Towards a unified framework for guided diffusion models

Yuchen Jiao* Yuxin Chen† Gen Li*

December 4, 2025

Abstract

Guided or controlled data generation with diffusion models has become a cornerstone of modern generative modeling. Despite substantial advances in diffusion model theory, the theoretical understanding of guided diffusion samplers remains severely limited. We make progress by developing a unified algorithmic and theoretical framework that accommodates both diffusion guidance and reward-guided diffusion. Aimed at fine-tuning diffusion models to improve certain rewards, we propose injecting a reward guidance term — constructed from the difference between the original and reward-reweighted scores — into the backward diffusion process, and rigorously quantify the resulting reward improvement over the unguided counterpart. As a key application, our framework shows that classifier-free guidance (CFG) decreases the expected reciprocal of the classifier probability, providing the first theoretical characterization of the specific performance metric that CFG improves for general target distributions. When applied to reward-guided diffusion, our framework yields a new sampler that is easy-to-train and requires no full diffusion trajectories during training. Numerical experiments further corroborate our theoretical findings.

Keywords: diffusion models, diffusion guidance, classifier-free guidance, reward-guided diffusion models

Contents

*Department of Statistics, the Chinese University of Hong Kong, Hong Kong.
†Department of Statistics and Data Science, the Wharton School, University of Pennsylvania.
Partial results of this work appeared in International Conference on Machine Learning 2025 (Li and Jiao, 2025).

1 Introduction

Score-based diffusion models have rapidly become a cornerstone of modern generative artificial intelligence, setting new state-of-the-art benchmarks across diverse domains such as image and video generation, medical imaging, genomics, among others (Sohl-Dickstein et al., 2015; Song and Ermon, 2019; Ho et al., 2020; Song et al., 2021b,a; Croitoru et al., 2023; Ramesh et al., 2022; Rombach et al., 2022; Shang et al., 2025; Saharia et al., 2022). In a nutshell, a diffusion model begins with a forward process that progressively corrupts samples from a target data distribution p_{data} in $\mathbb{R}^d$ into Gaussian noise, then learns to reverse this process, effectively transforming a standard Gaussian distribution back into the target data distribution. A key element to reverse the forward process is the estimation of the (Stein) score function, defined as the gradient of the log-likelihood of noisy data at corresponding noise levels. Informally speaking, if the forward process comprises N steps as follows:

$$X_0 \sim p_{\mathsf{data}}, \qquad X_0 \overset{\text{add noise}}{\longrightarrow} X_1 \overset{\text{add noise}}{\longrightarrow} \cdots \overset{\text{add noise}}{\longrightarrow} X_N, \tag{1}$$

then a common strategy to (approximately) reverse it involves iterative updates of the form:

$$Y_{n-1} = \text{linear-term}(Y_n) + \text{score}(Y_n) + \text{noise}, \qquad \text{for } n = N, \ldots, 1, \tag{2}$$

with score$(\cdot)$ indicating a suitable score function. Crucially, the score functions offer key information underlying the dynamics of the forward process, steering the trajectory $\{Y_n\}$ towards the target data distribution. When properly designed, each Y_n in (2) follows a distribution closely approximating that of X_n, as exemplified by the Denoising Diffusion Probabilistic Model (DDPM) (Ho et al., 2020).

1.1 Guided or controlled generation with diffusion models

As diffusion models gain widespread adoption in practice, there is a growing demand for controllable data generation, wherein the models can be guided to produce samples that align with specific objectives or user preferences. Two prominent paradigms for controllable generation are diffusion guidance (Dhariwal and Nichol, 2021; Ho and Salimans, 2021) and reward-guided diffusion models (Fan et al., 2023; Gao et al., 2024; Black et al., 2023; Huh and Mohapatra, 2025; Yuan et al., 2023; Zhao et al., 2024; Uehara et al., 2024a), which constitute the primary focus of the current paper.

Diffusion guidance. An important application of controlled generation is class-conditional sampling, in which the goal is to generate samples belonging to a specified class or category. For instance, a diffusion model trained on animal images may be instructed to produce images exclusively of cats or dogs based on the target class label, rather than arbitrary animals. A straightforward approach to achieving this is to replace the score function in (2) with the class-conditional score function (i.e., the gradient of the class-conditional log-likelihood in the forward process). Despite its conceptual simplicity, however, this standard approach often yields suboptimal sample quality in practice, for reasons that are not yet fully understood (Ho and Salimans, 2021; Rombach et al., 2022).

An alternative approach is diffusion guidance, originally proposed by Dhariwal and Nichol (2021) as *classifier guidance*, which augments the class-conditional version of the update rule (2) with a guidance term derived from the classifier probability (i.e., the classifier's predicted probability of a sample belonging to a target class) as follows:

$$Y_{n-1} = \text{linear-term}(Y_n) + \text{(conditional)-score}(Y_n) + \text{guidance}(Y_n) + \text{noise}, \qquad \text{for } n = N, \ldots, 1. \tag{3}$$

The intuition is to amplify the classifier's influence and steer sampling toward the target class. To alleviate the need for a separately trained classifier, Ho and Salimans (2021) proposed *classifier-free guidance (CFG)*, which replaces the classifier-based guidance term with the difference between conditional and unconditional scores, derived via the Bayes rule. CFG has since emerged as the prevailing paradigm for class-conditional diffusion models, achieving superior perceptual quality in practice. Nevertheless, the underlying working mechanism of diffusion guidance remains largely mysterious (Wu et al., 2024a; Bradley and Nakkiran, 2024; Chidambaram et al., 2024). Since incorporating guidance terms inevitably causes the generative process to deviate from the true conditional distribution, it leaves open fundamental theoretical questions such as what distribution CFG actually samples from, and which performance metrics it is expected to improve.

Reward-guided diffusion models. Another paradigm for controlled generation, which we term reward-guided (or reward-directed) diffusion models, seeks to generate samples that achieve higher values under a specified external reward function. Such a reward function may capture diverse objectives, such as alignment with human preferences, adherence to stylistic or compositional constraints, or performance on downstream tasks (e.g., image-text alignment). To fine-tune diffusion models with respect to a given reward, several approaches have been proposed, including but not limited to reinforcement learning (Black et al., 2023; Fan et al., 2023), direct preference optimization (Rafailov et al., 2023; Wallace et al., 2024), and supervised fine-tuning (Ziegler et al., 2019). These methods enable flexible, goal-directed control of data generation beyond traditional conditional sampling mechanisms. The theoretical underpinnings of existing approaches, however, remain largely elusive.

Given the conceptual similarity between the formulation of diffusion guidance and reward-guided diffusion models — both of which steer the sampling trajectory to enhance specific aspects of generation — a natural question arises: can diffusion guidance paradigms inspire new, general, and theoretically principled algorithms for reward-guided diffusion? For instance, For instance, a sampling strategy analogous to CFG may take the following form:

$$Y_{n-1} = \text{linear-term}(Y_n) + \text{score}(Y_n) + \text{reward-guidance}(Y_n) + \text{noise}, \qquad \text{for } n = N, \ldots, 1 \tag{4}$$

for some judiciously chosen reward guidance term. A deeper theoretical understanding about the inner workings of CFG could therefore illuminate the design and anaylsis of such methods, a direction we aim to pursue in this paper.

1.2 Main contributions: a unified algorithmic and theoretical framework

As discussed above, both diffusion guidance and reward-guided diffusion models share the common objective of enhancing specific aspects of the generated samples by fine-tuning pre-trained diffusion models. Motivated by this connection, the present paper develops a unified algorithmic and theoretical framework that accommodates and generalizes these two paradigms, while providing rigorous theoretical guarantees for the effectiveness of the proposed framework.

Our unified framework. In order to adapt diffusion models to improve a desired objective (e.g., increasing classifier probability as in diffusion guidance, and enhancing an external reward as in reward-guided diffusion), we propose to inject carefully designed guidance terms — constructed based on the difference of the original and adjusted scores — into the backward process of diffusion models. Within this framework, we rigorously quantify the improvement of the objective function relative to its unguided counterpart in the continuous-time limit. When instantiated for diffusion guidance and reward-guided diffusion, our unified framework yields the following key contributions.

- *Classifier-free diffusion guidance.* Our unified framework unveils that, for a broad class of target data distributions, CFG increases the average reciprocal of the classifier probability — an insight that clarifies, for the first time, the specific performance metric that CFG provably improves. This metric bears close connections to the Inception Score (IS) (Salimans et al., 2016), a widely used measure of sample quality that depends directly on classifier probabilities. In contrast, prior theoretical analyses of CFG were confined to restricted settings, such as mixtures of compactly supported distributions or isotropic Gaussian models (Wu et al., 2024a; Chidambaram et al., 2024; Bradley and Nakkiran, 2024). Our results extend beyond these special cases, providing the first theoretical guarantees for CFG under general data distribution.
- *Reward-guided diffusion models.* Our framework also leads to a practically effective approach for reward-guided diffusion models. The proposed algorithm utilizes the difference between reward-reweighted and original scores as a guidance term, eliminating the need of retraining when the guidance scale varies — unlike prior methods that often require retraining for each choice of the regularization parameter. In addition, the training procedure resembles denoising score matching combined with importance sampling; in each training iteration, only a single noise scale is used for gradient computation, which sidesteps the need to simulate the entire diffusion trajectory and hence reduces training overhead. Finally, we theoretically quantify the improvement in the expected reward, confirming the effectiveness of our proposed approach.

We conduct a series of numerical experiments to empirically validate our theoretical findings. Our analysis is further extended to more realistic settings that incorporate time discretization and imperfect score estimation. We prove that the corresponding discrete-time sampler can well approximate their continuous-time counterparts, thereby confirming the applicability and stability of our approach in practice.

Before proceeding, we note that preliminary results on CFG were presented in an early version of this paper (Li and Jiao, 2025). The present paper substantially extends those results by developing a unified framework that encompasses both CFG and reward-guided diffusion, and by proposing a new, general, and theoretically principled sampler for the reward-guided setting.

1.3 Organization

The remainder of this paper is organized as follows. Section 2 provides an overview of diffusion models, diffusion guidance, and reward-guided diffusion models, along with their continuous-time limits. Section 3 presents our unified framework and theoretical guarantees for the continuous-time samplers, and instantiates them to both classifier-free guidance and reward-guided diffusion. Detailed proofs are deferred to Section 6 and the appendices, whereas discussion about time discretization and score estimation errors is postponed to Section D. Section 4 reports our numerical experiments, whereas Section 5 discusses several additional prior work. We conclude this paper in Section 7.

2 Background

This section begins by reviewing the basics of diffusion models, followed by a brief introduction to diffusion guidance and reward-guided diffusion models. Here and throughout, we shall use the discrete variable $n = 1, 2, \cdots$ to index discrete-time steps, and the continuous variable $t \in [0, 1]$ to index continuous time.

2.1 Diffusion models

Diffusion generative models are composed of two stochastic processes: a *forward process*, which gradually transforms data into pure noise, and a *backward process*, which reverses this transformation to generate samples that approximate the target distribution.

Forward process. Concretely, denote by p_{data} the target data distribution in $\mathbb{R}^d$. Starting from an initial sample $X_0 \in \mathbb{R}^d$ drawn from p_{data}, the forward process evolves according to the following discrete-time dynamics:

$$X_0 \sim p_{\mathsf{data}}, \tag{5a}$$

$$X_n = \sqrt{1-\beta_n} X_{n-1} + \sqrt{\beta_n} W_n, \qquad n = 1, \cdots, N, \tag{5b}$$

where N denotes the total number of steps, $0 < \beta_1, \ldots, \beta_N < 1$ are coefficients, and $\{W_n\}_{1 \le n \le N} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_d)$ is a sequence of independent noise vectors drawn from standard Gaussian distributions. By defining

$$\overline{\alpha}_n = \prod_{k=1}^{n} (1-\beta_k), \tag{6}$$

we can easily see that each X_n can be expressed, for some random vector $W_n' \sim \mathcal{N}(0, I_d)$, as

$$X_n = \sqrt{\overline{\alpha}_n} X_0 + \sqrt{1-\overline{\alpha}_n} W_n'. \tag{7}$$

If $\overline{\alpha}_N \to 0$ as N grows, then the distribution of X_N converges to that of pure Gaussian noise.

Backward process. We now turn attention to the backward process, which is constructed by learning to reverse the forward process (5). In doing so, one can start from a standard Gaussian noise vector and iteratively transform it into a sample that approximately follows the target distribution p_{data}. A key ingredient that enables the reversal of the forward process is the so-called (Stein) score function, defined as the gradient of the log-density of each intermediate random vector X_n in (5):

$$s_n^\star(x) \coloneqq \nabla \log p_{X_n}(x), \qquad 1 \le n \le N. \tag{8}$$

Armed with estimates $\{s_n(\cdot)\}$ for this set of score functions, one can (approximately) reverse the forward process via, for instance, the following iterative update rule:

$$Y_N \sim \mathcal{N}(0, I_d), \tag{9a}$$

$$Y_{n-1} = \frac{1}{\sqrt{1-\beta_n}} \big(Y_n + \beta_n s_n(Y_n)\big) + \sqrt{\beta_n} Z_n, \qquad n = N, \ldots, 2, \tag{9b}$$

where $Z_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_d)$ is another sequence of Gaussian noise vectors independent of those used in the forward process (5), and Y_1 denotes the generated sample. This procedure (9) is known as the Denoising Diffusion Probabilistic Model (DDPM), one of the most widely used diffusion-based samplers (Ho et al., 2020).

Continuous-time limits through the lens of SDE. A widely adopted framework for understanding the working mechanism of diffusion models is through the lens of stochastic differential equations (SDEs). Specifically, the discrete-time forward process (5), when taken to the continuous-time limit, is intimately linked with the following SDE

$$\mathrm{d}X_t = -\frac{1}{2(1-t)} X_t \mathrm{d}t + \frac{1}{\sqrt{1-t}} \mathrm{d}B_t^{\mathsf{f}} \qquad \text{for } 0 \le t \le 1, \qquad \text{with } X_0 \sim p_{\mathsf{data}}, \tag{10a}$$

where B_t^{f} denotes a standard Brownian motion in $\mathbb{R}^d$. As $t \to 1$, the process specified by (10) converges to a standard Gaussian distribution. Classical SDE literature (Anderson, 1982; Haussmann and Pardoux, 1986) has established that the forward SDE (10a) can be reversed through another SDE as follows:

$$\mathrm{d}Y_t = \Big(\frac{1}{2} Y_t + \nabla \log p_{X_{1-t}}(Y_t)\Big) \frac{\mathrm{d}t}{t} + \frac{1}{\sqrt{t}} \mathrm{d}B_t, \quad \text{for } 0 \le t \le 1, \tag{10b}$$

where B_t is the standard Brownian motion. To formalize this connection, we state below a standard result concerning the distributional equivalence between the reverse and original SDE (see Song et al. (2021b)).

Lemma 1. *Consider any $0 \le \delta < 1$. For any τ and t obeying $0 \le \tau \le t \le 1-\delta$, one has*

$$X_t \,|\, X_\tau \sim \mathcal{N}\left(\sqrt{\frac{1-t}{1-\tau}} X_\tau, \frac{t-\tau}{1-\tau} I_d\right). \tag{11}$$

Moreover, if $Y_\delta \sim p_{X_{1-\delta}}$, then it holds that

$$Y_t \overset{\mathrm{d}}{=} X_{1-t} \quad \text{for all } t \in [\delta, 1]. \tag{12}$$

In words, if Y_δ is initialized according to the exact distribution of $X_{1-\delta}$, then the remaining trajectory of Y_t shares the same distribution as that of the corresponding forward process. Notably, the DDPM (9) can be interpreted as a suitable time discretization of the reverse SDE (10b) (see, e.g., Song et al. (2021b); Huang et al. (2024)).

2.2 Conditional sampling and diffusion guidance

Conditional diffusion models. Given a class label c, conditional diffusion models aim to generate samples from the conditional distribution $p_{X_0 \,|\, c}(\cdot \,|\, c)$ (also denoted by $p_{\mathsf{data} \,|\, c}(\cdot \,|\, c)$ throughout), where the class label c can encode, say, categorical information that specifies which part of the data distribution one should sample from. In contrast to unconditional diffusion models described in Section 2.1, class-conditional sampling targets more controllable data generation by guiding the diffusion process toward samples from a given class. Perhaps the most natural strategy is to replace the estimate $s_n(\cdot)$ of the unconditional score function $s_n^\star(\cdot)$ with an estimate $s_n(\cdot \,|\, c)$ of the following conditional score function

$$s_n^\star(\cdot \,|\, c) := \nabla \log p_{X_n \,|\, c}(\cdot \,|\, c), \qquad 1 \le n \le N, \tag{13}$$

thus resulting in the following sampling procedure:

$$Y_N \sim \mathcal{N}(0, I_d), \tag{14a}$$

$$Y_{n-1} = \frac{1}{\sqrt{1-\beta_n}}\big(Y_n + \beta_n s_n(Y_n \,|\, c)\big) + \sqrt{\beta_n} Z_n, \qquad n = N, \ldots, 2. \tag{14b}$$

As before, the Z_n's are independently drawn from $\mathcal{N}(0, I_d)$.

Similar to the unconditional case in (10a), one can introduce the class-conditional forward process in continuous time as follows:

$$\mathrm{d}X_t = -\frac{1}{2(1-t)} X_t \mathrm{d}t + \frac{1}{\sqrt{1-t}} \mathrm{d}B_t^{\mathsf{f}} \qquad \text{for } 0 \le t \le 1, \qquad \text{with } X_0 \sim p_{\mathsf{data} \,|\, c} \tag{15a}$$

where $p_{\mathsf{data} \,|\, c}$ stands for the target data distribution conditioned on class c. The corresponding reverse process is obtained by replacing the unconditional score function in (10b) with the conditional score function:

$$Y_0 \sim \mathcal{N}(0, I_d), \tag{15b}$$

$$\mathrm{d}Y_t = \left(\frac{1}{2} Y_t + \nabla \log p_{X_{1-t} \,|\, c}(Y_t \,|\, c)\right) \frac{\mathrm{d}t}{t} + \frac{1}{\sqrt{t}} \mathrm{d}B_t, \quad \text{for } 0 \le t \le 1, \tag{15c}$$

which can be regarded as the continuous-time limit of the discrete-time process (14).

Diffusion guidance: classifier guidance and classifier-free guidance. To enhance sampling performance in practice — particularly in terms of perceptual quality — Dhariwal and Nichol (2021) introduced the notion of "guidance," augmenting the conditional score in (14) with an additive term involving the classifer probability $p_{c \,|\, X_0}(c \,|\, \cdot)$. More precisely, this *classifier guidance* approach yields the following sampling procedure:

$$Y_{n-1}^w = \frac{1}{\sqrt{1-\beta_n}}\Big(Y_n^w + \beta_n\big(s_n(Y_n^w \,|\, c) + w \nabla \log p_{c \,|\, X_n}(c \,|\, Y_n^w)\big)\Big) + \sqrt{\beta_n} Z_n, \quad n = N, \cdots, 2, \tag{16}$$

where $w > 0$ is the guidance scale that modulates the influence of the classifier probability term during sampling. Here, we use the superscript w in Y_n^w to emphasize its dependence on the chosen guidance scale.

Motivated by the practical challenges in implementing classifier guidance (e.g., the substantial computational cost of training classifier models separately), Ho and Salimans (2021) proposed an alternative approach known as "classifier-free guidance (CFG)," which eliminates the need for separate classifier training. Leveraging the basic relation $\nabla \log p_{c\,|\,X_n}(c\,|\,x) = s_n^\star(x\,|\,c) - s_n^\star(x)$ (i.e., the Bayes rule), CFG replaces the guidance term in (16) with the score difference $s_n(x\,|\,c) - s_n(x)$, resulting in

$$Y_N^w \sim \mathcal{N}(0, I_d), \tag{17a}$$

$$Y_{n-1}^w = \frac{1}{\sqrt{1-\beta_n}}\Big(Y_n^w + \beta_n\big((1+w)s_n(Y_n^w\,|\,c) - w s_n(Y_n^w)\big)\Big) + \sqrt{\beta_n} Z_n, \quad n = N, \cdots, 2. \tag{17b}$$

Importantly, the unconditional score function $s_n^\star(\cdot)$ and the class-conditional score function $s_n^\star(\cdot\,|\,c)$ can be jointly learned using a single neural network (Ho and Salimans, 2021).

Further, by extending the aforementioned reverse SDE (15c) to incorporate the guidance term as in (17), we arrive at the following SDE $\{Y_t^w\}$:

$$\mathrm{d}Y_t^w = \Big(\frac{1}{2}Y_t^w + (1+w)\nabla \log p_{X_{1-t}\,|\,c}(Y_t^w\,|\,c) - w\nabla \log p_{X_{1-t}}(Y_t^w)\Big)\frac{\mathrm{d}t}{t} + \frac{1}{\sqrt{t}}\mathrm{d}B_t, \quad \text{for } 0 \le t \le 1, \tag{18}$$

where we again include $w > 0$ in the superscript of $\{Y_t^w\}$ to explicitly indicate the strength of guidance used to construct this SDE.

2.3 Reward-guided diffusion models

Moving beyond conditional diffusion models, another recent line of work has explored reward-guided (or reward-directed) diffusion models (Fan et al., 2023; Black et al., 2023; Uehara et al., 2024b; Gao et al., 2024; Zhao et al., 2024; Clark et al., 2023; Yuan et al., 2023; Huh and Mohapatra, 2025; Jiao et al., 2025a; Keramati et al., 2025), which leverage task-specific reward functions to fine-tune pre-trained diffusion models. For the most part, this reward-guided paradigm begins by learning a reward function that captures human preferences for a given task — often based on downstream objectives such as aesthetic quality or drug effectiveness. The pretrained diffusion model is then fine-tuned (or retrained) to enhance the expected reward of its generated samples, while maintaining proximity to the original sampler.

Specifically, suppose we are given an external reward function $r^{\mathsf{ext}} : \mathbb{R}^d \to \mathbb{R}$ that encodes preferences over generated samples for a specific task, along with a pre-trained (unguided) diffusion model $\mathcal{D}$. The goal is to

$$\text{improve } \mathbb{E}\big[r^{\mathsf{ext}}(Y^{\mathsf{sample}})\big] \quad \text{by fine-tuning the sampler } \mathcal{D}. \tag{19}$$

Here, we adopt the term "improve" rather than "maximize," because the reward function typically serves as guidance for steering the generative process rather than an objective to optimize in isolation. In practice, it is also important to prevent the adjusted sampler from deviating too far from the original diffusion model, which could otherwise lead to reward over-optimization and hence degraded sample quality.

