\mathbb{E}_t\left(\bar{\alpha}_{t-1}\sqrt{\alpha_t}\right)\boldsymbol{A}\boldsymbol{a}_i - \mathbb{E}_t\left(\frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)}\right)\mathbb{E}_t\left(\bar{\alpha}_{t-1}\sqrt{\alpha_t}\right)\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{a}_i \\ +\mathbb{E}_t\left(\frac{1}{(1-\bar{\alpha}_t)}\right)\mathbb{E}_t\left(\sqrt{\alpha}_t\left(1-\bar{\alpha}_{t-1}\right)\right)\boldsymbol{e}_i - \mathbb{E}_t\left(\frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)}\right)\mathbb{E}_t\left(\sqrt{\alpha}_t\left(1-\bar{\alpha}_{t-1}\right)\right)\left(\boldsymbol{A}\boldsymbol{A}^T\right)\boldsymbol{e}_i \\ 0 \end{bmatrix} =$$

$$\begin{bmatrix} \left(\mathbb{E}_t\left(\frac{1}{(1-\bar{\alpha}_t)}\right)\mathbb{E}_t\left(\bar{\alpha}_{t-1}\sqrt{\alpha_t}\right) - \mathbb{E}_t\left(\frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)}\right)\mathbb{E}_t\left(\bar{\alpha}_{t-1}\sqrt{\alpha_t}\right)\right)\boldsymbol{A}\boldsymbol{a}_i \\ +\mathbb{E}_t\left(\frac{1}{(1-\bar{\alpha}_t)}\right)\mathbb{E}_t\left(\sqrt{\alpha}_t\left(1-\bar{\alpha}_{t-1}\right)\right)\boldsymbol{e}_i - \mathbb{E}_t\left(\frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)}\right)\mathbb{E}_t\left(\sqrt{\alpha}_t\left(1-\bar{\alpha}_{t-1}\right)\right)\left(\boldsymbol{A}\boldsymbol{A}^T\right)\boldsymbol{e}_i \\ 0 \end{bmatrix}.$$

将所有行叠加在一起，即可得到生成建模的闭式解:

$$\boldsymbol{\theta}^* = \begin{bmatrix} \left(\mathbb{E}_t\left(\frac{1}{(1-\bar{\alpha}_t)}\right)\mathbb{E}_t\left(\bar{\alpha}_{t-1}\sqrt{\alpha_t}\right) - \mathbb{E}_t\left(\frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)}\right)\mathbb{E}_t\left(\bar{\alpha}_{t-1}\sqrt{\alpha_t}\right)\right)\boldsymbol{A}\boldsymbol{A}^T & \\ +\mathbb{E}_t\left(\frac{1}{(1-\bar{\alpha}_t)}\right)\mathbb{E}_t\left(\sqrt{\alpha}_t\left(1-\bar{\alpha}_{t-1}\right)\right)\boldsymbol{I} & \\ -\mathbb{E}_t\left(\frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)}\right)\mathbb{E}_t\left(\sqrt{\alpha}_t\left(1-\bar{\alpha}_{t-1}\right)\right)\boldsymbol{A}\boldsymbol{A}^T & \mathbf{0} \end{bmatrix} =$$

$$\begin{bmatrix} \left(\mathbb{E}_t\left(\frac{1}{(1-\bar{\alpha}_t)}\right)\mathbb{E}_t\left(\bar{\alpha}_{t-1}\sqrt{\alpha_t}\right) - \mathbb{E}_t\left(\frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)}\right)\mathbb{E}_t\left(\bar{\alpha}_{t-1}\sqrt{\alpha_t}\right)\right)\boldsymbol{A}\boldsymbol{A}^T & \\ +\mathbb{E}_t\left(\frac{1}{(1-\bar{\alpha}_t)}\right)\mathbb{E}_t\left(\sqrt{\alpha}_t\left(1-\bar{\alpha}_{t-1}\right)\right)\boldsymbol{I} & \\ -\mathbb{E}_t\left(\frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)}\right)\mathbb{E}_t\left(\sqrt{\alpha}_t\right)\boldsymbol{A}\boldsymbol{A}^T + \mathbb{E}_t\left(\frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)}\right)\mathbb{E}_t\left(\sqrt{\alpha}_t\bar{\alpha}_{t-1}\right)\boldsymbol{A}\boldsymbol{A}^T & \mathbf{0} \end{bmatrix} =$$