Several approaches have been proposed in prior literature to achieve this goal (19). For instance, supposing that the backward process proceeds in discrete time as $Y_N \to Y_{N-1} \to \cdots \to Y_1$, one common strategy employs, say, reinforcement learning techniques, to solve optimization problems of the form (Fan et al., 2023; Black et al., 2023)

$$\text{maximize}_\theta \quad \mathbb{E}_{\prod_{n=N}^2 p_\theta(Y_{n-1}\,|\,Y_n) p_{Y_N}(Y_N)}[r^{\mathsf{ext}}(Y_1)] + \gamma \mathsf{regularizer}\big(p_\theta(\cdot)\big), \tag{20}$$

where we parameterize the generative distribution p_θ via the parameter θ, γ denotes a regularization parameter, and $\mathsf{regularizer}(\cdot)$ denotes a regularization term (e.g., KL divergence) that forces p_θ to be reasonably close to the original diffusion model sampler. A key challenge arises from the fact that the reward function r^{ext} depends on the final sample Y_1, whose distribution depends on the entire denoising trajectory $Y_N, \ldots, Y_2$. Consequently, computing the gradient of the objective function in (20) w.r.t. θ during training requires sampling full trajectories, substantially increasing complexity and posing practical implementation hurdles. In addition, adjusting the regularization parameter γ typically necessitates full retraining, which reduces the flexibility of this approach. In this work, we propose an alternative route that directly modifies the backward diffusion process, to be detailed momentarily.

3 Main results

In this section, we present a unified framework — from the lens of continuous-time limits and SDEs — that accommodates both classifier-free guidance and reward-guided diffusion models. We then develop theoretical analysis to elucidate the effectiveness of these techniques, under fairly general assumptions on the target data distribution. Discussion about the effect of time discretization and score estimation error is postponed to Section D.

3.1 A unified framework: algorithm and theory

Recall that both diffusion guidance and reward-guided diffusion models adapt the original pretrained diffusion models to achieve specific objectives. For instance, diffusion guidance was originally introduced to enhance the classifier probability (Dhariwal and Nichol, 2021), whereas reward-guided diffusion seeks to improve certain external rewards evaluated on the generated samples. Building on this perspective, we consider the following unified objective that integrates both approaches:

$$\text{improve} \quad \mathbb{E}\left[r(Y^{\mathsf{sample}})\right] \tag{21}$$

where $r : \mathbb{R}^d \to \mathbb{R}_+$ is some *positive-valued* reward function, and Y^{sample} denotes the generated sample. Importantly, the goal is not to maximize $\mathbb{E}\left[r(Y^{\mathsf{sample}})\right]$, but rather to adapt the original diffusion models in a way that enhances a performance metric measured through this expected reward.

With this unified objective (21) in mind, if we were to employ a reverse SDE $\{Y_t\}_{0\le t\le 1}$ like (10b) to generate samples, then the goal would naturally become improving $\mathbb{E}[r(Y_1)]$, with Y_1 the endpoint of this reverse SDE. In practice, however, diffusion-based sampling typically applies early stopping to mitigate numerical instability, terminating the process $\{Y_t\}_{0\le t\le 1}$ at time $t = 1-\delta$ rather than $t = 1$, with $\delta > 0$ some sufficiently small quantity. To account for the effect of early stopping, we reformulate the objective (21) through posterior expectation as follows:

$$\text{improve} \quad \mathbb{E}\big[r_\delta(Y_{1-\delta})\big], \qquad \text{where} \quad r_\delta(y) \coloneqq \mathbb{E}\big[r(X_0)\,|\,X_\delta = y\big]. \tag{22}$$

Here, $\{X_t\}$ denotes the forward process (10a). Clearly, as $\delta \to 0$, this objective recovers the original form

$$\lim_{\delta\to 0} \mathbb{E}\big[r_\delta(Y_{1-\delta})\big] = \mathbb{E}\big[r(Y_1)\big],$$

thereby ensuring consistency with (21).

3.1.1 Intuition: the effect of an infinitesimal guided perturbation

We now present some algorithmic ideas to tackle the unified goal in (22), inspired by the classifier-free guidance approach and grounded in some basic analytical calculations. Specifically, recall the continuous-time class-conditional diffusion model in (15). The central idea of diffusion guidance is to augment the reverse process (15c) with an additive term proportional to $\nabla \log p_{X_{1-t}\,|\,c}(\cdot\,|\,c) - \nabla \log p_{X_{1-t}}(\cdot)$, yielding the guided diffusion process in (18). Motivated by this, it is natural to incorporate a guidance term — derived based on the reward function in (22) — into the reverse SDE of interest to steer the diffusion-based sampling towards improving the expected reward. This naturally raises the question of how such an additive guidance term influences sampling performance.

To make progress, we begin with some basic analysis. Consider perturbing the reverse SDE with a guidance term g over an infinitesimal time interval $[t, t+\Delta t]$ (for some vanishingly small Δt), and analyze the resulting effect. To be precise, consider the modified dynamics: for a given t and Δt obeying $0 < t < t + \Delta t \le 1 - \delta$,

$$\mathrm{d}Y_s^g = \Big(\frac{1}{2}Y_s^g + \nabla \log p_{X_{1-s}}(Y_s^g) + \underbrace{g}_{\text{guidance term}}\Big)\frac{\mathrm{d}s}{s} + \frac{1}{\sqrt{s}}\mathrm{d}B_s \quad \text{for } t \le s \le t+\Delta t, \tag{23a}$$

$$\mathrm{d}Y_s^g = \Big(\frac{1}{2}Y_s^g + \nabla \log p_{X_{1-s}}(Y_s^g)\Big)\frac{\mathrm{d}s}{s} + \frac{1}{\sqrt{s}}\mathrm{d}B_s \quad \text{for } 0 < s < t \ \text{ or } \ t+\Delta t < s \le 1, \tag{23b}$$

where g is assumed to be a time-invariant vector independent of the Brownian increments $\mathrm{d}B_s$ for $s \geq t$. In other words, the SDE governing the process $\{Y_s^g\}$ is identical to that of the reverse SDE (10b), except for the short time interval $[t, t+\Delta t]$ where the guidance term g is added to the drift.

To assess how the incorporation of the above guidance term g affects the resulting expected reward in (22), we characterize the limit, as $\Delta t \to 0$, of the following quantity:

$$\frac{1}{\Delta t}\left(\mathbb{E}\big[r_\delta(Y_{1-\delta}^g) \,|\, Y_t^g = y\big] - \mathbb{E}\big[r_\delta(Y_{1-\delta}) \,|\, Y_t = y\big]\right),$$

which captures the infinitesimal change in the expected reward when both the perturbed and unperturbed SDEs evolve from the same starting point. Our result is stated below; the proof is deferred to Section 6.2.

Lemma 2. *Suppose that the reward function $r(\cdot)$ satisfies $\mathbb{E}_{X_0 \sim p_{\mathsf{data}}}[r(X_0)^{1+\varepsilon}] < \infty$, where $\varepsilon > 0$ is some small constant. Recall the definition of r_δ in (22). One has*

$$\begin{aligned}
&\lim_{\Delta t \to 0} \frac{1}{\Delta t}\left(\mathbb{E}\big[r_\delta(Y_{1-\delta}^g) \,|\, Y_t^g = y\big] - \mathbb{E}\big[r_\delta(Y_{1-\delta}) \,|\, Y_t = y\big]\right) \\
&\quad = \frac{\mathbb{E}\big[r_\delta(Y_{1-\delta}) \,|\, Y_t = y\big]}{t}\left\langle \nabla \log p_{X_{1-t}^{r\text{-wt}}}(y) - \nabla \log p_{X_{1-t}}(y), g \right\rangle,
\end{aligned} \tag{24}$$

where the random vector $X_{1-t}^{r\text{-wt}}$ follows the reweighted distribution below for any $0 \leq t \leq 1$:

$$X_{1-t}^{r\text{-wt}} \,|\, X_0^{r\text{-wt}} \sim \mathcal{N}\left(\sqrt{t} X_0^{r\text{-wt}}, (1-t) I_d\right), \qquad p_{X_0^{r\text{-wt}}}(x_0) := \frac{r(x_0) p_{X_0}(x_0)}{\mathbb{E}_{X_0 \sim p_{\mathsf{data}}}[r(X_0)]} \text{ for any } x_0 \in \mathbb{R}^d. \tag{25}$$

Here, $X_0^{r\text{-wt}}$ is drawn from the reward-reweighted data distribution defined in (25), and $X_t^{r\text{-wt}}$ evolves from $X_0^{r\text{-wt}}$ following the same update rule as in the original forward process (10a). In words, this lemma reveals that the guided perturbation affects the infinitesimal change in the expected reward linearly. In particular, the influence of the guidance term on the resulting reward increases when it aligns more closely with a certain score difference — that is, the difference between the score w.r.t. the reward-reweighted distribution and the original score. As it turns out, this lemma plays the most critical role in our theoretical and algorithmic development.

3.1.2 A unified algorithmic framework

With the key finding in Lemma 2 in mind, a natural strategy to improve the expected reward in (22) is to align the guidance term with the score difference $\nabla \log p_{X_{1-t}^{r\text{-wt}}}(\cdot) - \nabla \log p_{X_{1-t}}(\cdot)$, and to incorporate it into the reverse-time SDE across the entire trajectory. Accordingly, we put forward the following guided reverse-time SDE $\{Y_t^w\}_{0 \leq t \leq 1}$ to tackle the unified objective in (22):

$$\begin{aligned}
\mathrm{d}Y_t^w &= \Big(\frac{1}{2}Y_t^w + \nabla \log p_{X_{1-t}}(Y_t^w) + \underbrace{w\big[\nabla \log p_{X_{1-t}^{r\text{-wt}}}(Y_t^w) - \nabla \log p_{X_{1-t}}(Y_t^w)\big]}_{\text{guidance term}}\Big)\frac{\mathrm{d}t}{t} + \frac{1}{\sqrt{t}}\mathrm{d}B_t \\
&= \Big(\frac{1}{2}Y_t^w + (1-w)\nabla \log p_{X_{1-t}}(Y_t^w) + w\nabla \log p_{X_{1-t}^{r\text{-wt}}}(Y_t^w)\Big)\frac{\mathrm{d}t}{t} + \frac{1}{\sqrt{t}}\mathrm{d}B_t, \qquad \text{for } 0 \leq t \leq 1,
\end{aligned} \tag{26}$$

where $w > 0$ represents the guidance scale, and B_t is a standard Brownian motion in $\mathbb{R}^d$. Clearly, when $w = 0$, this SDE (26) reduces to the original reverse process in (10b). The corresponding discrete-time diffusion-based sampling algorithm is thus given by

$$Y_{n-1}^w = \frac{1}{\sqrt{1-\beta_n}}\Big(Y_n^w + \beta_n\big((1-w)s_n(Y_n^w) + w s_n^{r\text{-wt}}(Y_n^w)\big)\Big) + \sqrt{\beta_n} Z_n, \quad n = N, \cdots, 2. \tag{27}$$

Here, $s_n^{r\text{-wt}}(\cdot)$ represents an estimate of the score function $s_n^{r\text{-wt},\star}(\cdot)$ w.r.t. the reward-reweighted distribution, defined as

$$s_n^{r\text{-wt},\star}(y) := \nabla \log p_{X_{1-\overline{\alpha}_n}^{r\text{-wt}}}(y), \tag{28}$$

where $\overline{\alpha}_n$ is defined in (6).

Score learning for the reward-reweighted distribution. The proposed guided approach in (26) and (27) involves the score function $\nabla \log p_{X_{1-t}^{r\text{-wt}}}(\cdot)$ w.r.t. the reward-reweighted distribution, raising the natural question of how to perform score learning in practice. As it turns out, such reward-reweighted score functions can be learned using techniques analogous to those standard score matching methods for estimating the original score functions (Hyvärinen, 2005). More precisely, we propose to optimize a reward-reweighted denoising score matching objective as follows:

$$\widehat{\theta}^{r\text{-wt}} = \arg\min_{\theta} \mathop{\mathbb{E}}_{t, x_0 \sim p_{\mathsf{data}}, \epsilon \sim \mathcal{N}(0,I), x_t = \sqrt{1-t}x_0 + \sqrt{t}\epsilon} \left[r(x_0) \left\| \epsilon - \mathsf{NN}_\theta(x_t, t) \right\|_2^2 \right], \tag{29}$$

where NN_θ denotes a neural network, parameterized by θ, used to learn the noise ϵ. Note that the expectation is also taken over some distribution of t, with one example being the uniform distribution over $[0, 1]$. Once $\widehat{\theta}^{r\text{-wt}}$ is obtained, we can take the score estimate to be

$$s_n^{r\text{-wt}}(x) = -\frac{\mathsf{NN}_{\widehat{\theta}^{r\text{-wt}}}\big(x, 1-\overline{\alpha}_n\big)}{\sqrt{1-\overline{\alpha}_n}}. \tag{30}$$

For completeness, the proofs of the validity of (29) and (30) are provided in Section A.

3.1.3 Theoretical guarantees

To validate the effectiveness of the proposed approach (cf. (26)) in enhancing the expected reward relative to the unguided approach, we present the following theorem, which pins down the resulting improvement in the expected reward. It is noteworthy that the infinitesimal characterization in Lemma 2 does not readily imply such an improvement, as the guided SDE involves time-varying guidance terms injected along the entire trajectory that needs to be carefully coped with.

In order to present our theoretical guarantees, we first extend the definition of $r_\delta(\cdot)$ in (22) as follows:

$$r_t(y) \coloneqq \mathbb{E}[r(X_0) \,|\, X_t = y] \qquad \text{for any } 0 \le t < 1. \tag{31}$$

Our main theorem is as follows.

Theorem 1. *Suppose that the reward function $r(\cdot)$ satisfies $\mathbb{E}_{X_0 \sim p_{\mathsf{data}}}[r(X_0)] < \infty$. Let $\{Y_t^w\}$ denote the guided reverse SDE in (26), and $\{Y_t\}$ the original reverse SDE in (10b). Then for any initialization $y \in \mathbb{R}^d$, the guided approach (26) achieves*

$$\begin{aligned} &\mathbb{E}\big[r_\delta(Y_{1-\delta}^w) \,|\, Y_0^w = y\big] - \mathbb{E}\big[r_\delta(Y_{1-\delta}) \,|\, Y_0 = y\big] \\ &\quad = \int_0^{1-\delta} \frac{w}{t} \mathbb{E}\left[r_{1-t}(Y_t^w) \left\| \nabla \log p_{X_{1-t}}(Y_t^w) - \nabla \log p_{X_{1-t}^{r\text{-wt}}}(Y_t^w) \right\|_2^2 \,\middle|\, Y_0^w = y \right] \mathrm{d}t. \end{aligned} \tag{32}$$

The proof of Theorem 1 is postponed to Section 6.3. In essence, Theorem 1 reveals that, under mild conditions, adding the guidance term into the drift results in a strictly higher expected reward compared to the unguided diffusion model, irrespective of the initialization (as long as the guided and unguided methods start from the same point). Crucially, this theorem clarifies the specific metric under which the guided approach offers its advantages — an insight that is far from obvious from the form of the SDE (26).

3.1.4 A closely related formulation: cost reduction

Thus far, our unified framework has focused on improving the expected reward, assuming a positive-valued reward function. A closely related objective is to reduce the expected value of a certain cost function associated with generated samples. Although reward improvement and cost reduction are intimately connected, the positivity assumption on the reward function makes it nontrivial to directly apply Theorem 1 to a positive-valued cost function. Fortunately, our algorithm and theoretical framework naturally extend to this setting without modification. For completeness, we present the corresponding results for the cost reduction formulation below, noting that the theoretical guarantees follow from the same analysis arguments and are therefore omitted for brevity.

More precisely, suppose that the new objective is to

$$\text{reduce} \quad \mathbb{E}\left[J(Y^{\mathsf{sample}})\right], \tag{33}$$

where $J : \mathbb{R}^d \to \mathbb{R}_+$ denotes a *positive-valued* cost function, and Y^{sample} is the generated sample. As before, we seek to generate samples via an SDE $\{Y_t^w\}_{0\le t\le 1-\delta}$ with early stopping, and reformulate the objective (33) accordingly as

$$\text{reduce} \quad \mathbb{E}\big[J_\delta(Y_{1-\delta})\big], \qquad \text{where} \quad J_\delta(y) \coloneqq \mathbb{E}\big[J(X_0)\,|\,X_\delta = y\big], \tag{34}$$

where $\{X_t\}$ represents the forward process (10a), and $Y_{1-\delta}$ is the output of the early-stopped SDE.

Algorithm. To achieve this goal, we propose the following guided reverse-time SDE $\{Y_t^w\}_{0\le t\le 1}$:

$$\begin{aligned}
\mathrm{d}Y_t^w &= \Big(\frac{1}{2}Y_t^w + \nabla \log p_{X_{1-t}}(Y_t^w) + \underbrace{w\big[\nabla \log p_{X_{1-t}}(Y_t^w) - \nabla \log p_{X_{1-t}^{J\text{-wt}}}(Y_t^w)\big]}_{\text{guidance term}}\Big)\frac{\mathrm{d}t}{t} + \frac{1}{\sqrt{t}}\mathrm{d}B_t \\
&= \Big(\frac{1}{2}Y_t^w + (1+w)\nabla \log p_{X_{1-t}}(Y_t^w) - w\nabla \log p_{X_{1-t}^{J\text{-wt}}}(Y_t^w)\Big)\frac{\mathrm{d}t}{t} + \frac{1}{\sqrt{t}}\mathrm{d}B_t, \qquad \text{for } 0\le t\le 1,
\end{aligned} \tag{35}$$

where B_t is a standard Brownian motion, $w > 0$ denotes the guidance scale, and $X_{1-t}^{J\text{-wt}}$ follows a cost-reweighted distribution defined, for each $0 \le t \le 1$, by

$$X_{1-t}^{J\text{-wt}} \,|\, X_0^{J\text{-wt}} \sim \mathcal{N}\left(\sqrt{t}X_0^{J\text{-wt}}, (1-t)I_d\right), \qquad p_{X_0^{J\text{-wt}}}(x_0) \coloneqq \frac{J(x_0)p_{X_0}(x_0)}{\mathbb{E}_{X_0\sim p_{\mathsf{data}}}[J(X_0)]} \text{ for any } x_0 \in \mathbb{R}^d. \tag{36}$$

In comparison to (26), the sign of the score difference term (i.e., the guidance term) in (35) is flipped, reflecting the objective of decreasing, rather than increasing, the expected cost. The corresponding discrete-time sampler is thus given by

$$Y_{n-1}^w = \frac{1}{\sqrt{1-\beta_n}}\Big(Y_n^w + \beta_n\big((1+w)s_n(Y_n^w) - ws_n^{J\text{-wt}}(Y_n^w)\big)\Big) + \sqrt{\beta_n}Z_n \quad n = N, \cdots, 2 \tag{37}$$

with the Z_n's independently drawn from $\mathcal{N}(0, I_d)$. Here, $s_n^{J\text{-wt}}(\cdot)$ represents an estimate of the cost-reweighted score function defined as

$$s_n^{J\text{-wt},\star}(y) \coloneqq \nabla \log p_{X_{1-\overline{\alpha}_n}^{J\text{-wt}}}(y), \tag{38}$$

with $\overline{\alpha}_n$ defined in (6).

Theoretical guarantees. Akin to Theorem 1, we have the following theoretical guarantees for the continuous-time sampler (35), rigorously quantifying the reduction in expected cost achieved by the guidance term and demonstrating the proven advantage of the guided approach over its unguided counterpart.

For any $0 \le t < 1$, define

$$J_t(y) \coloneqq \mathbb{E}[J(X_0)\,|\,X_t = y]. \tag{39}$$

Our theorem is stated as follows.

Theorem 2. *Suppose that the cost function $J(\cdot)$ satisfies $\mathbb{E}_{X_0\sim p_{\mathsf{data}}}[J(X_0)] < \infty$. Let $\{Y_t^w\}$ denote the guided reverse SDE in (35), and $\{Y_t\}$ the original reverse SDE in (10b). Then for any initialization $y \in \mathbb{R}^d$, the guided approach (35) achieves*

$$\begin{aligned}
&\mathbb{E}\big[J_\delta(Y_{1-\delta})\,|\,Y_0 = y\big] - \mathbb{E}\big[J_\delta(Y_{1-\delta}^w)\,|\,Y_0^w = y\big] \\
&\quad = \int_0^{1-\delta} \frac{w}{t}\mathbb{E}\left[J_{1-t}(Y_t^w)\left\|\nabla \log p_{X_{1-t}}(Y_t^w) - \nabla \log p_{X_{1-t}^{J\text{-wt}}}(Y_t^w)\right\|_2^2 \,\middle|\, Y_0^w = y\right]\mathrm{d}t.
\end{aligned} \tag{40}$$

3.2 Consequences for specific models

To illustrate the utility of our unified framework established in Section 3.1, we now develop its concrete consequences and implications for both classifier-free diffusion guidance and reward-guided diffusion models.

3.2.1 Classifier-free diffusion guidance

The first application of our unified framework concerns classifier-free diffusion guidance, as introduced in Section 2.2. One of the key challenges lies in identifying the underlying reward (or cost) function that the diffusion guidance mechanism seeks to improve, which is far from evident in the sampling dynamics described by (18).

Understanding classifier-free guidance through our unified framework. To leverage our unified framework in elucidating the effectiveness of CFG, a crucial step is to establish the precise connection between the SDE (18) underlying CFG and the guided SDE we propose in Section 3.1. Recognizing that the update rule (18) bears some resemblance to the SDE (35), we shall focus on situating CFG within the cost reduction framework in Section 3.1.4.

- *Selecting the forward process* $\{X_t\}$. Since CFG (18) is constructed by modifying the conditional diffusion process, we choose $\{X_t\}$ in this subsection to be conditional forward process defined in (15a), with $X_0 \sim p_{\mathsf{data}\,|\,c}$. This means that the score $\nabla \log p_{X_{1-t}}$ in (35) needs to be replaced with $\nabla \log p_{X_{1-t}\,|\,c}$.
- *Selecting the unguided reverse process* $\{Y_t\}$. With the forward process given by (15a), we shall take the corresponding unguided reverse process $\{Y_t\}$ to be (15c).
- *Selecting the cost function* $J(\cdot)$. Choosing $J(\cdot)$ suitably is essential to ensure that our algorithm (35) matches the CFG update rule in (18). As it turns out, the following choice — which is the the reciprocal of the classifier probability — satisfies this requirement:

$$J(y) = \frac{1}{p_{c\,|\,X_0}(c\,|\,y)}. \tag{41}$$

  In the presence of early stopping, the modified cost function $J_\delta(\cdot)$ admits a simplified expression

$$J_\delta(y) \coloneqq \mathbb{E}\left[\frac{1}{p_{c\,|\,X_0}(c\,|\,X_0)} \,\middle|\, X_\delta = y, c\right] = \frac{1}{p_{c\,|\,X_\delta}(c\,|\,y)}. \tag{42}$$

  In addition, the cost-reweighted distribution also takes a particularly simple form:

$$p_{X_t^{J\text{-wt}}}(y) = p_{X_t}(y) \qquad \text{for all } 0 \le t < 1, \tag{43}$$

  which, in effect, reduces to the unconditional distribution. To streamline presentation, the proofs of both (42) and (43) and are deferred to the end of this subsection.

With the above components in place, it is straightforward to verify that the proposed SDE (35) coincides exactly with the SDE (18) underlying CFG.

Theoretical guarantees and implications. Applying Theorem 2 thus leads to the following corollary, which rigorizes the advantage of classifier-free guidance in terms of improving the expected reciprocal of classifier probability. This is an immediate consequence of Theorem 2 as well as the description in this subsection.

Corollary 1 (Effectiveness of classifier-free diffusion guidance)**.** *Consider any given* $0 < \delta < 1$ *and any initialization* $y \in \mathbb{R}^d$*. Assume the prior probability* $p(c) > 0$ *is positive. Then the guided SDE* $\{Y_t^w\}$ *defined in* (18) *achieves*

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{p_{c\,|\,X_\delta}(c\,|\,Y_{1-\delta})} \,\middle|\, Y_0 = y\right] - \mathbb{E}\left[\frac{1}{p_{c\,|\,X_\delta}(c\,|\,Y^w_{1-\delta})} \,\middle|\, Y_0^w = y\right] \\
&= \int_0^{1-\delta} \frac{w}{t} \mathbb{E}\left[\frac{1}{p_{c\,|\,X_{1-t}}(c\,|\,Y_t^w)} \left\| \nabla \log p_{X_{1-t}\,|\,c}(Y_t^w) - \nabla \log p_{X_{1-t}}(Y_t^w) \right\|_2^2 \,\middle|\, Y_0^w = y\right] \mathrm{d}t.
\end{aligned} \tag{44}$$

Several remarks regarding the implications of Corollary 1 are in order.

- *Provable benefits of diffusion guidance and connections to the inception score.* Corollary 1 indicates that the average reciprocal of classifier probability decreases — often strictly — when non-zero guidance is applied, thereby providing a rigorous justification for the improvement targeted by CFG. The metric $p_{c\,|\,X_\delta}(c\,|\,y)^{-1}$ is oftentimes large for samples with low perceptual quality, suggesting that reducing $\mathbb{E}[p_{c\,|\,X_\delta}(c\,|\,Y_{1-\delta})^{-1}]$ effectively prioritizes the mitigation of low-quality or misclassified samples. This is consistent with the original motivation for introducing guidance (Dhariwal and Nichol, 2021). At a conceptual level, reducing the average reciprocal of classifier probability is also somewhat aligned with increasing the Inception Score (IS), the latter of which is associated with the expected logarithm of the classifier probability, i.e., $\mathbb{E}[\log p_{c\,|\,X_0}(c\,|\,Y^{\mathsf{sample}})]$.
- *Guidance improves overall quality but not individual samples.* Corollary 1 states that diffusion guidance improves the averaged reciprocal of the classifier probability, rather than guaranteeing improvement in classifier probability for individual samples. Therefore, while overall sample quality is enhanced via guidance, a small subset of samples might exhibit degradation in sample quality. This observation is corroborated by subsequent numerical experiments (to be reported in Figures 1 and 2).

Proof of expressions (42) and (43). To justify the validity of (42), we observe that

$$\begin{aligned}\mathbb{E}\left[\frac{1}{p_{c\,|\,X_0}(c\,|\,X_0)}\,\middle|\,X_\delta=y,c\right] &= \int \frac{p_{X_0\,|\,X_\delta,c}(x_0\,|\,y,c)}{p_{c\,|\,X_0}(c\,|\,x_0)}\mathrm{d}x_0 \overset{\text{(i)}}{=} \int \frac{p_{X_\delta,X_0\,|\,c}(y,x_0\,|\,c)}{p_{X_\delta\,|\,c}(y\,|\,c)}\frac{p_{X_0}(x_0)}{p_{X_0\,|\,c}(x_0\,|\,c)p(c)}\mathrm{d}x_0 \\ &\overset{\text{(ii)}}{=} \int \frac{p_{X_\delta\,|\,X_0}(y\,|\,x_0)p_{X_0\,|\,c}(x_0\,|\,c)}{p_{X_\delta\,|\,c}(y\,|\,c)}\frac{p_{X_0}(x_0)}{p_{X_0\,|\,c}(x_0\,|\,c)p(c)}\mathrm{d}x_0 \\ &= \frac{p_{X_\delta}(y)}{p_{X_\delta\,|\,c}(y\,|\,c)p(c)} = \frac{1}{p_{c\,|\,X_\delta}(c\,|\,y)},\end{aligned}$$

where (i) arises from the Bayes rule, and (ii) holds since $c \to X_0 \to X_\delta$ forms a Markov chain.

Regarding (43), it is readily seen from the definition (25) that

$$p_{X_{1-t}^{J\text{-wt}}}(y) = \frac{\int p_{c\,|\,X_0}(c\,|\,x_0)^{-1}p_{X_0\,|\,c}(x_0\,|\,c)p_{X_t\,|\,X_0}(y\,|\,x_0)\mathrm{d}x_0}{\int p_{c\,|\,X_0}(x\,|\,x_0)^{-1}p_{X_0\,|\,c}(x_0\,|\,c)\mathrm{d}x_0} = \frac{\int p_{X_0}(x_0)p_{X_t\,|\,X_0}(y\,|\,x_0)\mathrm{d}x_0}{\int p_{X_0}(x_0)\mathrm{d}x_0} = p_{X_t}(y),$$

where the second identity follows from the Bayes rule.

Proof of Corollary 1. Note that Theorem 2 applies to any positive function $J(x_0)$ satisfying $\mathbb{E}[J(X_0)] < \infty$. Thus, equipped with the description in this section, it suffices to verify that $\mathbb{E}[J(X_0)] < \infty$ is satisfied for this special case. Towards this, observe that the cost function $J(y)$ defined in (41) obeys

$$\mathop{\mathbb{E}}_{X_0\sim p_{\mathsf{data}\,|\,c}}[J(X_0)] = \mathop{\mathbb{E}}_{X_0\sim p_{\mathsf{data}\,|\,c}}\left[\frac{1}{p_{c\,|\,\mathsf{data}}(c\,|\,X_0)}\right] = \int \frac{p_{\mathsf{data}}(x_0)}{p(c)p_{\mathsf{data}\,|\,c}(x_0\,|\,c)}p_{\mathsf{data}\,|\,c}(x_0\,|\,c)\mathrm{d}x_0 = \frac{1}{p(c)} < \infty, \tag{45}$$

as long as $p(c) > 0$ for every class label c. This justifies the validity of Corollary 1.

3.2.2 Reward-guided diffusion models

We now turn attention to another application: reward-guided diffusion models, as described in Section 2.3. Rather than applying reinforcement learning techniques to optimize (20), we seek to directly modify the unguided backward process in (10b) to achieve the goal in (19).

Given a general external reward function $r^{\mathsf{ext}} : \mathbb{R}^d \to \mathbb{R}$, we define the following exponentiated rewards as objective functions:

$$r(y) = \exp\big(\beta r^{\mathsf{ext}}(y)\big), \qquad r_\delta(y) = \mathbb{E}\big[\exp\big(\beta r^{\mathsf{ext}}(X_0)\big)\,|\,X_\delta = y\big], \tag{46}$$

where $\beta > 0$ is a hyperparameter. This ensures non-negativity, as required in Theorem 1. Substituting the above choice (46) into the algorithm in (26) and Theorem 1, we immediately arrive at the following performance guarantees.

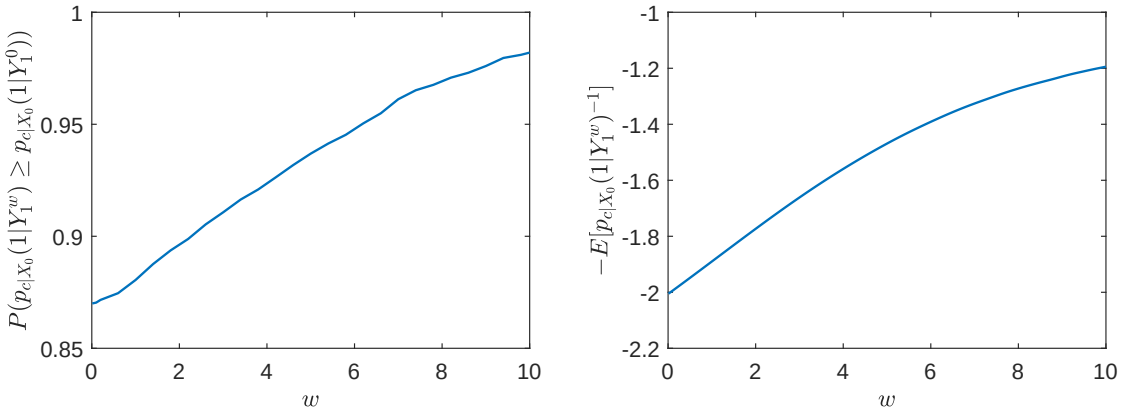

Figure 1: Experimental results under the Gaussian mixture model. (Left) Proportions of samples with improved classifier probabilities; (Right) Averages of $-p_{c\,|\,X_0}(1\,|\,Y_1^w)^{-1}$ for varying guidance scales w.

Corollary 2 (Effectiveness of reward-guided diffusion models)**.** *Assume that the reward function* $r^{\mathsf{ext}}(\cdot)$ *satisfies* $\mathbb{E}\big[\exp\big(\beta r^{\mathsf{ext}}(X_0)\big)\big] < \infty$. *Consider the guided SDE* $\{Y_t^w\}$ *constructed in* (26) *with the reward functions defined in* (46)*, along with the unguided backward SDE* $\{Y_t\}$ *defined in* (10b)*. Then for any initialization* $y \in \mathbb{R}^d$ *and any given quantity* $0 < \delta < 1$*, one has*

$$\begin{aligned}&\mathbb{E}\big[r_\delta(Y_{1-\delta}^w)\,|\,Y_0^w = y\big] - \mathbb{E}\big[r_\delta(Y_{1-\delta})\,|\,Y_0 = y\big]\\ &\quad= \int_0^{1-\delta} \frac{w}{t}\mathbb{E}\left[r_{1-t}(Y_t^w)\Big\|\nabla \log p_{X_{1-t}^{r\text{-}\mathsf{wt}}}(Y_t^w) - \nabla \log p_{X_{1-t}}(Y_t^w)\Big\|_2^2\,|\,Y_0^w = y\right]\mathrm{d}t.\end{aligned} \tag{47}$$

Remark 1. Note that the objective function is defined as the posterior expectation of the exponentiated reward to ensure that the density of $X_0^{r\text{-}\mathsf{wt}}$ remains positive. If the reward function $r^{\mathsf{ext}}(\cdot)$ is strictly positive, then Corollary 2 also directly applies to $r(y) = r^{\mathsf{ext}}(y)$ and $r_\delta(y) = \mathbb{E}[r^{\mathsf{ext}}(X_0)\,|\,X_\delta = y]$.

Importantly, the proposed sampler in (26) and (27) offers several practical advantages. In contrast to prior methods for reward-guided diffusion models — which often require sampling full trajectories to compute gradients (see Section 2.3) — our method enables training (29) using denoising score matching with only a single noise level t per update. This leads to a significant reduction in implementation complexity. Moreover, our sampler only requires training once (i.e., training both the score functions and the reward-reweighted score functions during the pretraining stage). After this, the pretrained scores can be reused for any choice of guidance strength w, eliminating the need for retraining when w varies.

4 Numerical experiments

To corroborate our theoretical findings in Section 3, this section conducts a series of numerical experiments for both classifier-free diffusion guidance and reward-guided diffusion models.

4.1 Classifier-free diffusion guidance

We now provide numerical validation of Corollary 1 using both synthetic and real-world datasets. A key takeaway from the experiments below is that: classifier-free guidance does not uniformly enhance the quality of every generated sample; instead, it improves the overall sample quality by reducing the expected reciprocal of the classifier probability.

The Gaussian mixture model. Starting with syntheic data, we consider a one-dimensional Gaussian Mixture Model (GMM) with two classes $c \in \{0, 1\}$, each having equal prior probability $p_c(0) = p_c(1) = 0.5$;

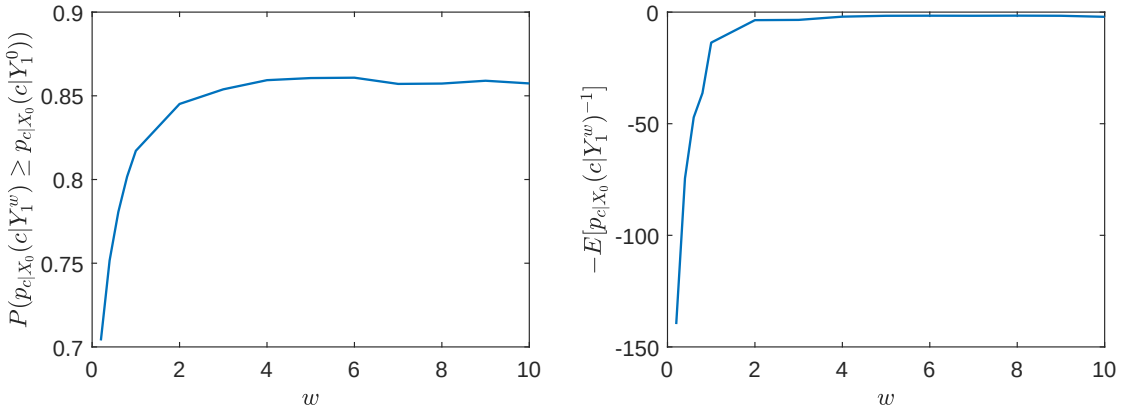

Figure 2: Experimental results on the ImageNet dataset. (Left) Proportions of samples with improved classifier probabilities; (Right) Averages of $-p_{c\,|\,X_0}(1\,|\,Y_1^w)^{-1}$ for varying guidance scales w.

here $p_c(\cdot)$ denotes the prior distribution of the class labels. The data distribution is defined as follows:

$$\begin{aligned} X_0\,|\,c=0 &\sim \mathcal{N}(0,1) \\ X_0\,|\,c=1 &\sim \frac{1}{2}\mathcal{N}(1,1)+\frac{1}{2}\mathcal{N}(-1,1). \end{aligned}$$

We focus on data generation for class $c=1$. The score functions $\nabla \log p_{X_{1-t}\,|\,c}(x\,|\,1)$, $\nabla \log p_{X_{1-t}}(x)$, and the classifier probability $p_{c\,|\,X_{1-t}}(1\,|\,x)$ all admit closed-form expressions, which are provided in Appendix B.1 (see (94), (95), and (96), respectively). To empirically verify our theoretical findings, we simulate the diffusion process with guidance scale w varying from 0.01 to 10, performing 10^4 independent trials for each w. In each trial, we initialize the sampler (17) with $Y_N^w \sim \mathcal{N}(0,1)$ and update according to the CFG update rule (17) with $N=4000$ steps. We obtain Y_1^w and its unguided counterpart Y_1^0 with the same initialization $Y_N^0 = Y_N^w$. The stepsizes $\{\beta_n\}$ are provided later (cf. (49)). For each trial, we use the classifier probabilities $p_{c\,|\,X_0}(1\,|\,Y_1^w)$ and $p_{c\,|\,X_0}(1\,|\,Y_1^0)$ to approximate the theoretically analyzed quantities $p_{c\,|\,X_\delta}(Y_{1-\delta}^w)$ and $p_{c\,|\,X_\delta}(Y_{1-\delta}^0)$, respectively. We then calculate two metrics for each choice of w:

- the proportion of trials satisfying $p_{c\,|\,X_0}(1\,|\,Y_1^w) \geq p_{c\,|\,X_0}(1\,|\,Y_1^0)$;
- the empirical average of $-\frac{1}{p_{c\,|\,X_0}(1\,|\,Y_1^w)}$.

The numerical results are presented in Figure 1.

The ImageNet dataset. Next, we conduct numerical experiments on the ImageNet dataset (Deng et al., 2009). Samples are generated using a pre-trained diffusion model (Rombach et al., 2021) with varying choices of guidance scales w, and classifier probabilities are evaluated using the Inception v3 classifier (Szegedy et al., 2016). For each value of w, we compute the aforementioned two metrics averaged over a total of 2×10^4 random trials — 20 trials for each of the 10^3 ImageNet categories. The stepsizes $\{\beta_n\}$ are

$$\beta_n \coloneqq \frac{c_1(1-\overline{\alpha}_n)\log N}{N}\left(1+\frac{c_1(1-\overline{\alpha}_n)\log N}{N}\right)^{-1}, \tag{49}$$

where the sequence $\{\overline{\alpha}_n\}$ is recursively defined as:

$$\begin{aligned} \overline{\alpha}_N &\coloneqq \frac{1}{N^{c_0}}, \\ \overline{\alpha}_{n-1} &\coloneqq \overline{\alpha}_n + \frac{c_1\overline{\alpha}_n(1-\overline{\alpha}_n)\log N}{N}. \end{aligned}$$

Here, we set the parameters to be $c_0 = 1, c_1 = 2$, and $N = 4000$. The numerical results are illustrated in Figure 2.

Numerical findings. In both of the above cases, it is observed that the proportion of samples with improved classifier probability remains strictly less than 1 for all tested values of w, thus indicating that CFG does not guarantee improvement for every individual sample. However, the average of $-p_{c\,|\,X_0}(1\,|\,Y_1^w)^{-1}$ increases with w, thereby explaining in part how guidance enhances sample quality, as predicted by Corollary 1. Additionally, we remark that in practical applications, the performance of guidede diffusion models is often assessed by two criteria: diversity and sample quality. The current paper focuses primarily on improvement on the classifier probability as the guidance scale w increases, which influences the sample quality; in contrast, prior work (e.g., Ho and Salimans (2021); Wu et al. (2024a)) also noted that large values of w may lead to degradation of diversity.

4.2 Reward-guided diffusion models

Next, we implement the proposed sampler (27) and conduct "proof-of-concept" experiments for two basic examples with synthetic data, one involving a Gaussian mixture distribution and the other a Swiss rolls dataset.

The Gaussian mixture model. Consider a mixture of two one-dimensional Gaussian distributions:

$$X_0 \sim \frac{1}{2}\mathcal{N}(-1,\sigma^2) + \frac{1}{2}\mathcal{N}(1,\sigma^2), \qquad \text{and} \quad X_\tau = \sqrt{1-\tau}X_0 + \sqrt{\tau}Z.$$

Define the reward function as

$$r^{\mathsf{ext}}(x) = -(x-2)^2.$$

which encourages the sampler to generate samples near 2, that is, on the right tail of the original distribution p_{X_0}. In this setting, we can derive closed-form expressions for the score functions $s_n^\star(\cdot)$ and $s_n^{r\text{-}\mathsf{wt},\star}(\cdot)$; detailed expressions are derived in Appendix B.2.

In our experiment, we evaluate the impacts of different values of w by implementing the sampler defined in (27), adopting the same stepsizes $\{\beta_n\}$ (cf. (49)) and the exact score functions $s_n^\star(\cdot)$ and $s_n^{r\text{-}\mathsf{wt},\star}(\cdot)$. Parameter $\beta = 1$. For each model with a specific value of w, we generate 10^5 independent samples. The corresponding empirical probability density functions are illustrated in Figure 3.

As shown in Figure 3, when $w = 0$, the proposed sampler in (27) reduces to the standard DDPM, and our sampler generates samples that approximately follow the original data distribution p_{X_0} when w is close to zero. In contrast, when w is large (e.g., $w \geq 1$), the guidance term becomes dominant. Consequently, the algorithm generates samples that concentrate around the value 2 to obtain higher rewards, resulting in a sample distribution that deviates substantially from p_{X_0}.

Swiss roll. We next consider a two-dimensional Swiss roll dataset, which follows the uniform distribution among 10^3 points in Figure 4. Consider the following reward function

$$r^{\mathsf{ext}}([x_1,x_2]) = 10\,\mathbb{1}(x_1 \in [-5,6]),$$

which encourages the sampler to generate samples whose first coordinate lies within the interval $[-5,6]$. The computation of the corresponding score functions in this setting is detailed in Appendix B.3. The experimental setup uses the same parameter as the above GMM experiments. For each specified value of w, we generate 10^3 samples, with the numerical results plotted in Figure 5.

Similar to the GMM case, when $w \approx 0$, the proposed reward-guided algorithm generates samples that closely match the data distribution illustrated in Figure 4. As the value of w increases, the generated samples tend to concentrate within the rectangular region $[-5,6] \times \mathbb{R}$. Moreover, when w is sufficiently large, nearly all generated samples are confined to this region in order to achieve a higher reward.