$$\begin{bmatrix} \left(\mathbb{E}_t\left(\frac{1}{(1-\bar{\alpha}_t)}\right)\mathbb{E}_t\left(\bar{\alpha}_{t-1}\sqrt{\alpha_t}\right) - \mathbb{E}_t\left(\frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)}\right)\mathbb{E}_t\left(\sqrt{\alpha}_t\right)\right)\boldsymbol{A}\boldsymbol{A}^T & \\ +\mathbb{E}_t\left(\frac{1}{(1-\bar{\alpha}_t)}\right)\mathbb{E}_t\left(\sqrt{\alpha}_t\left(1-\bar{\alpha}_{t-1}\right)\right)\boldsymbol{I} & \\ & \mathbf{0} \end{bmatrix} =$$

$$\begin{bmatrix} \nu\boldsymbol{A}\boldsymbol{A}^T + \gamma\boldsymbol{I} & \mathbf{0} \end{bmatrix}.$$

至此，定理得证。 □

A.6 推论证明10

利用附录A.5中的论证，我们将最小均方误差问题（12）的闭式解表示为：

$$\boldsymbol{\theta}_i^* = \left[\underset{\substack{\overrightarrow{\boldsymbol{x}_0}\sim q(\overrightarrow{\boldsymbol{x}_0})\\ \epsilon\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ t\sim\mathcal{U}(1,\ldots,T)}}{\mathbb{E}} \begin{bmatrix}\overleftarrow{\boldsymbol{x}_t}\\ t\end{bmatrix}\begin{bmatrix}\overleftarrow{\boldsymbol{x}_t}\\ t\end{bmatrix}^T\right]^{-1} \times \underset{\substack{\overrightarrow{\boldsymbol{x}_0}\sim q(\overrightarrow{\boldsymbol{x}_0})\\ \epsilon\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ t\sim\mathcal{U}(1,\ldots,T)}}{\mathbb{E}} \left[\left(\frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{a}_t}\overrightarrow{\boldsymbol{x}_0} + \frac{\sqrt{\alpha_t}\left(1-\bar{\alpha}_{t-1}\right)}{1-\bar{\alpha}_t}\overrightarrow{\boldsymbol{x}_t}\right)_i \begin{bmatrix}\overleftarrow{\boldsymbol{x}_t}\\ t\end{bmatrix}\right] =$$

$$\underset{t\sim\mathcal{U}(1,\ldots,T)}{\mathbb{E}} \begin{bmatrix}\left[\boldsymbol{A}\boldsymbol{A}^T\bar{\alpha}_t + \boldsymbol{I}\left(1-\bar{\alpha}_t\right)\right]^{-1} & \mathbf{0}\\ \mathbf{0} & \frac{3T}{T^3-1}\end{bmatrix}$$

$$\times \underset{\substack{\overrightarrow{\boldsymbol{x}_0}\sim q(\overrightarrow{\boldsymbol{x}_0})\\ \epsilon\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ t\sim\mathcal{U}(1,\ldots,T)}}{\mathbb{E}} \left[\left(\overrightarrow{\boldsymbol{x}_0}\sqrt{\bar{\alpha}_{t-1}} + \overrightarrow{\epsilon}\left(\frac{\sqrt{\alpha_t}\left(1-\bar{\alpha}_{t-1}\right)}{\sqrt{1-\bar{\alpha}_t}}\right)\right)_i \begin{bmatrix}\overrightarrow{\boldsymbol{x}_0}\sqrt{\bar{\alpha}_t} + \overleftarrow{\epsilon}\sqrt{1-\bar{\alpha}_t}\\ t\end{bmatrix}\right] =$$

$$\underset{t\sim\mathcal{U}(1,\ldots,T)}{\mathbb{E}} \begin{bmatrix}\left(1-\bar{\alpha}_t\right)^{-1}\left[\boldsymbol{I} - \left(\boldsymbol{A}\sqrt{\frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)}}\right)\boldsymbol{I}\left(1-\bar{\alpha}_t\right)\left(\boldsymbol{A}\sqrt{\frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)}}\right)^T\right] & \mathbf{0}\\ \mathbf{0} & \frac{3T}{T^3-1}\end{bmatrix}$$

$$\times \underset{\substack{\boldsymbol{z}_0\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ \epsilon\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\\ t\sim\mathcal{U}(1,\ldots,T)}}{\mathbb{E}} \begin{bmatrix}\left(\boldsymbol{A}\boldsymbol{z}_0(\boldsymbol{a}_i^T\boldsymbol{z}_0)\sqrt{\bar{\alpha}_{t-1}}\sqrt{\bar{\alpha}_t} + \overrightarrow{\epsilon_i}\left(\frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{\sqrt{1-\bar{\alpha}_t}}\right)\overleftarrow{\epsilon}\sqrt{1-\bar{\alpha}_t}\right)\\ 0\end{bmatrix} =$$