5 Other related work

We now briefly discuss several other recent papers related to this work. A number of works have proposed various guidance mechanisms to improve sample quality (Guo et al., 2024; Karras et al., 2024; Yu et al.,

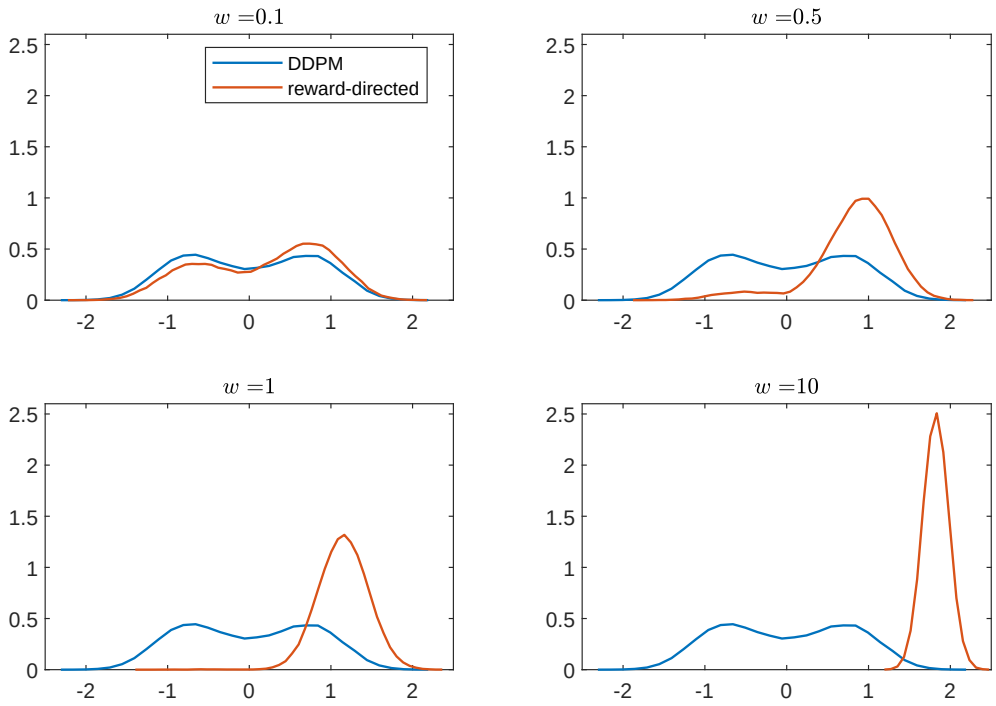

Figure 3: Empirical distributions of reward-guided sampler with specific values of w vs. DDPM ($w = 0$).

2023; Galashov et al., 2025). For instance, Guo et al. (2024) introduced a gradient-based guidance term and analyzed its effectiveness under a sort of linear score assumption; Karras et al. (2024) proposed using a "bad" version of the diffusion model itself as a guidance term; Yu et al. (2023) developed a training-free guidance method with the aid of an off-the-shelf classifier; Galashov et al. (2025) proposed to learn the guidance level w in CFG as continuous functions of both the conditioning and the noise level to further enhance sample quality, and extended this framework to reward-guided sampling. Note that guided diffusion models have been studied in other settings and applications as well, including inverse problems (Chung et al., 2022), discrete (state-space) diffusion models (Nisonoff et al., 2024; Schiff et al., 2024), and masked discrete diffusion (Rojas et al., 2025; Ye et al., 2025), among others.

A growing body of work has sought to theoretically understand the working mechanism of diffusion guidance. The recent papers Wu et al. (2024a); Chidambaram et al. (2024); Bradley and Nakkiran (2024) are perhaps most relevant to our work, although they focused primarily on more special families of target distributions. More specifically, Wu et al. (2024a) established that $p_{c\,|\,X_0}(1\,|\,Y_1^w) \geq p_{c\,|\,X_0}(1\,|\,Y_1^0)$ for the Gaussian mixture models (GMMs) under specific conditions; Chidambaram et al. (2024) argued that the guidance mechanism might degrade the performance of diffusion models by inducing mean overshoot and variance shrinkage, focusing on both GMMs and mixtures of compactly supported distributions. Moreover, Bradley and Nakkiran (2024) demonstrated that, at least for GMMs, diffusion guidance does not produce samples from the distribution $p_{X_0\,|\,c}(x\,|\,c)^w p_{X_0}(x)^{1-w}$, and established the intimate connection between diffusion guidance and an alternative scheme called the single-step predictor-corrector method. In addition to these papers, Pavasovic et al. (2025) proved that under certain conditions, the influence of guidance on the generated distribution diminishes as the data dimension increases; Li et al. (2025c) analyzed CFG within a simplified linear diffusion model and subsequently validated the theoretical insights on real-world nonlinear diffusion models; Jiao et al. (2025a) established some connection between diffusion guidance and test-time scaling (particularly a soft variant of the best-of-N sampling); Jin et al. (2025) investigated CFG in the context of Gaussian Mixture Models, while Tang and Xu (2024) studied the effectiveness of guidance for

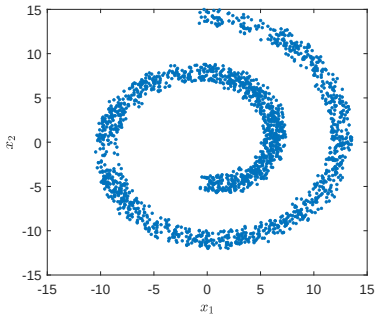

Figure 4: The target distribution is assumed to be the uniform distribution over the set of points shown in the figure.

steering generated samples toward a prescribed guidance set. On the score matching side, Fu et al. (2024) derived sample complexity bounds for estimating the conditional score function used in conditional sampling, and Liang et al. (2024) analyzed the impact of score mismatch and showed that it induces an asymptotic distributional bias between the target distribution and the sampling distribution.

In addition to the above work on guided diffusion models, a more principled introduction to diffusion models and recent development can be found in Lai et al. (2025); Chen et al. (2024a); Tang and Zhao (2024). It is also worth noting that the convergence properties of discrete-time diffusion-based samplers have been extensively studied in recent work, showing that the distribution of the generated samplers can well approximate the target data distribution under mild assumptions; see, e.g., Lee et al. (2022, 2023); Chen et al. (2022); Benton et al. (2023); Chen et al. (2023); Li et al. (2024b); Gupta et al. (2024); Chen et al. (2024b); Li et al. (2024a); Li and Yan (2024); Li and Jiao (2024); Li and Cai (2024); Huang et al. (2024); Cai and Li (2025); Liang et al. (2025); Wu et al. (2024b); Li et al. (2025b,a); Potaptchik et al. (2024); Azangulov et al. (2024); Jiao et al. (2025b); Zhang et al. (2025) and the references therein.

6 Analysis

In this section, we shall provide details of the proof of our main results. Note that Corollaries 1 and 2 can be derived immediately after establishing Theorem 1.

6.1 Preliminaries

Before embarking on the proofs of Lemma 2 and Theorem 1, let us single out a set of preliminary facts that shall be used throughout.

According to the definition of r_t (cf. (31)), it is readily seen that

$$\begin{aligned}
\mathbb{E}[r_\delta(Y_{1-\delta}) \,|\, Y_t = y] &= \mathbb{E}[r_\delta(X_\delta) \,|\, X_{1-t} = y] \\
&= \iint r(x_0) p_{X_0 \,|\, X_\delta}(x_0 \,|\, y') p_{X_\delta \,|\, X_{1-t}}(y' \,|\, y) \mathrm{d}y' \mathrm{d}x_0 \\
&= \int r(x_0) p_{X_0 \,|\, X_{1-t}}(x_0 \,|\, y) \mathrm{d}x_0 = \mathbb{E}[r(X_0) \,|\, X_{1-t} = y] = r_{1-t}(y), \qquad (50a) \\
r_{1-t}(y) &= \frac{1}{p_{X_{1-t}}(y)} \int r(x_0) p_{X_{1-t} \,|\, X_0}(y \,|\, x_0) p_{X_0}(x_0) \mathrm{d}x_0 \\
&= \frac{\mathbb{E}[r(X_0)] p_{X_{1-t}^{r\text{-wt}}}(y)}{p_{X_{1-t}}(y)}, \qquad (50b)
\end{aligned}$$

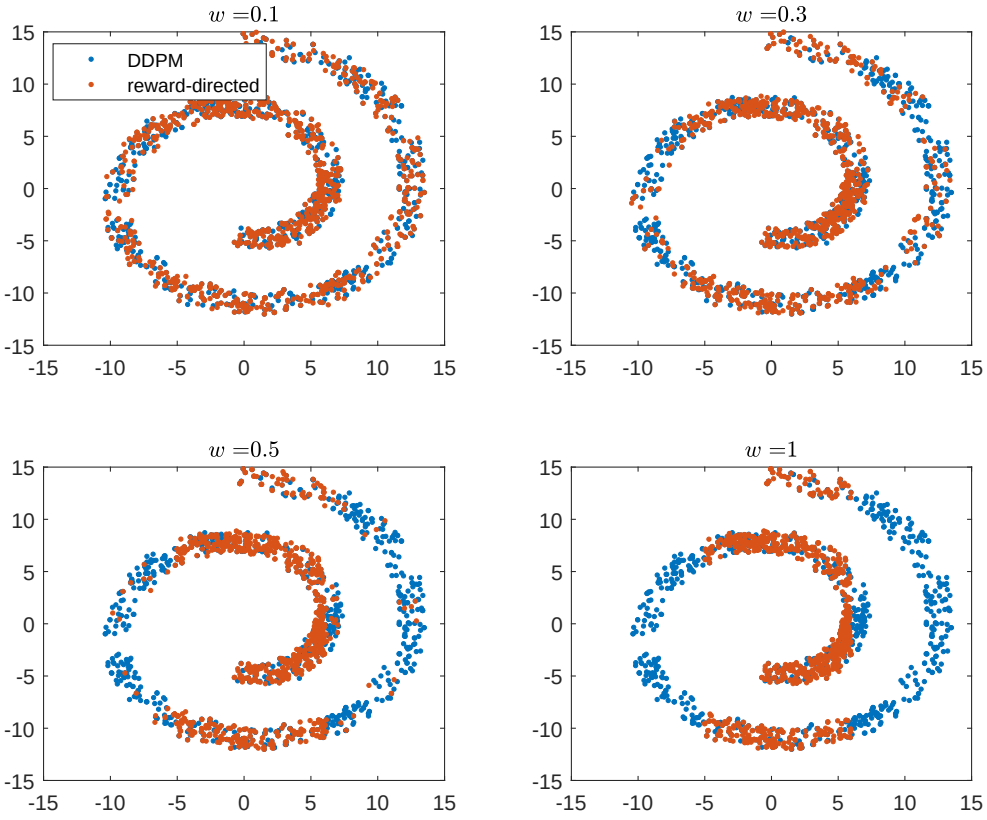

Figure 5: Generated samples of reward-guided sampler with specific values of w and DDPM ($w = 0$).

where we have made use of the definition of $p_{X_{1-t}^{r\text{-wt}}}$ (see (25)). Property (50b) also implies that the function $r_{1-t}(y)$ is infinitely differentiable w.r.t. (y, t) for any $0 < t < 1$, as long as $\mathbb{E}[r(X_0)] < \infty$.

Next, we present a key lemma concerning some useful property of $r_t(\cdot)$, whose proof can be found in Section C.1.

Lemma 3. *Consider function $r_t(\cdot)$ defined in (31). For any $\varepsilon > 0$ and any $0 \le \tau \le t \le 1 - \varepsilon$, we have*

$$r_t(x) = \mathbb{E}\big[r_\tau(X_\tau) \,|\, X_t = x\big]. \tag{51a}$$

Equivalently, for any $0 < \varepsilon \le \tau \le t \le 1$, we have

$$r_{1-\tau}(y) = \mathbb{E}\big[r_{1-t}(Y_t) \,|\, Y_\tau = y\big]. \tag{51b}$$

Here, X_t and Y_t are defined in (10).

Finally, we establish a few key properties of r_t. Let $R < \infty$ be some quantity such that

$$\mathbb{P}(\|X_0\|_2 < R) > \frac{1}{2} \quad \text{and} \quad \mathbb{P}(\|X_0^{r\text{-wt}}\|_2 < R) > \frac{1}{2}. \tag{52}$$

Then there exist some quantities $C_{t,d,k,R} > 0$ (resp. $C_{t,d,k,R,\mathbb{E}[r(X_0)]} > 0$), depending only on t, d, k, R (resp. $t, k, R, \mathbb{E}[r(X_0)]$), such that for J_{1-t} and r_{1-t} defined in (31), the following bounds hold:

$$\|\nabla^k \log p_{X_{1-t}}(y)\|_{\mathrm{F}}^2 \le \exp\big(C_{t,d,k,R}(1 + \|y\|_2^2)\big), \tag{53a}$$

$$\left\|\nabla^k r_{1-t}(y)\right\|_{\mathrm{F}}^2 \le \exp\left(C_{t,d,k,R,\mathbb{E}[r(X_0)]}(1+\|y\|_2^2)\right), \tag{53b}$$

$$\left\|\frac{\partial^k r_{1-t}(y)}{\partial t^k}\right\|_{\mathrm{F}}^2 \le \exp\left(C_{t,d,k,R,\mathbb{E}[r(X_0)]}(1+\|y\|_2^2)\right), \tag{53c}$$

$$\left\|\frac{\partial \nabla \log p_{X_{1-t}}(y)}{\partial t}\right\|_2^2 \le \exp\left(C_{t,d,1,R}(1+\|y\|_2^2)\right), \tag{53d}$$

where $\nabla^k r_t(y)$ denotes the k-th order derivative of function $r_t(y)$ with respect to y, and $\|\cdot\|_{\mathrm{F}}^2$ denotes the sum of squares of all entries. The proof is postponed to Appendix C.2.

6.2 Proof of Lemma 2

Denote by Y_t the continuous process in (10b). Since processes $\{Y_s\}$ (cf. (10b)) and $\{Y_s^g\}$ (cf. (23)) follow exactly the same SDE up to time $s = t$, it follows that

$$Y_t \overset{\mathrm{d}}{=} Y_t^g.$$

Conditioned on $Y_t = y$ and the Brownian motion $\{B_s\}$ for $s \in [t, t+\Delta t]$, we can quantify the difference between Y_s^g and Y_s by comparing (10b) with (23):

$$Y_s^g - Y_s = \frac{1}{2}\int_t^s (Y_u^g - Y_u)\frac{\mathrm{d}u}{u} + \int_t^s \left(\nabla \log p_{X_{1-u}}(Y_u^g) - \nabla \log p_{X_{1-u}}(Y_u)\right)\frac{\mathrm{d}u}{u} + g\log\frac{s}{t} \tag{54}$$

for all $t < s \le t+\Delta t$. In particular, taking $s = t+\Delta t$ and applying the triangle inequality yield

$$\begin{aligned}&\left\|Y_{t+\Delta t}^g - Y_{t+\Delta t} - g\log\frac{t+\Delta t}{t}\right\|_2 \\ &\quad\le \frac{1}{2}\int_t^{t+\Delta t} \left\|Y_u^g - Y_u\right\|_2 \frac{\mathrm{d}u}{u} + \int_t^{t+\Delta t} \left\|\nabla \log p_{X_{1-u}}(Y_u^g) - \nabla \log p_{X_{1-u}}(Y_u)\right\|_2 \frac{\mathrm{d}u}{u}.\end{aligned} \tag{55}$$

Before proceeding, we find it convenient to introduce the following event:

$$\mathcal{E} := \left\{\int_t^{t+\Delta t} \left(\|Y_s\|_2^2 + \|Y_s^g\|_2^2\right)\mathrm{d}s \le \widetilde{C}\sqrt{\Delta t}\right\}, \tag{56}$$

where $\widetilde{C} > 0$ is some sufficiently large constant. When Δt is taken to be sufficiently small, we claim that this event $\mathcal{E}$ satisfies

$$Y_{t+\Delta t}^g \mathbb{1}(\mathcal{E}) = \left(Y_{t+\Delta t} + g\frac{\Delta t}{t} + \widetilde{R}_w\right)\mathbb{1}(\mathcal{E}) \qquad \text{for some residual } \widetilde{R}_w \text{ with } \|\widetilde{R}_w\|_2 = o(\Delta t), \tag{57a}$$

and for any given y obeying $\|y\|_2 < \infty$,

$$\mathbb{P}\left(\mathcal{E}^{\mathrm{c}} \,|\, Y_t^g = Y_t = y\right) = O\left((\Delta t)^{2+1/\varepsilon}\right), \tag{57b}$$

where ε is a small constant such that $\mathbb{E}[r(X_0)^{1+\varepsilon}] < \infty$. Here, we focus solely on the scaling in Δt, with (d, t, y) and the distribution of X_0 regarded as fixed. The proofs of properties (57) are deferred to Appendix C.3. With properties (57) in place, we can proceed to the main steps of our proof.

Step 1: decomposing the reward difference based on event $\mathcal{E}$. To begin with, we observe that

$$\begin{aligned}\mathbb{E}\left[r_\delta(Y_{1-\delta}^g) \,|\, Y_t^g = y\right] &= \int r_\delta(y_{1-\delta}) p_{Y_{1-\delta}^g \,|\, Y_t^g}(y_{1-\delta} \,|\, y)\mathrm{d}y_{1-\delta} \\ &\overset{\text{(a)}}{=} \iint r(y_1) p_{X_0 \,|\, X_\delta}(y_1 \,|\, y_{1-\delta}) p_{Y_{1-\delta}^g \,|\, Y_t^g}(y_{1-\delta} \,|\, y)\mathrm{d}y_1\mathrm{d}y_{1-\delta} \\ &\overset{\text{(b)}}{=} \iint r(y_1) p_{Y_1 \,|\, Y_{1-\delta}}(y_1 \,|\, y_{1-\delta}) p_{Y_{1-\delta}^g \,|\, Y_t^g}(y_{1-\delta} \,|\, y)\mathrm{d}y_1\mathrm{d}y_{1-\delta}\end{aligned}$$

$$
\begin{aligned}
&\overset{\text{(c)}}{=} \iint r(y_1) p_{Y_1^g \,|\, Y_{1-\delta}^g}(y_1 \,|\, y_{1-\delta}) p_{Y_{1-\delta}^g \,|\, Y_t^g}(y_{1-\delta} \,|\, y) \mathrm{d}y_1 \mathrm{d}y_{1-\delta} = \int r(y_1) p_{Y_1^g \,|\, Y_t^g}(y_1 \,|\, y) \mathrm{d}y_1 \\
&\overset{\text{(d)}}{=} \iint r(y_1) p_{Y_1^g \,|\, Y_{t+\Delta t}^g}(y_1 \,|\, y') p_{Y_{t+\Delta t}^g \,|\, Y_t^g}(y' \,|\, y) \mathrm{d}y_1 \mathrm{d}y' \\
&\overset{\text{(e)}}{=} \int r_{1-t-\Delta t}(y') p_{Y_{t+\Delta t}^g \,|\, Y_t^g}(y' \,|\, y) \mathrm{d}y' = \mathbb{E}\big[r_{1-t-\Delta t}(Y_{t+\Delta t}^g) \,|\, Y_t^g = y\big],
\end{aligned}
$$

where (a) results from the definition of $r_\delta(\cdot)$ in (22), (b) holds since $\{Y_s\}$ (cf. (10b)) is the reverse process of $\{X_t\}$ obeying $Y_t \overset{\mathrm{d}}{=} X_{1-t}$, (c) follows since, by construction, $\{Y_s^g\}$ (cf. (23)) and $\{Y_s\}$ (cf. (10b)) follow the same SDE during the interval $[1-\delta, 1]$, (d) inserts the random vector $Y_{t+\Delta t}^g$ into the integral, and (e) is valid since, according to the definition of r_t (cf. (31)),

$$
\begin{aligned}
r_{1-t-\Delta t}(y') &\coloneqq \int r(y_1) p_{X_0 \,|\, X_{1-t-\Delta t}}(y_1 \,|\, y') \mathrm{d}y_1 = \int r(y_1) p_{Y_1 \,|\, Y_{t+\Delta t}}(y_1 \,|\, y') \mathrm{d}y_1 \\
&= \int r(y_1) p_{Y_1^g \,|\, Y_{t+\Delta t}^g}(y_1 \,|\, y') \mathrm{d}y_1. \qquad (58)
\end{aligned}
$$

Repeating the same arguments gives

$$
\mathbb{E}\big[r_\delta(Y_{1-\delta}) \,|\, Y_t = y\big] = \int r_{1-t-\Delta t}(y') p_{Y_{t+\Delta t} \,|\, Y_t}(y' | y) \mathrm{d}y' = \mathbb{E}\big[r_{1-t-\Delta t}(Y_{t+\Delta t}) \,|\, Y_t = y\big],
$$

which can be clearly seen since $p_{Y_1^g \,|\, Y_{t+\Delta t}^g}(y_1 \,|\, y') = p_{Y_1 \,|\, Y_{t+\Delta t}}(y_1 \,|\, y')$ under our construction. As a result,

$$
\begin{aligned}
&\mathbb{E}\big[r_\delta(Y_{1-\delta}^g) \,|\, Y_t^g = y\big] - \mathbb{E}\big[r_\delta(Y_{1-\delta}) \,|\, Y_t = y\big] = \mathbb{E}\big[r(Y_1^g) \,|\, Y_t^g = y\big] - \mathbb{E}\big[r(Y_1) \,|\, Y_t = y\big] \\
&\quad = \mathbb{E}\big[\big(r(Y_1^g) - r(Y_1)\big) \mathbb{1}(\mathcal{E}) \,|\, Y_t^g = Y_t = y\big] + \mathbb{E}\big[\big(r(Y_1^g) - r(Y_1)\big) \mathbb{1}(\mathcal{E}^{\mathrm{c}}) \,|\, Y_t^g = Y_t = y\big]. \qquad (59)
\end{aligned}
$$

Regarding the last term in (59), it follows from Hölder's inequality that

$$
\begin{aligned}
&\frac{1}{\Delta t}\Big|\mathbb{E}\big[\big(r(Y_1^g) - r(Y_1)\big) \mathbb{1}(\mathcal{E}^{\mathrm{c}}) \,|\, Y_t^g = Y_t = y\big]\Big| \\
&\quad \le \frac{1}{\Delta t}\mathbb{E}\big[\big(r(Y_1^g) + r(Y_1)\big) \mathbb{1}(\mathcal{E}^{\mathrm{c}}) \,|\, Y_t^g = Y_t = y\big] \\
&\quad \overset{\text{(a)}}{\le} \frac{1}{\Delta t}\left(\mathbb{E}\big[r^{1+\varepsilon}(Y_1^g) \,|\, Y_t^g = y\big] + \mathbb{E}\big[r^{1+\varepsilon}(Y_1) \,|\, Y_t = y\big]\right)^{1/(1+\varepsilon)} \left(\mathbb{P}(\mathcal{E}^{\mathrm{c}} \,|\, Y_t^g = Y_t = y)\right)^{\varepsilon/(1+\varepsilon)} \\
&\quad \overset{\text{(b)}}{\lesssim} (\Delta t)^{\varepsilon/(1+\varepsilon)} \left(\mathbb{E}\big[r^{1+\varepsilon}(Y_1^g) \,|\, Y_t^g = y\big] + \mathbb{E}\big[r^{1+\varepsilon}(Y_1) \,|\, Y_t = y\big]\right)^{1/(1+\varepsilon)}, \qquad (60)
\end{aligned}
$$

where (a) results from the Hölder inequality, and (b) follows from property (57b). Thus, it boils down to controlling the first term in (59). It is observed that, as $\Delta t \to 0$,

$$
\begin{aligned}
&\lim_{\Delta t \to 0} \mathbb{E}\big[r^{1+\varepsilon}(Y_1^g) \,|\, Y_t^g = y\big] = \mathbb{E}\big[r^{1+\varepsilon}(Y_1) \,|\, Y_t = y\big] = \mathbb{E}\big[r^{1+\varepsilon}(X_0) \,|\, X_{1-t} = y\big] \\
&\quad = \int r^{1+\varepsilon}(x_0) p_{X_0 \,|\, X_{1-t}}(x_0 \,|\, y) \mathrm{d}x_0 = \int r^{1+\varepsilon}(x_0) p_{X_0}(x_0) \frac{p_{X_{1-t} \,|\, X_0}(y \,|\, x_0)}{p_{X_{1-t}}(y)} \mathrm{d}x_0 \\
&\quad \le \frac{\mathbb{E}[r^{1+\varepsilon}(X_0)]}{(2\pi(1-t))^{d/2} p_{X_{1-t}}(y)} < \infty,
\end{aligned}
$$

where we have used the assumption that $\mathbb{E}[r^{1+\varepsilon}(X_0)] < \infty$. Substitution into (60) yields

$$
\frac{1}{\Delta t}\mathbb{E}\big[\big(r(Y_1^g) - r(Y_1)\big) \mathbb{1}(\mathcal{E}^{\mathrm{c}}) \,|\, Y_t^g = Y_t = y\big] \to 0, \qquad \text{as} \quad \Delta t \to 0.
$$

Taken together with (59), we see that

$$
\begin{aligned}
&\lim_{\Delta t \to 0} \frac{1}{\Delta t}\left(\mathbb{E}\big[r_\delta(Y_{1-\delta}^g) \,|\, Y_t^g = y\big] - \mathbb{E}\big[r_\delta(Y_{1-\delta}) \,|\, Y_t = y\big]\right) = \lim_{\Delta t \to 0} \frac{1}{\Delta t}\mathbb{E}\big[\big(r(Y_1^g) - r(Y_1)\big) \mathbb{1}(\mathcal{E}) \,|\, Y_t^g = Y_t = y\big] \\
&\quad = \lim_{\Delta t \to 0} \frac{1}{\Delta t}\left(\mathbb{E}[r_{1-t-\Delta t}(Y_{t+\Delta t}^g) \mathbb{1}(\mathcal{E}) \,|\, Y_t^g = y] - \mathbb{E}[r_{1-t-\Delta t}(Y_{t+\Delta t}) \mathbb{1}(\mathcal{E}) \,|\, Y_t = y]\right). \qquad (61)
\end{aligned}
$$

Step 2: computing the limit in (61). In view of (57a) and (61), as $\Delta t \to 0$, we have

$$\begin{aligned}
&\frac{1}{\Delta t}\left(\mathbb{E}[r_\delta(Y^g_{1-\delta}) \,|\, Y^g_t = y] - \mathbb{E}[r_\delta(Y_{1-\delta}) \,|\, Y_t = y]\right) \\
&\quad = \frac{1}{\Delta t}\mathbb{E}\left[\left(r_{1-t-\Delta t}\left(Y_{t+\Delta t} + g\frac{\Delta t}{t} + \widetilde{R}_w\right) - r_{1-t-\Delta t}(Y_{t+\Delta t})\right)\mathbb{1}(\mathcal{E}) \,|\, Y_t = y\right] \\
&\quad \overset{\text{(a)}}{=} \mathbb{E}\left[\left\langle \frac{g}{t} + \frac{\widetilde{R}_w}{\Delta t}, \nabla r_{1-t-\Delta t}\left(Y_{t+\Delta t}\right)\right\rangle \mathbb{1}(\mathcal{E}) \,|\, Y_t = y\right] \\
&\qquad + O(\Delta t)\mathbb{E}_{Y_{t+\Delta t}\,|\,Y_t}\left[\|\nabla^2 r_{1-t-\Delta t}\left(Y_{t+\Delta t}\right)\| \,\mathbb{1}(\mathcal{E}) \,|\, Y_t = y\right],
\end{aligned} \tag{62}$$

where (a) uses $\|\widetilde{R}_w\|_2 \,\mathbb{1}(\mathcal{E}) = o(\Delta t)$ as asserted by (57a). To proceed, we claim that the following inequalities hold, whose proof is postponed to Appendix C.4:

$$\mathbb{E}\left[\|\nabla r_{1-t-\Delta t}\left(Y_{t+\Delta t}\right)\|_2^{1+\varepsilon} \,|\, Y_t = y\right] < \infty, \tag{63a}$$

$$\mathbb{E}\left[\|\nabla^2 r_{1-t-\Delta t}\left(Y_{t+\Delta t}\right)\| \,|\, Y_t = y\right] < \infty. \tag{63b}$$

With inequalities (63) and the property $\|\widetilde{R}_w\|_2 \,\mathbb{1}(\mathcal{E}) = o(\Delta t)$ (cf. (57a)) in place, we can see from (62) that

$$\begin{aligned}
&\lim_{\Delta t\to 0}\frac{1}{\Delta t}\left(\mathbb{E}[r_\delta(Y^g_{1-\delta}) \,|\, Y^g_t = y] - \mathbb{E}[r_\delta(Y_{1-\delta}) \,|\, Y_t = y]\right) \\
&\quad = \lim_{\Delta t\to 0}\left\langle \frac{g}{t}, \mathbb{E}\left[\nabla r_{1-t-\Delta t}\left(Y_{t+\Delta t}\right)\mathbb{1}(\mathcal{E}) \,|\, Y_t = y\right]\right\rangle \\
&\quad \overset{\text{(a)}}{=} \lim_{\Delta t\to 0}\left\langle \frac{g}{t}, \mathbb{E}\left[\nabla r_{1-t-\Delta t}\left(Y_{t+\Delta t}\right) \,|\, Y_t = y\right]\right\rangle.
\end{aligned} \tag{64}$$

Here, to justify the validity of (a), we invoke (63a) and (57b) to show that

$$\begin{aligned}
&\left|\left\langle \frac{g}{t}, \mathbb{E}\left[\nabla r_{1-t-\Delta t}\left(Y_{t+\Delta t}\right)\mathbb{1}(\mathcal{E}) \,|\, Y_t = y\right]\right\rangle - \left\langle \frac{g}{t}, \mathbb{E}\left[\nabla r_{1-t-\Delta t}\left(Y_{t+\Delta t}\right) \,|\, Y_t = y\right]\right\rangle\right| \\
&= \left|\left\langle \frac{g}{t}, \mathbb{E}\left[\nabla r_{1-t-\Delta t}\left(Y_{t+\Delta t}\right)\mathbb{1}(\mathcal{E}^{\mathrm{c}}) \,|\, Y_t = y\right]\right\rangle\right| \\
&\le \frac{\|g\|_2}{t}\mathbb{E}\left[\|\nabla r_{1-t-\Delta t}\left(Y_{t+\Delta t}\right)\|_2 \,\mathbb{1}(\mathcal{E}^{\mathrm{c}}) \,|\, Y_t = y\right] \\
&\le \frac{\|g\|_2}{t}\mathbb{E}\left[\|\nabla r_{1-t-\Delta t}\left(Y_{t+\Delta t}\right)\|_2^{1+\varepsilon} \,|\, Y_t = y\right]^{\frac{1}{1+\varepsilon}} \mathbb{P}\big(\mathcal{E}^{\mathrm{c}} \,|\, Y_t = y\big)^{\frac{\varepsilon}{1+\varepsilon}} \\
&= O\big((\Delta t)^{\frac{1+2\varepsilon}{1+\varepsilon}}\big) \to 0, \qquad\qquad \text{as } \Delta t \to 0,
\end{aligned}$$

where the first inequality comes from the Cauchy-Schwarz inequality, and the second inequality comes from the Hölder inequality.

In order to complete the proof of this lemma, it comes down to establishing that

$$\begin{aligned}
&\lim_{\Delta\to 0}\mathbb{E}\left[\nabla r_{1-t-\Delta t}\left(Y_{t+\Delta t}\right) \,|\, Y_t = y\right] \\
&\quad = \lim_{\Delta\to 0}\mathbb{E}\big[r_\delta(Y_{1-\delta}) \,|\, Y_t = y\big]\big(\nabla \log p_{X^{\text{r-wt}}_{1-t}}(y) - \nabla \log p_{X_{1-t}}(y)\big),
\end{aligned} \tag{65}$$

which forms the content of the next step.

Step 3: connecting the limit (64) with score difference. Towards this, we decompose the expectation of interest in (64) as

$$\begin{aligned}
&\mathbb{E}\left[\nabla r_{1-t-\Delta t}\left(Y_{t+\Delta t}\right) \,|\, Y_t = y\right] \\
&\quad = \int \nabla r_{1-t-\Delta t}\left(y'\right) p_{X_{1-t-\Delta t}\,|\,X_{1-t}}(y' \,|\, y)\mathrm{d}y'
\end{aligned}$$

$$= \iint r(x_0)\nabla_{y'} p_{X_0 \mid X_{1-t-\Delta t}}(x_0 \mid y') p_{X_{1-t-\Delta t} \mid X_{1-t}}(y' \mid y)\mathrm{d}x_0\mathrm{d}y'$$
$$\overset{\text{(a)}}{=} \iint r(x_0)\nabla_{y'} \log p_{X_0 \mid X_{1-t-\Delta t}}(x_0 \mid y') p_{X_0 \mid X_{1-t-\Delta t}}(x_0 \mid y') p_{X_{1-t-\Delta t} \mid X_{1-t}}(y' \mid y)\mathrm{d}x_0\mathrm{d}y', \tag{66}$$

where (a) follows since $\nabla_{y'} p_{X_0 \mid X_{1-t-\Delta t}}(x_0 \mid y') = p_{X_0 \mid X_{1-t-\Delta t}}(x_0 \mid y')\nabla_{y'} \log p_{X_0 \mid X_{1-t-\Delta t}}(x_0 \mid y')$. Moreover, the gradient $\nabla_{y'} \log p_{X_0 \mid X_{1-t-\Delta t}}(x_0 \mid y')$ can be further decomposed using the Bayes rule as

$$\begin{aligned}\nabla_{y'} p_{X_0 \mid X_{1-t-\Delta t}}(x_0 \mid y') &= \nabla_{y'} \log \frac{p_{X_{1-t-\Delta t} \mid X_0}(y' \mid x_0) p_{X_0}(x_0)}{p_{X_{1-t-\Delta t}}(y')} \\ &= \nabla_{y'} \log p_{X_{1-t-\Delta t} \mid X_0}(y' \mid x_0) - \nabla \log p_{X_{1-t-\Delta t}}(y'). \end{aligned} \tag{67}$$

Substituting this identity into (66), we obtain

$$\begin{aligned}&\mathbb{E}\left[\nabla r_{1-t-\Delta t}\left(Y_{t+\Delta t}\right) \mid Y_t = y\right] \\ &= \underbrace{\iint r\left(x_0\right) p_{X_0 \mid X_{1-t-\Delta t}}(x_0 \mid y') p_{X_{1-t-\Delta t} \mid X_{1-t}}(y' \mid y)\nabla_{y'} \log p_{X_{1-t-\Delta t} \mid X_0}(y' \mid x_0)\mathrm{d}x_0\mathrm{d}y'}_{=: \mathcal{D}_1} \\ &\quad - \underbrace{\iint r\left(x_0\right) p_{X_0 \mid X_{1-t-\Delta t}}(x_0 \mid y') p_{X_{1-t-\Delta t} \mid X_{1-t}}(y' \mid y)\nabla \log p_{X_{1-t-\Delta t}}(y')\mathrm{d}x_0\mathrm{d}y'}_{=: \mathcal{D}_2}. \end{aligned} \tag{68}$$

Below we shall calculate these two terms $\mathcal{D}_1$ and $\mathcal{D}_2$ separately.

Step 4: analysis of the term $\mathcal{D}_1$ in (68). Towards this, we start with the integral in $\mathcal{D}_1$ with respect to x_0 for a fixed y'; namely, we would like to evaluate

$$\mathcal{I}(y') \coloneqq \int r\left(x_0\right) p_{X_0 \mid X_{1-t-\Delta t}}(x_0 \mid y')\nabla_{y'} \log p_{X_{1-t-\Delta t} \mid X_0}(y' \mid x_0)\mathrm{d}x_0,$$

which satisfies

$$\mathcal{D}_1 = \int \mathcal{I}(y') p_{X_{1-t-\Delta t} \mid X_{1-t}}(y' \mid y)\mathrm{d}y'.$$

By virtue of the equation $\nabla_{y'} \log p_{X_{1-t-\Delta t} \mid X_0}(y' \mid x_0) = \frac{\nabla_{y'} p_{X_{1-t-\Delta t} \mid X_0}(y' \mid x_0)}{p_{X_{1-t-\Delta t} \mid X_0}(y' \mid x_0)}$, we can simplify

$$\begin{aligned}\mathcal{I}(y') &= \int r\left(x_0\right) p_{X_0 \mid X_{1-t-\Delta t}}(x_0 \mid y') \frac{\nabla_{y'} p_{X_{1-t-\Delta t} \mid X_0}(y' \mid x_0)}{p_{X_{1-t-\Delta t} \mid X_0}(y' \mid x_0)}\mathrm{d}x_0 \\ &= \int r\left(x_0\right) \nabla_{y'} p_{X_{1-t-\Delta t} \mid X_0}(y' \mid x_0) \frac{p_{X_0}(x_0)}{p_{X_{1-t-\Delta t}}(y')}\mathrm{d}x_0 \\ &= \frac{1}{p_{X_{1-t-\Delta t}}(y')}\nabla_{y'} \int r\left(x_0\right) p_{X_{1-t-\Delta t} \mid X_0}(y' \mid x_0) p_{X_0}(x_0)\mathrm{d}x_0, \end{aligned} \tag{69}$$

where the second line comes from the Bayes rule. In view of the definition (25) of $p_{X_{1-t}^{r\text{-wt}}}$, we can further simplify the above integral as:

$$\begin{aligned}\mathcal{I}(y') &= \frac{1}{p_{X_{1-t-\Delta t}}(y')}\nabla_{y'} \int r(x_0) p_{X_{1-t-\Delta t} \mid X_0}(y' \mid x_0) p_{X_0}(x_0)\mathrm{d}x_0 \\ &= \frac{\mathbb{E}[r(X_0)]\nabla p_{X_{1-t-\Delta t}^{r\text{-wt}}}(y')}{p_{X_{1-t-\Delta t}}(y')} = \frac{\mathbb{E}[r(X_0)] p_{X_{1-t-\Delta t}^{r\text{-wt}}}(y')\nabla \log p_{X_{1-t-\Delta t}^{r\text{-wt}}}(y')}{p_{X_{1-t-\Delta t}}(y')} \\ &= r_{1-t-\Delta t}(y')\nabla \log p_{X_{1-t-\Delta t}^{r\text{-wt}}}(y'), \end{aligned} \tag{70}$$

where the last line arises from (50). Substituting (70) into (69) gives

$$\begin{aligned}\mathcal{D}_1 &= \int r_{1-t-\Delta t}(y')\nabla \log p_{X^{r\text{-wt}}_{1-t-\Delta t}}(y') p_{X_{1-t-\Delta t}\,|\,X_{1-t}}(y'\,|\,y)\mathrm{d}y' \\ &= \int r_{1-t-\Delta t}(y')\nabla \log p_{X^{r\text{-wt}}_{1-t-\Delta t}}(y) p_{X_{1-t-\Delta t}\,|\,X_{1-t}}(y'\,|\,y)\mathrm{d}y' \\ &\quad + \int r_{1-t-\Delta t}(y')\left(\nabla \log p_{X^{r\text{-wt}}_{1-t-\Delta t}}(y') - \nabla \log p_{X^{r\text{-wt}}_{1-t-\Delta t}}(y)\right) p_{X_{1-t-\Delta t}\,|\,X_{1-t}}(y'\,|\,y)\mathrm{d}y' \\ &=: \mathcal{D}_{1,1} + \mathcal{D}_{1,2}. \end{aligned} \tag{71}$$

This leaves us with two terms to control, which we accomplish separately in the sequel.

- With regards to the term $\mathcal{D}_{1,1}$, by virtue of the definition of $r_{1-t}(y)$ (cf. (31)), we have

$$\begin{aligned}\mathcal{D}_{1,1} &= \nabla \log p_{X^{r\text{-wt}}_{1-t-\Delta t}}(y) \int r_{1-t-\Delta t}(y') p_{X_{1-t-\Delta t}\,|\,X_{1-t}}(y'\,|\,y)\mathrm{d}y' = r_{1-t}(y)\nabla \log p_{X^{r\text{-wt}}_{1-t-\Delta t}}(y) \\ &= \mathbb{E}[r_\delta(Y_{1-\delta})\,|\,Y_t = y]\nabla \log p_{X^{r\text{-wt}}_{1-t-\Delta t}}(y), \end{aligned} \tag{72}$$

  where the last line applies (50a).

- With regards to the term $\mathcal{D}_{1,2}$, it follows from the fundamental theorem of calculus that

$$\begin{aligned}\left\|\nabla \log p_{X^{r\text{-wt}}_{1-t-\Delta t}}(y') - \nabla \log p_{X^{r\text{-wt}}_{1-t-\Delta t}}(y)\right\| &\le \max_{0\le\gamma\le 1} \|\nabla^2 \log p_{X^{r\text{-wt}}_{1-t-\Delta t}}(\gamma y' + (1-\gamma)y)\|\, \|y-y'\|_2 \\ &\overset{\text{(a)}}{\le} C_{y,t,d,R}(\|y-y'\|_2^2 + 1)\|y-y'\|_2 \\ &\le C_{y,t,d,R}\|y-y'\|_2^3 + C_{y,t,d,R}\|y-y'\|_2. \end{aligned}$$

  Here, (a) invokes (112) which tells us that

$$\begin{aligned}\max_{0\le\gamma\le 1} \|\nabla^2 \log p_{X^{r\text{-wt}}_{1-t-\Delta t}}(\gamma y' + (1-\gamma)y)\| &\le C \max_{0\le\gamma\le 1} \frac{\|\gamma y' + (1-\gamma)y\|_2^2 + (t+\Delta t)R^2}{(1-t-\Delta t)^2} + \frac{Cd^2}{1-t-\Delta t} \\ &\le C\frac{2\|y'-y\|_2^2 + 2\|y\|_2^2 + (t+\Delta t)R^2}{(1-t-\Delta t)^2} + \frac{Cd^2}{1-t-\Delta t}, \end{aligned}$$

  where C is a sufficiently large constant. As a consequence, we can bound $\mathcal{D}_{1,2}$ as

$$\begin{aligned}\|\mathcal{D}_{1,2}\|_2 &\le \int r_{1-t-\Delta t}\,(y')\, p_{X_{1-t-\Delta t}\,|\,X_{1-t}}(y'\,|\,y) \left\|\nabla \log p_{X^{r\text{-wt}}_{1-t-\Delta t}}(y') - \nabla \log p_{X^{r\text{-wt}}_{1-t-\Delta t}}(y)\right\|_2 \mathrm{d}y' \\ &\le C_{y,t,d,R} \int r_{1-t-\Delta t}\,(y')\, p_{X_{1-t-\Delta t}\,|\,X_{1-t}}(y'\,|\,y) \left(\|y-y'\|_2 + \|y-y'\|_2^3\right) \mathrm{d}y'. \end{aligned} \tag{73}$$

  To continue the bound, we note that $\|y-y'\|_2^n$ for $n = 1, 3$ satisfies

$$\begin{aligned}\|y-y'\|_2^n &\le \left(\sqrt{\frac{t+\Delta t}{t}}\left\|y - \sqrt{\frac{t}{t+\Delta t}}y'\right\|_2 + \left(\sqrt{\frac{t+\Delta t}{t}} - 1\right)\|y\|_2\right)^n \\ &\le 4\left(\left(\frac{t+\Delta t}{t}\right)^{n/2}\left\|y - \sqrt{\frac{t}{t+\Delta t}}y'\right\|_2^n + \left(\sqrt{\frac{t+\Delta t}{t}} - 1\right)^n \|y\|_2^n\right). \end{aligned} \tag{74}$$