$$\underset{t\sim\mathcal{U}(1,\ldots,T)}{\mathbb{E}} \begin{bmatrix}\left(1-\bar{\alpha}_t\right)^{-1}\left[\boldsymbol{I} - \boldsymbol{A}\boldsymbol{A}^T\bar{\alpha}_t\right] & \mathbf{0}\\ \mathbf{0} & \frac{3T}{T^3-1}\end{bmatrix} \underset{t\sim\mathcal{U}(1,\ldots,T)}{\mathbb{E}} \begin{bmatrix}\boldsymbol{A}\boldsymbol{a}_i\bar{\alpha}_{t-1}\sqrt{\alpha_t}\\ 0\end{bmatrix} =$$

$$\begin{bmatrix}\mathbb{E}_t\left(\frac{1}{(1-\bar{\alpha}_t)}\right)\boldsymbol{I} - \mathbb{E}_t\left(\frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)}\right)\boldsymbol{A}\boldsymbol{A}^T & \mathbf{0}\\ \mathbf{0} & \frac{3T}{T^3-1}\end{bmatrix}\begin{bmatrix}\mathbb{E}_t\left(\bar{\alpha}_{t-1}\sqrt{\alpha_t}\right)\boldsymbol{A}\boldsymbol{a}_i\\ 0\end{bmatrix} =$$

$$\begin{bmatrix}\left(\mathbb{E}_t\left(\frac{1}{(1-\bar{\alpha}_t)}\right)\mathbb{E}_t\left(\bar{\alpha}_{t-1}\sqrt{\alpha_t}\right) - \mathbb{E}_t\left(\frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)}\right)\mathbb{E}_t\left(\bar{\alpha}_{t-1}\sqrt{\alpha_t}\right)\right)\boldsymbol{A}\boldsymbol{a}_i\\ 0\end{bmatrix}$$

现在，如附录 A.5 中那样将所有行堆叠在一起，我们得到 DDPM 生成器的闭式解 $\boldsymbol{\theta}^* = \begin{bmatrix}\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T & \mathbf{0}\end{bmatrix}$，推论的证明至此完成。□

A.7 定理证明11

为了证明 DDPM 学得到了分布 $q\left(\overrightarrow{\boldsymbol{x}_0}\right)$，只需证明 $\overleftarrow{\boldsymbol{x}_0} \sim p\left(\overleftarrow{\boldsymbol{x}_0}\right)$ 具有如下结构 $\boldsymbol{A}\overleftarrow{\boldsymbol{z}_0}$，其中 $\overleftarrow{\boldsymbol{z}_0} \sim$ 正态分布 $(\mathbf{0}, \boldsymbol{I})$。利用 **推论 10** 中的闭式解，我们从 $\overleftarrow{\boldsymbol{x}_T} \sim$ 正态分布 $(\mathbf{0}, \boldsymbol{I})$ 开始执行一步反向过程。然后，我们校正反向过程中的漂移和弥散错位。对所有状态递归应用反向步骤后，我们得到

$$\overleftarrow{\boldsymbol{x}_0} = \left(\prod_{t=1}^{T}\left(\omega_t\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\right)\right)\overleftarrow{\boldsymbol{x}_T} + \left(\prod_{t=1}^{T-1}\left(\omega_t\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\right)\right)\xi_T\sqrt{\beta_T}\overleftarrow{\epsilon_T} + \left(\prod_{t=1}^{T-2}\left(\omega_t\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\right)\right)\xi_T\sqrt{\beta_{T-1}}\overleftarrow{\epsilon_{T-1}}$$
$$+ \cdots + \left(\omega_1\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\right)\left(\omega_2\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\right)\xi_3\sqrt{\beta_3}\overleftarrow{\epsilon_3} + \left(\omega_1\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\right)\xi_2\sqrt{\beta_2}\overleftarrow{\epsilon_2}.$$

由于$\{\overleftarrow{\epsilon_t}\}_{t>1}$ 是独立同分布高斯随机变量，我们将上述表达式改写为：

$$\overleftarrow{\boldsymbol{x}_0} = \left(\prod_{t=1}^{T}\left(\omega_t\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\right)\right)\overleftarrow{\boldsymbol{x}_T} + \bar{\Sigma}^{1/2}\overleftarrow{\epsilon},$$

其中 $\overleftarrow{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$，$\bar{\Sigma}$ 表示包含$\{\overleftarrow{\epsilon_t}\}_{t>1}$的各项的方差之和。回顾一下，$\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{I}$、漂移系数 $\omega_t = \omega = \frac{1}{\bar{\nu}\sqrt{2}^{1/T}}$和弥散 $\xi_t = \sqrt{2\beta_t \frac{2(t-1)/T}{T-1}()}$。现在，我们来计算各项的方差：

$$\mathbb{E}\left[\left(\left(\omega_1\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\right)\xi_2\sqrt{\beta_2}\overleftarrow{\epsilon_2}\right)\left(\left(\omega_1\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\right)\xi_2\sqrt{\beta_2}\overleftarrow{\epsilon_2}\right)^T\right] = \omega_1^2\bar{\nu}^2\xi_2^2\beta_2\boldsymbol{A}\boldsymbol{A}^T\mathbb{E}\left[\overleftarrow{\epsilon_2}\overleftarrow{\epsilon_2}^T\right]\boldsymbol{A}\boldsymbol{A}^T = \omega_1^2\bar{\nu}^2\xi_2^2\beta_2\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{A}^T$$
$$= \omega_1^2\bar{\nu}^2\xi_2^2\beta_2\boldsymbol{A}\boldsymbol{A}^T = \left(\frac{1}{\bar{\nu}\sqrt{2}^{1/T}}\right)^2\bar{\nu}^2\left(\frac{2^{1/T}}{2\beta_2(T-1)}\right)\beta_2\boldsymbol{A}\boldsymbol{A}^T = \frac{\boldsymbol{A}\boldsymbol{A}^T}{2(T-1)};$$
$$\mathbb{E}\left[\left(\left(\omega_1\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\right)\left(\omega_2\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\right)\xi_3\sqrt{\beta_3}\overleftarrow{\epsilon_3}\right)\left(\left(\omega_1\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\right)\left(\omega_2\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\right)\xi_3\sqrt{\beta_3}\overleftarrow{\epsilon_3}\right)^T\right] = \frac{\boldsymbol{A}\boldsymbol{A}^T}{2(T-1)};\cdots;$$
$$\mathbb{E}\left[\left(\left(\prod_{t=1}^{T-1}\left(\omega_t\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\right)\right)\xi_T\sqrt{\beta_T}\overleftarrow{\epsilon_T}\right)\left(\left(\prod_{t=1}^{T-1}\left(\omega_t\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\right)\right)\xi_T\sqrt{\beta_T}\overleftarrow{\epsilon_T}\right)^T\right] = \frac{\boldsymbol{A}\boldsymbol{A}^T}{2(T-1)}.$$

注意，$\left(\boldsymbol{A}\boldsymbol{A}^T\right)^2 := \boldsymbol{A}\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{A}\boldsymbol{A}^T$。由于共有（$T-1$）项，总协方差为 $\bar{\Sigma} = \frac{\boldsymbol{A}\boldsymbol{A}^T}{2}$，这等价于反向过程结束时生成的以下样本：

$$\overleftarrow{\boldsymbol{x}_0} = \left(\prod_{t=1}^{T}\left(\omega_t\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\right)\right)\overleftarrow{\boldsymbol{x}_T} + \frac{\boldsymbol{A}\boldsymbol{A}^T}{\sqrt{2}}\overleftarrow{\epsilon} = \boldsymbol{A}\boldsymbol{A}^T\left(\prod_{t=1}^{T}\left(\frac{1}{\bar{\nu}\sqrt{2}^{1/T}}\right)\bar{\nu}\right)\overleftarrow{\boldsymbol{x}_T} + \frac{\boldsymbol{A}\boldsymbol{A}^T}{\sqrt{2}}\overleftarrow{\epsilon}$$
$$= \frac{\boldsymbol{A}\boldsymbol{A}^T}{\sqrt{2}}\overleftarrow{\boldsymbol{x}_T} + \frac{\boldsymbol{A}\boldsymbol{A}^T}{\sqrt{2}}\overleftarrow{\epsilon} = \boldsymbol{A}\left(\frac{\boldsymbol{A}^T\overleftarrow{\boldsymbol{x}_T}}{\sqrt{2}} + \frac{\boldsymbol{A}^T\overleftarrow{\epsilon}}{\sqrt{2}}\right) = \boldsymbol{A}\overleftarrow{\boldsymbol{z}_0}.$$