  Additionally, it is seen that

$$\begin{aligned}r_{1-t-\Delta t}\,(y')\, p_{X_{1-t-\Delta t}\,|\,X_{1-t}}(y'\,|\,y) &= \int r(x_0) p_{X_0\,|\,X_{1-t-\Delta t}}(x_0\,|\,y') p_{X_{1-t-\Delta t}\,|\,X_{1-t}}(y'\,|\,y)\mathrm{d}x_0 \\ &= \frac{1}{p_{X_{1-t}}(y)} \int r(x_0) p_{X_0}(x_0) p_{X_{1-t-\Delta t}\,|\,X_0}(y'\,|\,x_0) p_{X_{1-t}\,|\,X_{1-t-\Delta t}}(y\,|\,y')\mathrm{d}x_0 \end{aligned}$$

$$
\begin{aligned}
&= \frac{\mathbb{E}[r(X_0)]}{p_{X_{1-t}}(y)} \int p_{X_0^{r\text{-wt}}}(x_0) p_{X_{1-t-\Delta t}^{r\text{-wt}} \,|\, X_0^{r\text{-wt}}}(y' \,|\, x_0) p_{X_{1-t}^{r\text{-wt}} \,|\, X_{1-t-\Delta t}^{r\text{-wt}}}(y \,|\, y') \mathrm{d}x_0 \\
&= \frac{\mathbb{E}[r(X_0)] p_{X_{1-t}^{r\text{-wt}}}(y)}{p_{X_{1-t}}(y)} \int p_{X_0^{r\text{-wt}}, X_{1-t-\Delta t}^{r\text{-wt}} \,|\, X_{1-t}^{r\text{-wt}}}(x_0, y' \,|\, y) \mathrm{d}x_0 \\
&= \frac{\mathbb{E}[r(X_0)] p_{X_{1-t}^{r\text{-wt}}}(y)}{p_{X_{1-t}}(y)} p_{X_{1-t-\Delta t}^{r\text{-wt}} \,|\, X_{1-t}^{r\text{-wt}}}(y' \,|\, y) \\
&= r_{1-t}(y) p_{X_{1-t-\Delta t}^{r\text{-wt}} \,|\, X_{1-t}^{r\text{-wt}}}(y' \,|\, y), \qquad (75)
\end{aligned}
$$

where the last equation uses (50). Substituting (74) and (75) into (73), we obtain

$$
\begin{aligned}
\|\mathcal{D}_{1,2}\| \le\ & 4C_{y,t,d,R} r_{1-t}(y) \left(\frac{t+\Delta t}{t}\right)^{3/2} \\
& \cdot \int \left(\left\| y - \sqrt{\frac{t}{t+\Delta t}} y' \right\|_2 + \left\| y - \sqrt{\frac{t}{t+\Delta t}} y' \right\|_2^3 \right) p_{X_{1-t-\Delta t}^{r\text{-wt}} \,|\, X_{1-t}^{r\text{-wt}}}(y' \,|\, y) \mathrm{d}y' \\
& + 4C_{y,t,d,R} r_{1-t}(y) \left(\sqrt{\frac{t+\Delta t}{t}} - 1 \right)^2 (\|y\|_2 + \|y\|_2^3) \int p_{X_{1-t-\Delta t}^{r\text{-wt}} \,|\, X_{1-t}^{r\text{-wt}}}(y' \,|\, y) \mathrm{d}y'. \qquad (76)
\end{aligned}
$$

Now, we claim that (with the proof postponed to Appendix C.5)

$$
\int \left\| y - \sqrt{\frac{t}{t+\Delta t}} y' \right\|_2 p_{X_{1-t-\Delta t}^{r\text{-wt}} \,|\, X_{1-t}^{r\text{-wt}}}(y' \,|\, y) \mathrm{d}y' \lesssim \sqrt{\frac{\Delta t}{t+\Delta t}} \left(\frac{\|y\|_2 + \sqrt{t} R}{\sqrt{1-t}} + \sqrt{d} \log^{\frac{1}{2}} \frac{1}{\Delta t} \right), \qquad (77a)
$$

$$
\int \left\| y - \sqrt{\frac{t}{t+\Delta t}} y' \right\|_2^3 p_{X_{1-t-\Delta t}^{r\text{-wt}} \,|\, X_{1-t}^{r\text{-wt}}}(y' \,|\, y) \mathrm{d}y' \lesssim \left(\frac{\Delta t}{t+\Delta t} \right)^{3/2} \left(\frac{\|y\|_2^3 + t^{3/2} R^3}{(1-t)^{3/2}} + d^{3/2} \log^{3/2} \frac{1}{\Delta t} \right), \qquad (77b)
$$

and hence both integrals tend to 0 as $\Delta t \to 0$. It is also seen that

$$
\left(\sqrt{\frac{t+\Delta t}{t}} - 1 \right)^2 \le \frac{(\Delta t)^2}{4t^2} \to 0, \quad \text{as } \Delta t \to 0. \qquad (78)
$$

Substituting (77) and (78) into (76), we arrive at

$$
\|\mathcal{D}_{1,2}\|_2 \le C_{y,t,d,R} \sqrt{\Delta t}, \qquad (79)
$$

where $C_{y,t,d,R}$ denotes a quantity depending on y, t, d, and R but independent of Δt.

Taking (72) and (79) together with (71) then yields

$$
\lim_{\Delta t \to 0} \mathcal{D}_1 = \mathbb{E}[r_\delta(Y_{1-\delta}) \,|\, Y_t = y] \lim_{\Delta t \to 0} \nabla \log p_{X_{1-t-\Delta t}^{r\text{-wt}}}(y) = \mathbb{E}[r_\delta(Y_{1-\delta}) \,|\, Y_t = y] \nabla \log p_{X_{1-t}^{r\text{-wt}}}(y), \qquad (80)
$$

where the limit uses the continuity of $\log p_{X_{1-t}}(y)$ w.r.t. t.

Step 5: analysis of the term $\mathcal{D}_2$ in (68). Towards this end, we start with the following decomposition:

$$
\begin{aligned}
\mathcal{D}_2 &= \iint r(x_0) \, p_{X_0 \,|\, X_{1-t-\Delta t}}(x_0 \,|\, y') \mathrm{d}x_0 p_{X_{1-t-\Delta t} \,|\, X_{1-t}}(y' \,|\, y) \nabla \log p_{X_{1-t-\Delta t}}(y') \mathrm{d}y' \\
&\overset{\text{(a)}}{=} \int r_{1-t-\Delta t}(y') \, p_{X_{1-t-\Delta t} \,|\, X_{1-t}}(y' \,|\, y) \nabla \log p_{X_{1-t-\Delta t}}(y') \mathrm{d}y' \\
&= \int r_{1-t-\Delta t}(y') \, p_{X_{1-t-\Delta t} \,|\, X_{1-t}}(y' \,|\, y) \nabla \log p_{X_{1-t-\Delta t}}(y) \mathrm{d}y'
\end{aligned}
$$

$$+\int r_{1-t-\Delta t}(y')\, p_{X_{1-t-\Delta t}\,|\,X_{1-t}}(y'\,|\,y)\left(\nabla \log p_{X_{1-t-\Delta t}}(y') - \nabla \log p_{X_{1-t-\Delta t}}(y)\right)\mathrm{d}y'$$
$$=: \mathcal{D}_{2,1} + \mathcal{D}_{2,2},$$

where (a) applies the definition of $r_{1-t-\Delta t}(\cdot)$ (cf. (31)). This leaves us with two terms to control.

- With regards to the first term $\mathcal{D}_{2,1}$, we can demonstrate that

$$\begin{aligned}\mathcal{D}_{2,1} &= \nabla \log p_{X_{1-t-\Delta t}}(y)\int r_{1-t-\Delta t}(y')\, p_{X_{1-t-\Delta t}\,|\,X_{1-t}}(y'\,|\,y)\mathrm{d}y' \\ &\overset{\text{(a)}}{=} \nabla \log p_{X_{1-t-\Delta t}}(y)\mathbb{E}[r_\delta(Y_{1-\delta})\,|\,Y_t = y] \\ &\to \nabla \log p_{X_{1-t}}(y)\mathbb{E}[r_\delta(Y_{1-\delta})\,|\,Y_t = y], \quad \text{as } \Delta t \to 0, \end{aligned} \tag{81}$$

  where the limit uses the continuity of $\log p_{X_{1-t}}(y)$ w.r.t. t, and in (a) we invoke (70) and (50a) to derive

$$\begin{aligned}\int r_{1-t-\Delta t}(y')\, p_{X_{1-t-\Delta t}\,|\,X_{1-t}}(y'\,|\,y)\mathrm{d}y' &= \frac{1}{p_{X_{1-t}}(y)}\int r_{1-t-\Delta t}(y')p_{X_{1-t}\,|\,X_{1-t-\Delta t}}(y\,|\,y')p_{X_{1-t-\Delta t}}(y')\mathrm{d}y' \\ &= \frac{1}{p_{X_{1-t}}(y)}\int r(x_0)p_{X_{1-t}\,|\,X_0}(y\,|\,x_0)p_{X_0}(x_0)\mathrm{d}x_0 \\ &= \int r_0(x)\, p_{X_0\,|\,X_{1-t}}(x\,|\,y)\mathrm{d}x = r_{1-t}(y) = \mathbb{E}_{Y_{1-\delta}}[r_\delta(Y_{1-\delta})\,|\,Y_t = y].\end{aligned}$$

- Next, let us turn attention to the term $\mathcal{D}_{2,2}$. By virtue of (111), we have

$$\begin{aligned}\left\|\nabla \log p_{X_{1-t-\Delta t}}(y') - \nabla \log p_{X_{1-t-\Delta t}}(y)\right\| &\le \max_{0\le\gamma\le 1}\|\nabla^2 \log p_{X_{1-t-\Delta t}}(\gamma y' + (1-\gamma)y)\|\,\|y - y'\|_2 \\ &\le C_{y,t,d,R}(\|y-y'\|_2^2 + 1)\|y-y'\|_2 \\ &\le C_{y,t,d,R}\|y-y'\|_2^3 + C_{y,t,d,R}\|y-y'\|_2.\end{aligned}$$

  Repeating the same argument as in the analysis of $\mathcal{D}_{1,2}$, we reach

$$\|\mathcal{D}_{2,2}\|_2 \le C_{y,t,d,R}\sqrt{\Delta t}. \tag{82}$$

Combining (81) and (82) then yields

$$\lim_{\Delta t\to 0}\mathcal{D}_2 = \nabla \log p_{X_{1-t}}(y)\mathbb{E}_{Y_{1-\delta}}[r_\delta(Y_{1-\delta})\,|\,Y_t = y]. \tag{83}$$

Step 6: putting all pieces together. Combining (80) and (83) with (68) results in

$$\begin{aligned}&\lim_{\Delta t\to 0}\mathbb{E}_{Y_{t+\Delta t}\,|\,Y_t}\left[\nabla r_{1-t-\Delta t}(Y_{t+\Delta t})\,\mathbb{1}(\mathcal{E})\,|\,Y_t = y\right] \\ &\quad = \mathbb{E}_{Y_{1-\delta}}[r_\delta(Y_{1-\delta})\,|\,Y_t = y]\left(\nabla \log p_{X_{1-t}^{r\text{-wt}}}(y) - \nabla \log p_{X_{1-t}}(y)\right).\end{aligned}$$

Substitution into (64) then yields

$$\begin{aligned}&\lim_{\Delta t\to 0}\frac{1}{\Delta t}\left(\mathbb{E}_{Y_{1-\delta}^g}[r_\delta(Y_{1-\delta}^g)\,|\,Y_t^g = y] - \mathbb{E}_{Y_{1-\delta}}[r_\delta(Y_{1-\delta})\,|\,Y_t = y]\right) \\ &\quad = \frac{\mathbb{E}_{Y_{1-\delta}}[r_\delta(Y_{1-\delta})\,|\,Y_t = y]}{t}\left\langle g, \nabla \log p_{X_{1-t}^{r\text{-wt}}}(y) - \nabla \log p_{X_{1-t}}(y)\right\rangle\end{aligned} \tag{84}$$

as claimed.

6.3 Proof of Theorems 1 and 2

In this subsection, we present the proof of our main result in Theorem 1. Note that Theorem 2 can be established using the same argument as for Theorem 1, and hence we omit the proof of Theorem 2 for the sake of brevity.

Step 1: computing the limit of reward improvement. We claim that there exists some quantity $\delta_R > 0$ — depending only on y_t, t, and the distributional property of X_0 — such that for any $\delta < \delta_R$, the following property holds:

$$\begin{aligned} 0 &= \frac{1}{\delta}\Big\{\mathbb{E}\big[r_{1-t-\delta}(Y_{t+\delta}) - r_{1-t}(Y_t)\,|\,Y_t = y_t\big]\Big\} \\ &= \frac{\partial r_{1-t}(y)}{\partial t}\Big|_{y=y_t} + \frac{1}{2t}\mathsf{Tr}\big(\nabla^2 r_{1-t}(y_t)\big) + \Big\langle \nabla r_{1-t}(y_t), \Big(\frac{1}{2}y_t + \nabla \log p_{X_{1-t}}(y_t)\Big)\frac{1}{t}\Big\rangle + O(\delta), \end{aligned} \tag{85}$$

whose proof is postponed to Section C.6. Similarly, it holds that

$$\begin{aligned} &\frac{1}{\delta}\Big\{\mathbb{E}\big[r_{1-t-\delta}(Y^w_{t+\delta}) - r_{1-t}(Y^w_t)\,|\,Y^w_t = y_t\big]\Big\} \\ &\quad= \frac{\partial r_{1-t}(y)}{\partial t}\Big|_{y=y_t} + \frac{1}{2t}\mathsf{Tr}\big(\nabla^2 r_{1-t}(y_t)\big) + \Big\langle \nabla r_{1-t}(y_t), \Big(\frac{1}{2}y_t + \nabla \log p_{X_{1-t}}(y_t) + w g^\star_t(y_t)\Big)\frac{1}{t}\Big\rangle + O(\delta), \end{aligned} \tag{86}$$

where $g^\star_t$ denotes the following guidance term:

$$g^\star_t(y_t) \coloneqq \nabla \log p_{X^{r\text{-wt}}_{1-t}}(y_t) - \nabla \log p_{X_{1-t}}(y_t). \tag{87}$$

Comparing the above two relations leads to

$$\begin{aligned} &\mathbb{E}\big[r_{1-t-\delta}(Y^w_{t+\delta})\,|\,Y^w_t = y_t\big] - \mathbb{E}\big[r_{1-t}(Y^w_t)\,|\,Y^w_t = y_t\big] \\ &\quad= \Big(\mathbb{E}\big[r_{1-t-\delta}(Y^w_{t+\delta})\,|\,Y^w_t = y_t\big] - r_{1-t}(y_t)\Big) - \Big(\mathbb{E}\big[r_{1-t-\delta}(Y_{t+\delta})\,|\,Y_t = y_t\big] - r_{1-t}(y_t)\Big) \\ &\quad= \delta\frac{w}{t}\big\langle \nabla r_{1-t}(y_t), g^\star_t(y_t)\big\rangle + O(\delta^2). \end{aligned} \tag{88}$$

Step 2: computing the gradient of conditional rewards. In order to further evaluate (88), let us calculate the gradient of interest as follows

$$\begin{aligned} \nabla r_{1-t}(y) = \nabla_y \int r(x_0) p_{X_0\,|\,X_{1-t}}(x_0\,|\,y)\mathrm{d}x_0 &\overset{\text{(a)}}{=} \int r(x_0) p_{X_0\,|\,X_{1-t}}(x_0\,|\,y)\nabla_y \log p_{X_0\,|\,X_{1-t}}(x_0\,|\,y)\mathrm{d}x_0 \\ &\overset{\text{(b)}}{=} \int r(x_0) p_{X_0\,|\,X_{1-t}}(x_0\,|\,y)\big(\nabla_y \log p_{X_{1-t}\,|\,X_0}(y\,|\,x_0) - \nabla \log p_{X_{1-t}}(y)\big)\mathrm{d}x_0 \\ &\overset{\text{(c)}}{=} \int r(x_0) p_{X_0\,|\,X_{1-t}}(x_0\,|\,y)\nabla_y \log p_{X_{1-t}\,|\,X_0}(y\,|\,x_0)\mathrm{d}x_0 - r_{1-t}(y)\nabla \log p_{X_{1-t}}(y) \\ &\overset{\text{(d)}}{=} r_{1-t}(y)\left(\nabla \log p_{X^{r\text{-wt}}_{1-t}}(y) - \nabla \log p_{X_{1-t}}(y)\right) = r_{1-t}(y) g^\star_t(y). \end{aligned} \tag{89}$$

Here, (a) holds since $\nabla_y \log p_{X_0\,|\,X_{1-t}}(x_0\,|\,y) = \frac{\nabla_y p_{X_0\,|\,X_{1-t}}(x_0\,|\,y)}{p_{X_0\,|\,X_{1-t}}(x_0\,|\,y)}$; (b) arises from the identity $\nabla_y \log p_{X_0\,|\,X_{1-t}}(x_0\,|\,y) = \nabla_y \log p_{X_{1-t}\,|\,X_0}(y\,|\,x_0) - \nabla \log p_{X_{1-t}}(y)$ (i.e., the Bayes rule); (c) is valid due to the following relation

$$\int r(x_0) p_{X_0\,|\,X_{1-t}}(x_0\,|\,y)\nabla \log p_{X_{1-t}}(y)\mathrm{d}x_0 = \nabla \log p_{X_{1-t}}(y)\int r(x_0) p_{X_0\,|\,X_{1-t}}(x_0\,|\,y)\mathrm{d}x_0 = r_{1-t}(y)\nabla \log p_{X_{1-t}}(y);$$

and (d) arises from the following property:

$$\begin{aligned} &\int r(x_0) p_{X_0\,|\,X_{1-t}}(x_0\,|\,y)\nabla_y \log p_{X_{1-t}\,|\,X_0}(y\,|\,x_0)\mathrm{d}x_0 \\ &\quad= \int r(x_0)\frac{p_{X_0\,|\,X_{1-t}}(x_0\,|\,y)}{p_{X_{1-t}\,|\,X_0}(y\,|\,x_0)}\nabla_y p_{X_{1-t}\,|\,X_0}(y\,|\,x_0)\mathrm{d}x_0 \\ &\quad= \int r(x_0)\frac{p_{X_0}(x_0)}{p_{X_{1-t}}(y)}\nabla_y p_{X_{1-t}\,|\,X_0}(y\,|\,x_0)\mathrm{d}x_0 \\ &\quad= \frac{1}{p_{X_{1-t}}(y)}\nabla_y \int r(x_0) p_{X_{1-t}\,|\,X_0}(y\,|\,x_0) p_{X_0}(x_0)\mathrm{d}x_0 \end{aligned}$$

$$= \frac{\mathbb{E}[r(X_0)]}{p_{X_{1-t}}(y)} \nabla_y \int p_{X_{1-t}^{r\text{-wt}} \,|\, X_0^{r\text{-wt}}}(y \,|\, x_0) p_{X_0^{r\text{-wt}}}(x_0) \mathrm{d}x_0 = \frac{\mathbb{E}[r(X_0)]}{p_{X_{1-t}}(y)} \nabla p_{X_{1-t}^{r\text{-wt}}}(y)$$
$$= \frac{\mathbb{E}[r(X_0)] p_{X_{1-t}^{r\text{-wt}}}(y)}{p_{X_{1-t}}(y)} \nabla \log p_{X_{1-t}^{r\text{-wt}}}(y) \overset{(50b)}{=} r_{1-t}(y) \nabla \log p_{X_{1-t}^{r\text{-wt}}}(y).$$

Step 3: putting all pieces together. Substituting (89) into (88) yields

$$\mathbb{E}\big[r_{1-t-\delta}(Y_{t+\delta}^w) \,|\, Y_t^w = y_t\big] - \mathbb{E}\big[r_{1-t}(Y_t^w) \,|\, Y_t^w = y_t\big] = \delta \frac{w}{t} r_{1-t}(y_t) \|g_t^\star(y_t)\|_2^2 + O(\delta^2). \tag{90}$$

With this relation in place, we can deduce that

$$\begin{aligned}
&\mathbb{E}\big[r_{1-t-\delta}(Y_{t+\delta}^w) \,|\, Y_0^w = y_0\big] - \mathbb{E}\big[r_{1-t}(Y_t^w) \,|\, Y_0^w = y_0\big] \\
&\quad = \mathbb{E}\Big[\mathbb{E}\big[r_{1-t-\delta}(Y_{t+\delta}^w) \,|\, Y_t^w\big] \,\Big|\, Y_0^w = y_0\Big] - \mathbb{E}\Big[\mathbb{E}\big[r_{1-t}(Y_t^w) \,|\, Y_t^w\big] \,\Big|\, Y_0^w = y_0\Big] \\
&\quad = \delta \frac{w}{t} \mathbb{E}\big[r_{1-t}(Y_t^w) \|g_t^\star(Y_t^w)\|_2^2 \,|\, Y_0^w = y_0\big] + O(\delta^2),
\end{aligned}$$

which in turn allows one to derive

$$\begin{aligned}
\frac{\partial}{\partial t} \mathbb{E}\big[r_{1-t}(Y_t^w) \,|\, Y_0^w = y_0\big] &= \lim_{\delta \to 0} \frac{1}{\delta} \Big(\mathbb{E}\big[r_{1-t-\delta}(Y_{t+\delta}^w) \,|\, Y_0^w\big] - \mathbb{E}\big[r_{1-t}(Y_t^w) \,|\, Y_0^w = y_0\big] \Big) \\
&= \frac{w}{t} \mathbb{E}\big[r_{1-t}(Y_t^w) \|g_t^\star(Y_t^w)\|_2^2 \,|\, Y_0^w = y_0\big].
\end{aligned}$$

Integrating from $t = 0$ to $1 - \delta$, we arrive at

$$\mathbb{E}\big[r_\delta(Y_{1-\delta}^w) \,|\, Y_0^w = y_0\big] - r_1(y_0) = \int_0^{1-\delta} \frac{w}{t} \mathbb{E}\big[r_{1-t}(Y_t^w) \|g_t^\star(Y_t^w)\|_2^2 \,|\, Y_0^w = y_0\big] \mathrm{d}t. \tag{91}$$

Meanwhile, it follows from (85) that

$$\mathbb{E}\big[r_\delta(Y_{1-\delta}) \,|\, Y_0 = y_0\big] - r_1(y_0) = 0.$$

The above two identities taken collectively conclude the proof of Theorem 1.

7 Conclusion

In this paper, we have developed an algorithmic and theoretical framework that unifies the design and analysis of two widely used paradigms in guided diffusion modeling: (classifier-free) diffusion guidance and reward-guided diffusion. When applied to classifier-free guidance, our results have provided the first theoretical characterization — for general target data distributions — of the specific performance metric that CFG improves, namely the expected reciprocal of the classifier probability. When applied to reward-guided diffusion, our framework has led to a simple yet effective algorithm with two notable advantages: (i) its training procedure closely resembles denoising score matching and hence can be effectively accomplished in practice; (ii) it does not require simulating full diffusion trajectories during training. The numerical experiments on both synthetic and real-world data have further validated our theoretical findings.

Acknowledgments

Y. Chen is supported in part by the Alfred P. Sloan Research Fellowship, the ONR grants N00014-22-1-2354 and N00014-25-1-2344, the NSF grants 2221009 and 2218773, the Wharton AI & Analytics Initiative's AI Research Fund, and the Amazon Research Award. G. Li is supported in part by the Chinese University of Hong Kong Direct Grant for Research and the Hong Kong Research Grants Council ECS 2191363.