现在，我们来验证 $\overleftarrow{\boldsymbol{x}_0}$ 是否位于数据流形上。为此，只需证明 $\overleftarrow{\boldsymbol{z}_0}$ 的均值为零且协方差为单位协方差。因此，我们计算

$$\mathbb{E}\left[\overleftarrow{\boldsymbol{z}_0}\right] = \mathbb{E}\left[\frac{\boldsymbol{A}^T\overleftarrow{\boldsymbol{x}_T}}{\sqrt{2}} + \frac{\boldsymbol{A}^T\overleftarrow{\epsilon}}{\sqrt{2}}\right] = \frac{\boldsymbol{A}^T\mathbb{E}\left[\overleftarrow{\boldsymbol{x}_T}\right]}{\sqrt{2}} + \frac{\boldsymbol{A}^T\mathbb{E}\left[\overleftarrow{\epsilon}\right]}{\sqrt{2}} = \mathbf{0}$$

and

$$\mathbb{E}\left[\overleftarrow{\boldsymbol{z}_0}\overleftarrow{\boldsymbol{z}_0}^T\right] = \mathbb{E}\left[\left(\frac{\boldsymbol{A}^T\overleftarrow{\boldsymbol{x}_T}}{\sqrt{2}} + \frac{\boldsymbol{A}^T\overleftarrow{\epsilon}}{\sqrt{2}}\right)\left(\frac{\boldsymbol{A}^T\overleftarrow{\boldsymbol{x}_T}}{\sqrt{2}} + \frac{\boldsymbol{A}^T\overleftarrow{\epsilon}}{\sqrt{2}}\right)^T\right] \overset{(i)}{=} \frac{\boldsymbol{A}^T\mathbb{E}\left[\overleftarrow{\boldsymbol{x}_T}\overleftarrow{\boldsymbol{x}_T}^T\right]\boldsymbol{A}}{2} + \frac{\boldsymbol{A}^T\mathbb{E}\left[\overleftarrow{\epsilon}\overleftarrow{\epsilon}^T\right]\boldsymbol{A}}{2} \overset{(ii)}{=} \boldsymbol{I},$$

其中，(i) 利用了 $\overleftarrow{\boldsymbol{x}_T}$ 和 $\overleftarrow{\epsilon}$ 是相互独立且均值为零的随机变量这一事实；(ii) 则依据 $\overleftarrow{\boldsymbol{x}_T}$、$\overleftarrow{\epsilon}$ 具有单位协方差，以及 $\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{I}$ 这一性质。因此，证明完成。□

A.8 定理12的证明

除了在每个中间状态执行图像修复步骤外，证明还沿用了附录 A.7中的论证。从高斯先验 $\overleftarrow{\boldsymbol{x}_T} \sim$ 正态分布 $(\mathbf{0}, \boldsymbol{I})$开始，我们首先计算

$$\overleftarrow{\boldsymbol{x}_{T-1}} = \omega_T\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\overleftarrow{\boldsymbol{x}_T} + \xi_T\sqrt{\beta_T}\overleftarrow{\epsilon_T}.$$

在下一步中，我们运行正向随机微分方程，以计算给定图像中已知部分在噪声水平为（$T-1$）时的结果：

$$\overrightarrow{\boldsymbol{x}_{T-1}} = \sqrt{\bar{\alpha}_{T-1}}\overrightarrow{\boldsymbol{x}_0} + \sqrt{1-\bar{\beta}_{T-1}}\overrightarrow{\epsilon_T}.$$

然后，我们通过将已知部分与图像中生成的未知部分相结合来更新 $\overleftarrow{\boldsymbol{x}_{T-1}}$：

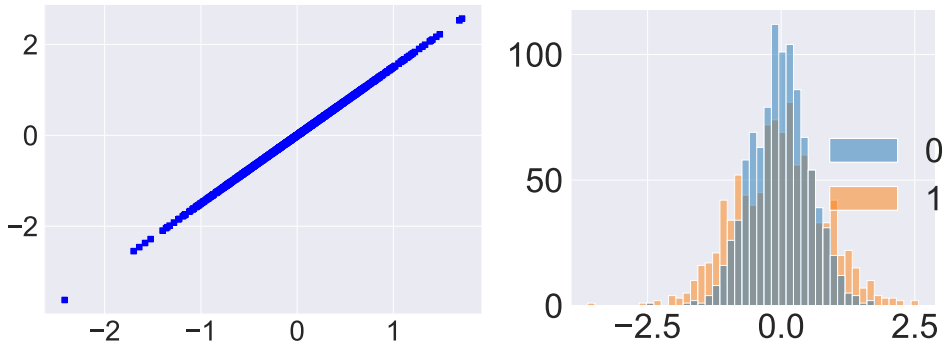
$$\overleftarrow{\boldsymbol{x}_{T-1}} = \boldsymbol{D}(\boldsymbol{m})\ \overleftarrow{\boldsymbol{x}_{T-1}} + \boldsymbol{D}(1-\boldsymbol{m})\ \overrightarrow{\boldsymbol{x}_{T-1}}$$

至此，反向过程的状态（$T-1$）处的图像修复完成。对所有状态递归应用这一过程，我们得到

$$\begin{aligned}
\overleftarrow{\boldsymbol{x}_0} = &\left(\prod_{t=T-1}^{1} \omega_{T-t}\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\right)\left(\omega_T\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\right)\overleftarrow{\boldsymbol{x}_T} + \\
&\left(\prod_{t=T-1}^{2} \omega_{T-t}\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\right)\left(\omega_{T-1}\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{1}-\boldsymbol{m})\sqrt{\bar{\alpha}_{T-1}}\overrightarrow{\boldsymbol{x}_0}\right) + \\
&\left(\prod_{t=T-1}^{3} \omega_{T-t}\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\right)\left(\omega_{T-2}\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{1}-\boldsymbol{m})\sqrt{\bar{\alpha}_{T-2}}\overrightarrow{\boldsymbol{x}_0}\right) + \cdots + \\
&\left(\omega_1\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\sqrt{\bar{\alpha}_1}\overrightarrow{\boldsymbol{x}_0}\right)\left(\omega_2\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{1}-\boldsymbol{m})\sqrt{\bar{\alpha}_2}\overrightarrow{\boldsymbol{x}_0}\right) + \left(\omega_1\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{1}-\boldsymbol{m})\sqrt{\bar{\alpha}_1}\overrightarrow{\boldsymbol{x}_0}\right) + \\
&\left(\prod_{t=T-1}^{2} \omega_{T-t}\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\right)\left(\omega_{T-1}\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{1}-\boldsymbol{m})\sqrt{1-\bar{\alpha}_{T-1}}\overrightarrow{\epsilon_{T-1}}\right) + \\
&\left(\prod_{t=T-1}^{3} \omega_{T-t}\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\right)\left(\omega_{T-2}\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{1}-\boldsymbol{m})\sqrt{1-\bar{\alpha}_{T-2}}\overrightarrow{\epsilon_{T-2}}\right) + \cdots + \\
&\left(\omega_1\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\sqrt{1-\bar{\alpha}_1}\overrightarrow{\epsilon_1}\right)\left(\omega_2\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{1}-\boldsymbol{m})\sqrt{1-\bar{\alpha}_2}\overrightarrow{\epsilon_2}\right) + \\
&\left(\omega_1\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{1}-\boldsymbol{m})\sqrt{1-\bar{\alpha}_1}\overrightarrow{\epsilon_1}\right) + \\
&\left(\prod_{t=T-1}^{1} \omega_{T-t}\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\right)\xi_T\sqrt{\beta_T}\overleftarrow{\epsilon_T} + \\
&\left(\prod_{t=T-1}^{2} \omega_{T-t}\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\right)\xi_{T-1}\sqrt{\beta_{T-1}}\overleftarrow{\epsilon_{T-1}} + \\
&\left(\omega_1\bar{\nu}\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\right)\xi_2\sqrt{\beta_2}\overleftarrow{\epsilon_2}.
\end{aligned}$$

现在，让我们化简涉及 $\overleftarrow{\boldsymbol{x}_T}$ 的第一项。代入漂移系数，即 $\omega_t = \omega = \frac{1}{\bar{\nu}\sqrt{2}^{1/T}}$，可化简为 $\frac{1}{\sqrt{2}}\left(\prod_{T-1}^{1}\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\right)\boldsymbol{A}\boldsymbol{A}^T\overleftarrow{\boldsymbol{x}_T}$。类似地，将涉及 $\overrightarrow{\boldsymbol{x}_0}$ 的项合并后，得到 $\left(\sum_{t=1}^{T-1}\left(\frac{1}{\sqrt{2}}\right)^{t/T}\sqrt{\bar{\alpha}_t}\left(\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\right)^{t-1}\right)\times\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{1}-\boldsymbol{m})\overrightarrow{\boldsymbol{x}_0}$。现在，我们采用如下假设：$\{\overleftarrow{\epsilon_t}\}_{t=2}^{T}$ 和 $\{\overleftarrow{\epsilon_t}\}_{t=1}^{T-1}$ 是独立同分布高斯随机变量。我们将前向过程中包含 $\overrightarrow{\epsilon_t}$ 的离散项的合并方差记为 $\overrightarrow{\Sigma}$，将反向过程中包含 $\overleftarrow{\epsilon_t}$ 的项的合并方差记为 Σ。因此，在每个中间状态进行图像修复后，反向过程结束时生成的最终样本为：