A Score matching (proof of (29) and (30))

When performing standard score matching w.r.t. $p_{X_0^{r\text{-wt}}}(\cdot)$ (see Hyvärinen (2005) and Chen et al. (2022, Appendix A)), the optimal parameter of the network is calculated by

$$\begin{aligned}
&\arg\min_\theta \mathop{\mathbb{E}}_{t,x_0\sim p_{X_0^{r\text{-wt}}}(\cdot),\epsilon\sim\mathcal{N}(0,I),x_t=\sqrt{1-t}x_0+\sqrt{t}\epsilon}\left[\|\epsilon-\mathsf{NN}_\theta(x_t,t)\|_2^2\right]\\
&=\arg\min_\theta \mathbb{E}_t\left[\int\left\|\frac{x_t-\sqrt{1-t}x_0}{\sqrt{t}}-\mathsf{NN}_\theta(x_t,t)\right\|_2^2 p_{X_{1-t}^{r\text{-wt}}\,|\,X_0^{r\text{-wt}}}(x_t\,|\,x_0)p_{X_0^{r\text{-wt}}}(x_0)\mathrm{d}x_0\mathrm{d}x_t\right]\\
&=\arg\min_\theta \frac{1}{\mathbb{E}[r(X_0)]}\mathbb{E}_t\left[\iint\left\|\frac{x_t-\sqrt{1-t}x_0}{\sqrt{t}}-\mathsf{NN}_\theta(x_t,t)\right\|_2^2 r(x_0)p_{X_{1-t}\,|\,X_0}(x_t\,|\,x_0)p_{X_0}(x_0)\mathrm{d}x_0\mathrm{d}x_t\right]\\
&=\arg\min_\theta \mathop{\mathbb{E}}_{t,x_0\sim p_{X_0}(\cdot),\epsilon\sim\mathcal{N}(0,I),x_t=\sqrt{1-t}x_0+\sqrt{t}\epsilon}\left[r(x_0)\,\|\epsilon-\mathsf{NN}_\theta(x_t,t)\|_2^2\right],
\end{aligned}$$

which coincides exactly with the definition of $\widehat{\theta}^{r\text{-wt}}$ in (29).

When the class of neural networks $\{\mathsf{NN}_\theta\}$ is sufficiently expressive, then one has

$$\begin{aligned}
\mathsf{NN}_{\widehat{\theta}^{r\text{-wt}}}(x_t,t)&=\mathop{\mathbb{E}}_{x_0\sim p_{X_0^{r\text{-wt}}}(\cdot),\epsilon\sim\mathcal{N}(0,I)}\left[\epsilon\,|\,\sqrt{1-t}x_0+\sqrt{t}\epsilon=x_t\right]\\
&=\mathop{\mathbb{E}}_{x_0\sim p_{X_0^{r\text{-wt}}}(\cdot),\epsilon\sim\mathcal{N}(0,I)}\left[\frac{x_t-\sqrt{1-t}x_0}{\sqrt{t}}\,|\,\sqrt{1-t}x_0+\sqrt{t}\epsilon=x_t\right]\\
&=\sqrt{t}\int\frac{x_t-\sqrt{1-t}x_0}{t}p_{X_0^{r\text{-wt}}\,|\,X_t^{r\text{-wt}}}(x_0\,|\,x_t)\mathrm{d}x_0\\
&=-\sqrt{t}\nabla\log p_{X_t^{r\text{-wt}}}(x_t),
\end{aligned}$$

where the last line follows since

$$\begin{aligned}
\nabla\log p_{X_t^{r\text{-wt}}}(x_t)&=\frac{\nabla p_{X_t^{r\text{-wt}}}(x_t)}{p_{X_t^{r\text{-wt}}}(x_t)}=\frac{1}{p_{X_t^{r\text{-wt}}}(x_t)}\int\nabla_{x_t}p_{X_t^{r\text{-wt}}\,|\,X_0^{r\text{-wt}}}(x_t\,|\,x_0)p_{X_0^{r\text{-wt}}}(x_0)\mathrm{d}x_0\\
&=-\frac{1}{p_{X_t^{r\text{-wt}}}(x_t)}\int\left(\frac{x_t-\sqrt{1-t}x_0}{t}\right)p_{X_t^{r\text{-wt}}\,|\,X_0^{r\text{-wt}}}(x_t\,|\,x_0)p_{X_0^{r\text{-wt}}}(x_0)\mathrm{d}x_0\\
&=-\int\left(\frac{x_t-\sqrt{1-t}x_0}{t}\right)p_{X_0^{r\text{-wt}}\,|\,X_t^{r\text{-wt}}}(x_0\,|\,x_t)\mathrm{d}x_0.
\end{aligned}$$

According to the definition of $s_n^{r\text{-wt},\star}(x)$ (cf. (28)), we can take $t=1-\overline{\alpha}_n$ to complete the proof of (30).

B Settings and basic calculation for numerical experiments

B.1 Setup for the numerical experiments in Section 4.1

We intend to calculate score functions for a general one-dimensional GMM with data distribution

$$X_0\sim\sum_{k=1}^K\pi_k\mathcal{N}(\mu_k,1), \tag{92}$$

where π_k is the prior probability, and μ_k is the mean of the k-th component. In the numerical setup, the data distribution p_{data} corresponds to a GMM with $K=3$ components and parameters

$$\pi_1=1/2,\quad \pi_2=\pi_3=1/4,\quad \mu_1=0,\quad \mu_2=-1,\quad \text{and}\quad \mu_3=1.$$

The conditional distribution for class $c=1$ is a GMM with $K=2$ components and parameters

$$\pi_1=\pi_2=1/2,\quad \mu_1=-1\quad\text{and}\quad \mu_1=1.$$

In view of Lemma 1, the random variable X_{1-t} follows the distribution

$$X_{1-t} \sim \sum_{k=1}^{K} \pi_k \mathcal{N}\big(\sqrt{t}\mu_k, 1\big),$$

with the corresponding density function given by

$$p_{X_{1-t}}(x) = \sum_{k=1}^{K} \pi_k (2\pi)^{-1/2} \exp\bigg(- \frac{(x - \sqrt{t}\mu_k)^2}{2} \bigg).$$

The gradient of the log-density $\log p_{X_{1-t}}(x)$ can be directly computed as:

$$\nabla \log p_{X_{1-t}}(x) = \frac{\nabla p_{X_{1-t}}(x)}{p_{X_{1-t}}(x)} = -\sum_{k=1}^{K} \pi_k^t \big(x - \sqrt{t}\mu_k\big) = -x + \sqrt{t} \sum_{k=1}^{K} \pi_k^t \mu_k, \tag{93}$$

where

$$\pi_k^t \coloneqq \frac{\pi_k \exp\big(- \frac{(x - \sqrt{t}\mu_k)^2}{2} \big)}{\sum_{i=1}^{K} \pi_i \exp\big(- \frac{(x - \sqrt{t}\mu_i)^2}{2} \big)}.$$

Under the aforementioned numerical setup, the conditional and unconditional score functions $s_n(\cdot \,|\, c)$ and $s_n(\cdot)$ at $t = 1 - \prod_{k=1}^{n}(1 - \beta_k)$ are given respectively by

$$s_n(x \,|\, 1) = \nabla \log p_{X_{1-t} \,|\, c}(x \,|\, 1) = -x + \frac{\sqrt{t}(1 - \exp(-2\sqrt{t}x))}{1 + \exp(-2\sqrt{t}x)}; \tag{94}$$

$$s_n(x) = \nabla \log p_{X_{1-t}}(x) = -x + \frac{\sqrt{t}(1 - \exp(-2\sqrt{t}x))}{1 + \exp(-2\sqrt{t}x) + 2\exp\big(\frac{t}{2} - \sqrt{t}x\big)}. \tag{95}$$

In addition, the classifier probability $p_{c \,|\, X_{1-t}}(1 \,|\, x)$ is given by

$$p_{c \,|\, X_{1-t}}(1 \,|\, x) = \frac{p_{X_{1-t} \,|\, c}(x \,|\, c) p(c)}{p_{X_{1-t}}(x)} = \frac{1 + \exp(-2\sqrt{t}x)}{1 + \exp(-2\sqrt{t}x) + 2\exp\big(\frac{t}{2} - \sqrt{t}x\big)}. \tag{96}$$

B.2 Basic calculation for the Gaussian mixture model

For the external reward function $r^{\mathsf{ext}}(x) = -(x-2)^2$, the exponentiated reward $r(x)$ is proportional to a Gaussian distribution, i.e.,

$$r(x) = \exp\big(\beta r^{\mathsf{ext}}(x)\big) = \exp\big(- \beta (x-2)^2 \big) \propto \phi\left(x \,\middle|\, 2, \frac{1}{2\beta} \right),$$

where $\phi(\cdot \,|\, \mu, \sigma^2)$ denotes the densify function of a Gaussian distribution with mean μ and variance σ^2. Substituting this into the expression for the distribution $p_{X_0^{r\text{-}\mathsf{wt}}}(\cdot)$ gives

$$\begin{aligned} p_{X_0^{r\text{-}\mathsf{wt}}}(x) &= \frac{p_{X_0}(x)}{Z_{\beta,\sigma}} \phi\big(x \,|\, 2, (2\beta)^{-1}\big) = \frac{\phi(x \,|\, -1, \sigma^2)\phi(x \,|\, 2, (2\beta)^{-1}) + \phi(x \,|\, 1, \sigma^2)\phi(x \,|\, 2, (2\beta)^{-1})}{2 Z_{\beta,\sigma}} \\ &= \frac{1}{2 Z_{\beta,\sigma}} \left[\rho_{-1} \phi(x \,|\, \mu_{-1}, \widehat{\sigma}^2) + \rho_1 \phi(x \,|\, \mu_1, \widehat{\sigma}^2) \right], \end{aligned}$$

where $Z_{\beta,\sigma}$ is a normalization factor dependent on β and σ, and the mean and variance parameters of the resulting Gaussian mixture distribution are given by

$$\widehat{\sigma}^2 = \frac{\sigma^2}{1 + 2\beta\sigma^2}, \qquad \mu_{-1} = \frac{4\beta\sigma^2 - 1}{1 + 2\beta\sigma^2}, \qquad \mu_1 = \frac{4\beta\sigma^2 + 1}{1 + 2\beta\sigma^2}, \qquad \rho_{-1} = \exp\left(- \frac{8\beta}{1 + 2\beta\sigma^2} \right), \qquad \rho_1 = 1.$$

Here, $X_0^{r\text{-}\mathsf{wt}}$ also follows a two-component Gaussian mixture distribution:

$$X_0^{r\text{-}\mathsf{wt}} \sim \frac{\rho_{-1}}{\rho_{-1} + \rho_1} \mathcal{N}(\mu_{-1}, \widehat{\sigma}^2) + \frac{\rho_1}{\rho_{-1} + \rho_1} \mathcal{N}(\mu_1, \widehat{\sigma}^2). \tag{97}$$

Calculation of the score function. The score function is given by

$$s_n^\star(\cdot) = \nabla \log p_{X_{1-t}}(\cdot), \qquad \text{with } t = 1 - \prod_{k=1}^{n}(1-\beta_k).$$

By virtue of Lemma 1, the random variable X_{1-t} follows the distribution

$$X_{1-t} \sim \frac{1}{2}\mathcal{N}(-\sqrt{t}, t\sigma^2 + 1 - t) + \frac{1}{2}\mathcal{N}(\sqrt{t}, t\sigma^2 + 1 - t), \quad X_{1-t} \,|\, X_0 \sim \mathcal{N}(\sqrt{t}X_0, 1-t).$$

Direct calculation gives

$$\begin{aligned}
\nabla \log p_{X_{1-t}}(x) &= \frac{\nabla \int p_{X_{1-t}\,|\,X_0}(x) p_{X_0}(x_0) \mathrm{d}x_0}{p_{X_{1-t}}(x)} = \mathbb{E}\left[\frac{\sqrt{t}X_0 - x}{1-t} \,\middle|\, X_{1-t} = x\right] \\
&= -\frac{x}{1-t} + \frac{\sqrt{t}}{1-t}\mathbb{E}[X_0 \,|\, X_{1-t} = x] \\
&= -\frac{x}{t(\sigma^2-1)+1} + \frac{\sqrt{t}}{t(\sigma^2-1)+1}\big(1 - 2p_t(x)\big).
\end{aligned}$$

Here, the last equation makes use of the posterior expectation derived from

$$\begin{aligned}
\mathbb{E}[X_0 \,|\, X_{1-t} = x] &= \left[-1 + \frac{\sqrt{t}\sigma^2(x+\sqrt{t})}{t\sigma^2 + 1 - t}\right] p_t(x) + \left[1 + \frac{\sqrt{t}\sigma^2(x-\sqrt{t})}{t\sigma^2 + 1 - t}\right]\big(1 - p_t(x)\big) \\
&= \frac{\sqrt{t}\sigma^2 x}{t(\sigma^2-1)+1} + \frac{1-t}{t(\sigma^2-1)+1}\big(1 - 2p_t(x)\big),
\end{aligned}$$

with the posterior probability $p_t(x)$ given by

$$p_t(x) \coloneqq \frac{\frac{1}{2}\phi(x \,|\, -\sqrt{t}, t(\sigma^2-1)+1)}{\frac{1}{2}\phi(x \,|\, -\sqrt{t}, t(\sigma^2-1)+1) + \frac{1}{2}\phi(x \,|\, \sqrt{t}, t(\sigma^2-1)+1)} = \frac{\exp(-\frac{2\sqrt{t}x}{t(\sigma^2-1)+t})}{1 + \exp(-\frac{2\sqrt{t}x}{t(\sigma^2-1)+t})}.$$

Calculation of the reward-reweighted score. The reward-reweighted score function is given by

$$s_n^{r\text{-wt},\star}(\cdot) = \nabla \log p_{X_{1-t}^{r\text{-wt}}}(\cdot), \qquad \text{with} \quad t = 1 - \prod_{k=1}^{n}(1-\beta_k).$$

By Lemma 2 and (97), the random variable $X_{1-t}^{r\text{-wt}}$ follows the distribution

$$X_{1-t}^{r\text{-wt}} \sim \frac{\rho_{-1}}{\rho_{-1}+\rho_1}\mathcal{N}(\sqrt{t}\mu_{-1}, t\widehat{\sigma}^2 + 1 - t) + \frac{\rho_1}{\rho_{-1}+\rho_1}\mathcal{N}(\sqrt{t}\mu_1, t\widehat{\sigma}^2 + 1 - t).$$

Direct calculation gives

$$\begin{aligned}
\nabla \log p_{X_{1-t}^{r\text{-wt}}}(x) &= -\frac{x}{1-t} + \frac{\sqrt{t}}{1-t}\mathbb{E}\big[X_0^{r\text{-wt}} \,|\, X_{1-t}^{r\text{-wt}} = x\big] \\
&= -\frac{x}{t(\widehat{\sigma}^2-1)+1} - \frac{\sqrt{t}}{t(\widehat{\sigma}^2-1)+1}\frac{2p_t^{\mathrm{r}}(x) - 4\beta\sigma^2 - 1}{1 + 2\beta\sigma^2},
\end{aligned}$$

where the last equation uses the following posterior expectation:

$$\begin{aligned}
\mathbb{E}[X_0^{r\text{-wt}} \,|\, X_{1-t}^{r\text{-wt}} = x] &= \left[\mu_{-1} + \frac{\sqrt{t}\widehat{\sigma}^2(x - \sqrt{t}\mu_{-1})}{t(\widehat{\sigma}^2-1)+1}\right] p_t^{r\text{-wt}}(x) + \left[\mu_1 + \frac{\sqrt{t}\widehat{\sigma}^2(x - \sqrt{t}\mu_1)}{t(\widehat{\sigma}^2-1)+1}\right]\big(1 - p_t^{r\text{-wt}}(x)\big) \\
&= \frac{\sqrt{t}\widehat{\sigma}^2 x}{t(\widehat{\sigma}^2-1)+1} - \frac{1-t}{t(\widehat{\sigma}^2-1)+1}\big(p_t^{r\text{-wt}}(x)(\mu_1 - \mu_{-1}) - \mu_1\big)
\end{aligned}$$

$$= \frac{\sqrt{t}\widehat{\sigma}^2 x}{t(\widehat{\sigma}^2-1)+1} - \frac{1-t}{t(\widehat{\sigma}^2-1)+1}\left(\frac{2}{1+2\beta\sigma^2}p_t^{r\text{-wt}}(x) - \frac{1+4\beta\sigma^2}{1+2\beta\sigma^2}\right),$$

with the posterior probability given by

$$\begin{aligned} p_t^{r\text{-wt}}(x) &= \frac{\rho_{-1}\phi(x \,|\, \sqrt{t}\mu_{-1}, t(\widehat{\sigma}^2-1)+1)}{\rho_1\phi(x \,|\, \sqrt{t}\mu_1, t(\widehat{\sigma}^2-1)+1) + \rho_{-1}\phi(x \,|\, \sqrt{t}\mu_{-1}, t(\widehat{\sigma}^2-1)+1)} \\ &= \frac{\exp\left(\frac{\sqrt{t}(\mu_{-1}-\mu_1)x}{t(\widehat{\sigma}^2-1)+1} - \frac{8\beta(1-t)}{(1+2\beta\sigma^2)(t(\widehat{\sigma}^2-1)+1)}\right)}{1+\exp\left(\frac{\sqrt{t}(\mu_{-1}-\mu_1)x}{t(\widehat{\sigma}^2-1)+1} - \frac{8\beta(1-t)}{(1+2\beta\sigma^2)(t(\widehat{\sigma}^2-1)+1)}\right)}. \end{aligned}$$

B.3 Basic calculation for the Swiss roll

Let $\mathcal{S} = \{s_1, \cdots, s_N\} \subset \mathbb{R}^2$ denote the Swiss roll dataset, and assume that the target distribution $p_{X_0} = p_{\mathsf{data}}$ is uniform over $\mathcal{S}$. We define a reward region $\widetilde{\mathcal{S}} \subset \mathcal{S}$ as the subset of samples whose first coordinate lies in the interval $[-5, 6]$:

$$\widetilde{\mathcal{S}} = \{s \in \mathcal{S} : \text{the first coordinate of } s \text{ resides within the intervel } [-5, 6]\}.$$

According to the definition of score functions, the gradients of the log densities can be computed as

$$\begin{aligned} \nabla \log p_{X_{1-t}}(x) &= -\frac{x}{1-t} + \frac{\sqrt{t}}{1-t}\mathbb{E}[X_0 \,|\, X_{1-t} = x] = -\frac{x}{1-t} + \frac{\sqrt{t}}{N(1-t)}\sum_{n=1}^N s_i p_{X_0 \,|\, X_{1-t}}(s_i \,|\, x), \\ \nabla \log p_{X_{1-t}^{r\text{-wt}}}(x) &= -\frac{x}{1-t} + \frac{\sqrt{t}}{1-t}\mathbb{E}[X_0 \,|\, X_{1-t}^{r\text{-wt}} = x] = -\frac{x}{1-t} + \frac{\sqrt{t}}{N(1-t)}\sum_{i=1}^N s_i p_{X_0^{r\text{-wt}} \,|\, X_{1-t}^{r\text{-wt}}}(s_i \,|\, x), \end{aligned}$$

where the posterior probabilities $p_{X_0 \,|\, X_{1-t}}(s \,|\, x)$ and $p_{X_0^{r\text{-wt}} \,|\, X_{1-t}^{r\text{-wt}}}(s \,|\, x)$ are given respectively by

$$\begin{aligned} p_{X_0 \,|\, X_{1-t}}(s \,|\, x) &= \frac{p_{X_{1-t} \,|\, X_0}(x|s)}{N p_{X_{1-t}}(x)} = \frac{\exp\left(-\frac{\|x-\sqrt{t}s\|_2^2}{2(1-t)}\right)}{\sum_{i=1}^N \exp\left(-\frac{\|x-\sqrt{t}s_i\|_2^2}{2(1-t)}\right)}, \\ p_{X_0^{r\text{-wt}} \,|\, X_{1-t}^{r\text{-wt}}}(s \,|\, x) &= \frac{p_{X_{1-t}^{r\text{-wt}} \,|\, X_0^{r\text{-wt}}}(x \,|\, s)\exp\left(10\beta\, \mathbb{1}(s \in \widetilde{\mathcal{S}})\right)}{\left(|\widetilde{\mathcal{S}}|(\mathrm{e}^{10\beta}-1) + |\mathcal{S}|\right) p_{X_{1-t}^{r\text{-wt}}}(x)} \\ &= \frac{\exp\left(-\frac{\|x-\sqrt{t}s\|_2^2}{2(1-t)} + 10\beta\, \mathbb{1}(s \in \widetilde{\mathcal{S}})\right)}{\sum_{s' \in \widetilde{\mathcal{S}}}\exp\left(-\frac{\|x-\sqrt{t}s'\|_2^2}{2(1-t)} + 10\beta\right) + \sum_{s' \in \widetilde{\mathcal{S}}^{\mathrm{c}} \cap \mathcal{S}}\exp\left(-\frac{\|x-\sqrt{t}s'\|_2^2}{2(1-t)}\right)}. \end{aligned}$$

C Proof of some auxiliary results in Section 6

C.1 Proof of Lemma 3

According to the equivalence between X_t and Y_t (see (12) in Lemma 1), it is sufficient to focus on establishing the first relation (51a).

In view of the definition of $r_t(\cdot)$ in (31), we can show that, for $0 \le \tau \le t$,

$$\begin{aligned} \mathbb{E}\big[r_\tau(X_\tau) \,|\, X_t = x\big] &= \int r_\tau(x_\tau) p_{X_\tau \,|\, X_t}(x_\tau \,|\, x)\mathrm{d}x_\tau = \iint r(x_0) p_{X_0 \,|\, X_\tau}(x_0 \,|\, x_\tau) p_{X_\tau \,|\, X_t}(x_\tau \,|\, x)\mathrm{d}x_\tau \mathrm{d}x_0 \\ &= \iint r(x_0)\frac{p_{X_0}(x_0)p_{X_\tau \,|\, X_0}(x_\tau \,|\, x_0)}{p_{X_\tau}(x_\tau)}\frac{p_{X_t \,|\, X_\tau}(x \,|\, x_\tau)p_{X_\tau}(x_\tau)}{p_{X_t}(x)}\mathrm{d}x_\tau \mathrm{d}x_0 \\ &= \iint r(x_0)\frac{p_{X_0}(x_0)p_{X_\tau \,|\, X_0}(x_\tau \,|\, x_0)p_{X_t \,|\, X_\tau}(x \,|\, x_\tau)}{p_{X_t}(x)}\mathrm{d}x_\tau \mathrm{d}x_0 \end{aligned}$$

$$\overset{\text{(a)}}{=} \int r(x_0)\frac{p_{X_0}(x_0)p_{X_t\,|\,X_0}(x\,|\,x_0)}{p_{X_t}(x)}\mathrm{d}x_0 = \int r(x_0)p_{X_0\,|\,X_t}(x_0\,|\,x)\mathrm{d}x_0 = r_t(x),$$

where (a) uses the Markovian property of sequence X_t and hence

$$\int p_{X_\tau\,|\,X_0}(x_\tau\,|\,x_0)p_{X_t\,|\,X_\tau}(x\,|\,x_\tau)\mathrm{d}x_\tau = p_{X_t\,|\,X_0}(x\,|\,x_0).$$

This completes the proof.