$$\begin{aligned}
\overleftarrow{\boldsymbol{x}_0} = \frac{1}{\sqrt{2}}\left(\prod_{T-1}^{1}\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\right)\boldsymbol{A}\boldsymbol{A}^T\overleftarrow{\boldsymbol{x}_T} + &\left(\sum_{t=1}^{T-1}\left(\frac{1}{\sqrt{2}}\right)^{t/T}\sqrt{\bar{\alpha}_t}\left(\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{m})\right)^{t-1}\right) \\
&\times\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{D}(\boldsymbol{1}-\boldsymbol{m})\overrightarrow{\boldsymbol{x}_0} + \overrightarrow{\Sigma}\overrightarrow{\epsilon} + \overleftarrow{\Sigma}\overleftarrow{\epsilon},
\end{aligned}$$



(a) 数据流形　　(b) 边缘分布

图3：数据生成分布支撑在线性流形上，其边缘分布为高斯边缘分布。

其中，$\overleftarrow{\epsilon}$ 和 $\overrightarrow{\epsilon}$ 是独立同分布高斯变量，服从 正态分布 $(\mathbf{0},\ \boldsymbol{I})$。我们可以将 $\overleftarrow{\epsilon}$ 和 $\overrightarrow{\epsilon}$ 合并，表示为单个高斯随机变量。我们将它们分开写，以明确突出各自的贡献。证明至此结束。□

B 实验

本节提供实验证据，以支持我们的理论结果。由于本文旨在提供理论依据，我们只进行一些玩具实验，以验证主要的理论论断。我们考虑一种可以实现完美恢复的情形。所谓完美恢复，是指均方根误差（RMSE）小到可以忽略不计，如下所示 10^{-5}。关于使用基于扩散的图像修复进行的大规模实验，请参阅 Lugmayr、Danelljan、Romero、Yu、Timofte 和 Van Gool 最近发表的工作 [LDRYTV22]。

B.1 实现细节

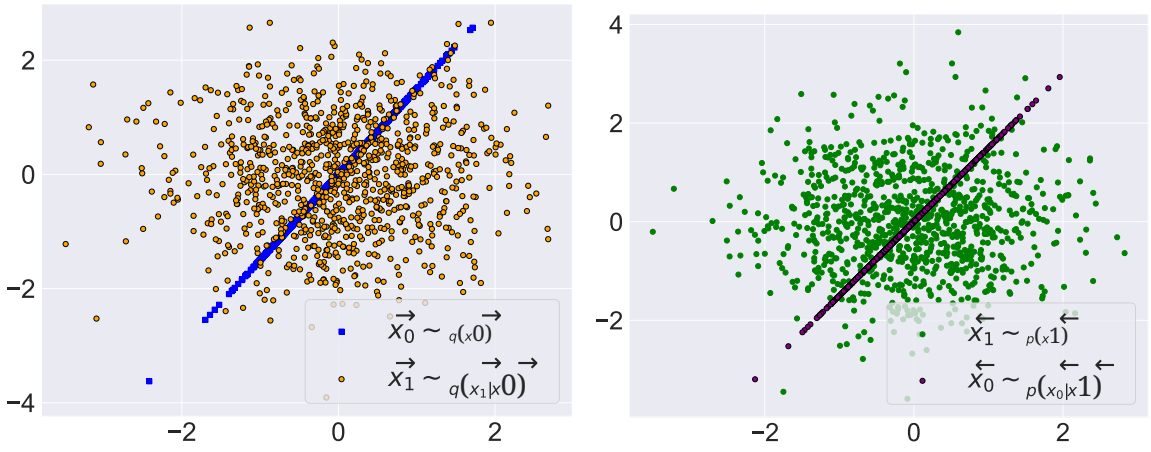
如图3a所示，数据生成分布 $q\left(\overrightarrow{\boldsymbol{x}_0}\right)$ 支撑于 $\mathbb{R}^2$中的一个线性流形上，即 $k=1$ 和 $d=2$。我们选择 $\boldsymbol{A}=\begin{bmatrix}2/\sqrt{13}\\3/\sqrt{13}\end{bmatrix}$ 和 $\boldsymbol{m}=\begin{bmatrix}0\\1\end{bmatrix}$，满足假设 1 和~~2~~。样本 $\overrightarrow{\boldsymbol{x}_0}\sim q(\overrightarrow{\boldsymbol{x}_0})$遵循以下结构：$\overrightarrow{\boldsymbol{x}_0}=\boldsymbol{A}z_0$，其中 $z_0\sim$ 服从 N (0, 1)。在图3b中，$\overrightarrow{\boldsymbol{x}_0}$ 第一和第二个坐标的边缘分布分别以蓝色和橙色条形表示。我们考虑$n=1000$ 个样本、方差 $\beta=0.9$和 $R=100$ 个重采样步骤。

B.2 实验结果

图4a 显示了 $n=1000$ 从数据生成分布中抽取的样本 $q\left(\overrightarrow{\boldsymbol{x}_0}\right)$ 以及正向随机微分方程 $\overrightarrow{\boldsymbol{x}_1}$的输出。这验证了正向随机微分方程收敛到参考先验。图4b 展示了 $n=1000$ 从高斯先验中抽取的样本 $p\left(\overleftarrow{\boldsymbol{x}_1}\right)$，以及通过反向随机微分方程成功投影到数据流形上的样本。在图2中，我们比较了基线算法 RePaint（●）和所提算法 RePaint$^+$ （$*$）恢复的信号。原始样本（■）与 RePaint 修复的样本$^+$ 之间的距离为 ϵ，均方根误差为 3.28×10^{-8}，如表1。另一方面　　另一方面，RePaint 修复样本的均方根误差为 22.97。正如讨论的那样

表1：图像修复的均方根误差。

方法	$d=2, k=1$
RePaint	22.97
RePaint+反向随机微分方程	27.61
RePaint$^+$	3.28×10^{-8}



(a) 正向随机微分方程　　(b) 反向随机微分方程

图 4：DDPM 中正向过程和反向过程的可视化。正向随机微分方程将从数据流形中抽取的样本推向高斯先验的支撑集。反向随机微分方程则发现数据分布之下的子空间。

在我们的理论分析中（§3.2），错位导致的偏差会迫使 RePaint 偏离数据流形。

有人可能会问，将这些修复样本再经过一个反向随机微分方程步骤，是否会实现完美恢复。RePaint 的这一修改设置会将生成的样本 $\overleftarrow{\boldsymbol{x}_0}$ 投影到数据流形上，如图 5中的红色箭头所示。然而，这并不会将 $\overleftarrow{\boldsymbol{x}_0}$ 推入 $\overrightarrow{\boldsymbol{x}_0}$ 周围的微小子空间，因为它们的均方根误差仍为27.61，比原始设置更差。此外，从图 5 可以清楚看出，即使是向数据流形进行最小二乘投影，也无法恢复原始样本。因此，原始 RePaint 算法中存在一种非平凡偏差，会对完美恢复产生不利影响。与之相反，RePaint$^+$ 消除了这种偏差，并且在重采样步骤无限多时能够实现完美恢复。步骤数有限时，它会收敛到原始样本的一个极小 ϵ邻域。这进一步印证了我们从论文正文分析中得出的算法见解。

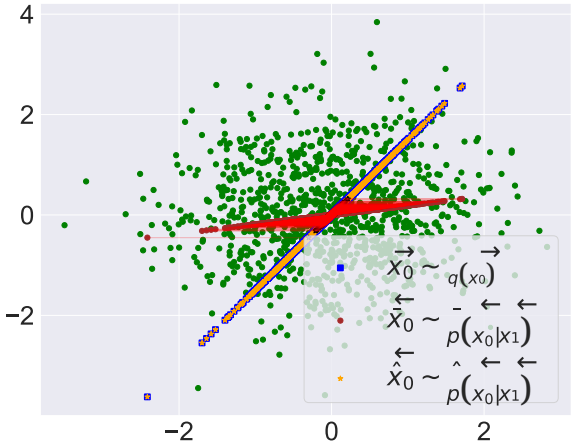


图 5：在 RePaint 生成的修复样本上运行反向随机微分方程。红色箭头标出了最后的反向步骤（→）。该方法未能恢复真实底层样本。