C.2 Proof of Claim (53)

Before delving into the proof, we first provide the following lower bound on $p_{X_{1-t}}(y)$. Recalling the definition of R in (52), we have

$$\begin{aligned} p_{X_{1-t}}(y) &\geq \int_{\|x_0\|_2\leq R} p_{X_0}(x_0)\big(2\pi(1-t)\big)^{-d/2}\exp\left(-\frac{\|y-\sqrt{t}x_0\|_2^2}{2(1-t)}\right)\mathrm{d}x_0 \\ &\geq \frac{1}{2}\big(2\pi(1-t)\big)^{-d/2}\exp\left(-\frac{(\|y\|_2+\sqrt{t}R)^2}{2(1-t)}\right). \end{aligned} \tag{98}$$

In addition, we can see that $r_{1-t}(y)$ is finite for any given y, since

$$\begin{aligned} r_{1-t}(y) &= \int r(x_0)p_{X_0\,|\,X_{1-t}}(y)\mathrm{d}x_0 = \int r(x_0)\frac{p_{X_0}(x_0)p_{X_{1-t}\,|\,X_0}(y\,|\,x_0)}{p_{X_{1-t}}(y)}\mathrm{d}x_0 \\ &\overset{\text{(a)}}{\leq} 2\exp\left(\frac{(\|y\|_2+\sqrt{t}R)^2}{2(1-t)}\right)\int r(x_0)p_{X_0}(x_0)\mathrm{d}x_0 = 2\mathbb{E}[r(X_0)]\exp\left(\frac{(\|y\|_2+\sqrt{t}R)^2}{2(1-t)}\right). \end{aligned} \tag{99}$$

Here, (a) makes use of (98) as well as the fact that $p_{X_{1-t}\,|\,X_0}(y\,|\,x_0)\leq(2\pi(1-t))^{-d/2}$.

Now we are ready to present the proof of (53).

Proof of Claim (53a). Here, we shall focus on establishing this claim for $k=1$ and $k=2$; the proof for a larger k is similar and is hence omitted for brevity.

Recalling the definition of R in (52), we have the following lower bound for density $p_{X_{1-t}}(y)$:

$$\begin{aligned} p_{X_{1-t}}(y) &\geq p_{X_{1-t},\mathbb{1}\{\|X_0\|_2<R\}}(y,1) \\ &= \mathbb{P}(\|X_0\|_2<R)p_{X_{1-t}\,|\,\|X_0\|_2<R}(y) \\ &\geq \frac{1}{2}\inf_{x_0:\|x_0\|_2<R}\big(2\pi(1-t)\big)^{-d/2}\exp\left(-\frac{\|y-\sqrt{t}x_0\|_2^2}{2(1-t)}\right) \qquad (100) \\ &\geq \frac{1}{2}\big(2\pi(1-t)\big)^{-d/2}\exp\left(-\frac{(\|y\|_2+\sqrt{t}R)^2}{2(1-t)}\right), \qquad (101) \end{aligned}$$

where $p_{X_{1-t},\mathbb{1}\{\|X_0\|_2<R\}}$ denotes the joint probability density of X_{1-t} and the indicator variable $\mathbb{1}\{\|X_0\|_2<R\}$, and $p_{X_{1-t}\,|\,\|X_0\|_2<R}(y)$ denotes the probability density of X_{1-t} conditioned on the event $\|X_0\|_2<R$. For $t<1$, given that the $X_{1-t}\,|\,X_0$ follows the Gaussian distribution $\mathcal{N}(\sqrt{t}X_0,(1-t)I)$, the score function admits the following expression:

$$\begin{aligned} \nabla\log p_{X_{1-t}}(y) &= -p_{X_{1-t}}(y)^{-1}\int_{x_0}p_{X_0}(x_0)\big(2\pi(1-t)\big)^{-d/2}\exp\Big(-\frac{\|y-\sqrt{t}x_0\|_2^2}{2(1-t)}\Big)\frac{y-\sqrt{t}x_0}{1-t}\mathrm{d}x_0 \\ &= -\int_{x_0}p_{X_0\,|\,X_{1-t}}(x_0\,|\,y)\frac{y-\sqrt{t}x_0}{1-t}\mathrm{d}x_0. \end{aligned} \tag{102}$$

Moreover, observing that for any $D>0$, one has

$$\big\|\nabla\log p_{X_{1-t}}(y)\big\|_2 \leq \int_{x_0:\left\|\frac{y-\sqrt{t}x_0}{\sqrt{1-t}}\right\|_2\leq D} p_{X_0\,|\,X_{1-t}}(x_0\,|\,y)\left\|\frac{y-\sqrt{t}x_0}{1-t}\right\|_2\mathrm{d}x_0$$

$$+\left\|p_{X_{1-t}}(y)^{-1}\int_{x_0:\left\|\frac{y-\sqrt{t}x_0}{\sqrt{1-t}}\right\|_2>D}p_{X_0}(x_0)\big(2\pi(1-t)\big)^{-d/2}\exp\Big(-\frac{\|y-\sqrt{t}x_0\|_2^2}{2(1-t)}\Big)\frac{y-\sqrt{t}x_0}{1-t}\mathrm{d}x_0\right\|_2. \tag{103}$$

The first term above can be bounded by

$$\int_{x_0:\left\|\frac{y-\sqrt{t}x_0}{\sqrt{1-t}}\right\|_2\le D}p_{X_0\,|\,X_{1-t}}(x_0\,|\,y)\left\|\frac{y-\sqrt{t}x_0}{1-t}\right\|_2\mathrm{d}x_0\le\frac{D}{\sqrt{1-t}}. \tag{104}$$

Regarding the second term above, substituting (98) into the second term on the right-hand-side of (103) yields

$$\begin{aligned}
&\left\|p_{X_{1-t}}(y)^{-1}\int_{x_0:\left\|\frac{y-\sqrt{t}x_0}{\sqrt{1-t}}\right\|_2>D}p_{X_0}(x_0)(2\pi(1-t))^{-d/2}\exp\Big(-\frac{\|y-\sqrt{t}x_0\|_2^2}{2(1-t)}\Big)\frac{y-\sqrt{t}x_0}{1-t}\mathrm{d}x_0\right\|_2\\
&\le 2\exp\Big(\frac{(\|y\|_2+\sqrt{t}R)^2}{2(1-t)}\Big)\int_{x_0:\left\|\frac{y-\sqrt{t}x_0}{\sqrt{1-t}}\right\|_2>D}p_{X_0}(x_0)\exp\left(-\frac{\|y-\sqrt{t}x_0\|_2^2}{2(1-t)}\right)\left\|\frac{y-\sqrt{t}x_0}{1-t}\right\|_2\mathrm{d}x_0\\
&\lesssim\frac{2}{\sqrt{1-t}}\exp\left(\frac{(\|y\|_2+\sqrt{t}R)^2}{2(1-t)}-cD^2+cd\right),
\end{aligned}\tag{105}$$

where c is some universal constant. By choosing

$$D=C'\left(\frac{\|y\|_2+\sqrt{t}R}{\sqrt{1-t}}+d\right)$$

for some constant $C'>0$ large enough, we can see from the above bounds that

$$\big\|\nabla\log p_{X_{1-t}}(y)\big\|_2\le\frac{2D}{\sqrt{1-t}}\lesssim\frac{\|y\|_2+\sqrt{t}R}{1-t}+\frac{d}{\sqrt{1-t}}. \tag{106}$$

In addition, by replacing X_0 with $X_0\,|\,c$ and $X_0^{\text{r-wt}}$, we obtain the following two inequalities via similar arguments:

$$\begin{aligned}
\|\nabla p_{X_{1-t}\,|\,c}(y\,|\,c)\|_2&\lesssim\frac{\|y\|_2+\sqrt{t}R}{1-t}+\frac{d}{\sqrt{1-t}},\\
\|\nabla p_{X_{1-t}^{\text{r-wt}}}(y)\|_2&\lesssim\frac{\|y\|_2+\sqrt{t}R}{1-t}+\frac{d}{\sqrt{1-t}}.
\end{aligned}\tag{107a}$$

Note that we can bound the Jacobian matrix in a similar way. Direct calculation gives

$$\begin{aligned}
\nabla^2\log p_{X_{1-t}}(y)=&-\frac{1}{(1-t)p_{X_{1-t}}(y)}\int_{x_0}p_{X_0}(x_0)(2\pi(1-t))^{-d/2}\exp\left(-\frac{\|y-\sqrt{t}x_0\|_2^2}{2(1-t)}\right)\frac{(y-\sqrt{t}x_0)(y-\sqrt{t}x_0)}{1-t}\mathrm{d}x_0\\
&+\frac{1}{1-t}I_d+\nabla\log p_{X_{1-t}}(y)\big(\nabla\log p_{X_{1-t}}(y)\big)^\top.
\end{aligned}$$

Invoking (106), we immediately obtain

$$\begin{aligned}
\|\nabla^2\log p_{X_{1-t}}(y)\|\le&\frac{1}{(1-t)p_{X_{1-t}}(y)}\int_{x_0}p_{X_0}(x_0)\big(2\pi(1-t)\big)^{-d/2}\exp\left(-\frac{\|y-\sqrt{t}x_0\|_2^2}{2(1-t)}\right)\frac{\|y-\sqrt{t}x_0\|_2^2}{1-t}\mathrm{d}x_0\\
&+\frac{1}{1-t}+O\left(\frac{d^2}{1-t}\right)+O\left(\frac{\|y\|_2^2+tR^2}{(1-t)^2}\right).
\end{aligned}\tag{108}$$

It suffices to analyze the first term above. Applying similar arguments as for (104)-(105), we reach

$$\frac{1}{p_{X_{1-t}}(y)}\int_{x_0}p_{X_0}(x_0)\big(2\pi(1-t)\big)^{-d/2}\exp\left(-\frac{\|y-\sqrt{t}x_0\|_2^2}{2(1-t)}\right)\frac{\|y-\sqrt{t}x_0\|_2^2}{1-t}\mathrm{d}x_0$$

$$\leq D^2 + 2\exp\left(\frac{(\|y\|_2 + \sqrt{t}R)^2}{2(1-t)} - cD^2 + cd\right). \tag{109}$$

By taking $D \asymp (\|y\|_2 + \sqrt{t}R)/\sqrt{1-t} + d$, we have

$$\frac{1}{p_{X_{1-t}}(y)} \int_{x_0} p_{X_0}(x_0)(2\pi(1-t))^{-d/2} \exp\left(-\frac{\|y-\sqrt{t}x_0\|_2^2}{2(1-t)}\right) \frac{\|y-\sqrt{t}x_0\|_2^2}{1-t} \mathrm{d}x_0 \lesssim \frac{\|y\|_2^2 + tR^2}{1-t} + d^2. \tag{110}$$

Substitution into (108) yields

$$\|\nabla^2 \log p_{X_{1-t}}(y)\| \lesssim \frac{\|y\|_2^2 + tR^2}{(1-t)^2} + \frac{d^2}{1-t}. \tag{111}$$

In addition, by replacing X_0 with $X_0^{r\text{-wt}}$, we obtain the following inequality via similar arguments:

$$\|\nabla^2 \log p_{X_{1-t}^{r\text{-wt}}}(y)\| \lesssim \frac{\|y\|_2^2 + tR^2}{(1-t)^2} + \frac{d^2}{1-t}. \tag{112}$$

Proof of Claims (53b). As before, we shall focus on proving the claim for the case with $k = 1$; the proof for the case with a larger k follows from similar arguments and is hence omitted for brevity.

In view of the definition of r_{1-t} (cf. (31)), we can derive

$$\begin{aligned}
\|\nabla r_{1-t}(y)\|_2 &= \left\|\nabla \int r(x_0) p_{X_0 \,|\, X_{1-t-\Delta t}}(x_0 \,|\, y') \mathrm{d}x_0\right\|_2 \\
&= \left\|\int r(x_0) p_{X_0 \,|\, X_{1-t}}(x_0 \,|\, y) \nabla_y \log p_{X_0 \,|\, X_{1-t}}(x_0 \,|\, y) \mathrm{d}x_0\right\|_2 \\
&\leq \int r(x_0) p_{X_0 \,|\, X_{1-t}}(x_0 \,|\, y) \left\|\nabla_y \log p_{X_0 \,|\, X_{1-t}}(x_0 \,|\, y)\right\|_2 \mathrm{d}x_0.
\end{aligned} \tag{113}$$

By virtue of (106) and the Bayes rule, we have

$$\begin{aligned}
\left\|\nabla_y \log p_{X_0 \,|\, X_{1-t}}(x_0 \,|\, y)\right\|_2 &= \left\|\nabla_y \log p_{X_{1-t} \,|\, X_0}(y \,|\, x_0) + \nabla_y \log p_{X_{1-t}}(y)\right\|_2 \\
&\leq \left\|\nabla_y \log p_{X_{1-t} \,|\, X_0}(y \,|\, x_0)\right\|_2 + \left\|\nabla_y \log p_{X_{1-t}}(y)\right\|_2 \\
&\lesssim \frac{\|y-\sqrt{t}x_0\|_2 + \|y\|_2 + \sqrt{t}R}{1-t} + \frac{d}{\sqrt{1-t}}.
\end{aligned}$$

Substituting into (113), we arrive at

$$\begin{aligned}
\|\nabla r_{1-t}(y)\|_2 &\lesssim \int r(x_0) p_{X_0 \,|\, X_{1-t}}(x_0 \,|\, y) \left(\frac{\|y-\sqrt{t}x_0\|_2 + \|y\|_2 + \sqrt{t}R}{1-t} + \frac{d}{\sqrt{1-t}}\right) \mathrm{d}x_0 \\
&= r_{1-t}(y) \int \frac{\|y-\sqrt{t}x_0\|_2}{1-t} p_{X_0^{r\text{-wt}} \,|\, X_{1-t}^{r\text{-wt}}}(x_0 \,|\, y) \mathrm{d}x_0 \\
&\quad + r_{1-t}(y) \left(\frac{\|y\|_2 + \sqrt{t}R}{1-t} + \frac{d}{\sqrt{1-t}}\right).
\end{aligned}$$

Here, we have used the following identity:

$$\begin{aligned}
r(x_0) p_{X_0 \,|\, X_{1-t}}(x_0 \,|\, y) &= r(x_0) \frac{p_{X_{1-t} \,|\, X_0}(y \,|\, x_0) p_{X_0}(x_0)}{p_{X_{1-t}}(y)} \\
&= \frac{p_{X_{1-t} \,|\, X_0}(y \,|\, x_0) p_{X_0^{r\text{-wt}}}(x_0)}{p_{X_{1-t}^{r\text{-wt}}}(y)} \frac{p_{X_{1-t}^{r\text{-wt}}}(y)}{p_{X_{1-t}}(y')} \mathbb{E}[r(X_0)] \\
&= p_{X_0^{r\text{-wt}} \,|\, X_{1-t}^{r\text{-wt}}}(x_0 \,|\, y) r_{1-t}(y),
\end{aligned}$$

where the second line uses the definition of $p_{X_0}^{r\text{-wt}}$ (cf. (25)), and the last line comes from (50). Repeating the proof of (106) results in

$$\int \frac{\|y-\sqrt{t}x_0\|_2}{1-t} p_{X_0^{r\text{-wt}}\,|\,X_{1-t}^{r\text{-wt}}}(x_0\,|\,y)\mathrm{d}x_0 \lesssim \frac{\|y\|_2+\sqrt{t}R}{1-t}+\frac{d}{\sqrt{1-t}}. \tag{114}$$

Taking the preceding bounds together leads to

$$\|\nabla r_{1-t}(y)\|_2 \lesssim \mathbb{E}[r(X_0)\,|\,X_{1-t}=y]\left(\frac{\|y\|_2+\sqrt{t}R}{1-t}+\frac{d}{\sqrt{1-t}}\right). \tag{115}$$

Further, it follows from (99) and (50) that $\mathbb{E}[r(X_0)\,|\,X_{1-t}=y]$ is finite, thus completing the proof.

Proof of Claims (53c) Let us again focus on proving the claim when $k=1$; the proof for the case with a larger k is similar and is hence omitted.

Define $\widetilde{X}_{1-t}=X_{1-t}/\sqrt{t}$. According to the definition of r_t (cf. (31)), we have

$$r_{1-t}(y)=\int r(x_0)p_{X_0|X_{1-t}}(x_0\,|\,y)\mathrm{d}x_0=\int r(x_0)p_{X_0|\widetilde{X}_{1-t}}(x_0\,|\,y/\sqrt{t})\mathrm{d}x_0.$$

Denote $\widetilde{y}=y/\sqrt{t}$. It follows from the basic calculus that

$$\frac{\partial}{\partial t}r_{1-t}(y)=\underbrace{-\frac{1}{2t^{3/2}}\int r(x_0)\left\langle\nabla_{\widetilde{y}}p_{X_0|\widetilde{X}_{1-t}}(x_0\,|\,\widetilde{y}),y\right\rangle\mathrm{d}x_0}_{=:\mathcal{I}_1}+\underbrace{\int r(x_0)\frac{\partial}{\partial t}p_{X_0|\widetilde{X}_{1-t}}(x_0\,|\,\widetilde{y})\mathrm{d}x_0}_{=:\mathcal{I}_2}. \tag{116}$$

In the sequel, we shall look at the terms $\mathcal{I}_1$ and $\mathcal{I}_2$ separately.

Regarding $\mathcal{I}_1$, we observe that

$$\begin{aligned}\mathcal{I}_1&=\frac{1}{2t^{3/2}}\int r(x_0)\left\langle\nabla_{\widetilde{y}}y\nabla_y p_{X_0|X_{1-t}}(x_0\,|\,y),y\right\rangle\mathrm{d}x_0\\&=\frac{1}{2t}\int r(x_0)\left\langle\nabla_y p_{X_0|X_{1-t}}(x_0\,|\,y),y\right\rangle\mathrm{d}x_0.\end{aligned}$$

Moreover, notice that

$$\begin{aligned}\int r(x_0)\nabla_y p_{X_0|X_{1-t}}(x_0\,|\,y)\mathrm{d}x_0&=\int r(x_0)p_{X_0|X_{1-t}}(x_0\,|\,y)\nabla\log p_{X_0|X_{1-t}}(x_0\,|\,y)\mathrm{d}x_0\\&=\int r(x_0)p_{X_0|X_{1-t}}(x_0\,|\,y)\left(\nabla\log p_{X_{1-t}|X_0}(y\,|\,x_0)-\nabla\log p_{X_{1-t}}(y)\right)\mathrm{d}x_0\\&=\int r(x_0)p_{X_0|X_{1-t}}(x_0\,|\,y)\left(-\frac{y-\sqrt{t}x_0}{1-t}\nabla\log p_{X_{1-t}|X_0}(y\,|\,x_0)-\nabla\log p_{X_{1-t}}(y)\right)\mathrm{d}x_0,\end{aligned}$$

which in turn allows one to derive

$$\begin{aligned}\|\mathcal{I}_1\|_2&\le\frac{\|y\|_2}{2t}\int r(x_0)p_{X_0|X_{1-t}}(x_0\,|\,y)\frac{\|y-\sqrt{t}x_0\|_2}{1-t}\mathrm{d}x_0+\frac{\|y\|_2}{2t}\int r(x_0)p_{X_0|X_{1-t}}(x_0\,|\,y)\|\nabla\log p_{X_{1-t}}(y)\|_2\mathrm{d}x_0\\&\overset{\text{(a)}}{\lesssim}\frac{\|y\|_2\mathbb{E}[r(X_0)]p_{X_{1-t}^{r\text{-wt}}}(y)}{2tp_{X_{1-t}}(y)}\left(\frac{\|y\|_2+\sqrt{t}R}{1-t}+\frac{d}{\sqrt{1-t}}\right)+\frac{\|y\|_2 r_{1-t}(y)}{2t}\|\nabla\log p_{X_{1-t}}(y)\|_2\\&\overset{\text{(b)}}{\lesssim}\frac{\|y\|_2\mathbb{E}[r(X_0)]p_{X_{1-t}^{r\text{-wt}}}(y)}{2tp_{X_{1-t}}(y)}\left(\frac{\|y\|_2+\sqrt{t}R}{1-t}+\frac{d}{\sqrt{1-t}}\right)+\frac{\|y\|_2 r_{1-t}(y)}{2t}\left(\frac{\|y\|_2+\sqrt{t}R}{1-t}+\frac{d}{\sqrt{1-t}}\right)\\&\overset{\text{(c)}}{\lesssim}\frac{\|y\|_2 r_{1-t}(y)}{2t}\left(\frac{\|y\|_2+\sqrt{t}R}{1-t}+\frac{d}{\sqrt{1-t}}\right).\end{aligned}$$

Here, (b) uses (106), (c) results from (50), and the first term on the right-hand side of (a) arises from (114):

$$\begin{aligned}\int r(x_0)p_{X_0|X_{1-t}}(x_0\,|\,y)\frac{\|y-\sqrt{t}x_0\|_2}{1-t}\mathrm{d}x_0 &= \frac{\mathbb{E}[r(X_0)]p_{X_{1-t}^{r\text{-wt}}}(y)}{p_{X_{1-t}}(y)}\int p_{X_0^{r\text{-wt}}|X_{1-t}^{r\text{-wt}}}(x_0\,|\,y)\frac{\|y-\sqrt{t}x_0\|_2}{1-t}\mathrm{d}x_0\\ &\lesssim \frac{\mathbb{E}[r(X_0)]p_{X_{1-t}^{r\text{-wt}}}(y)}{p_{X_{1-t}}(y)}\left(\frac{\|y\|_2+\sqrt{t}R}{1-t}+\frac{d}{\sqrt{1-t}}\right).\end{aligned}$$

When it comes to $\mathcal{I}_2$, we see that

$$\begin{aligned}\frac{\partial}{\partial t}p_{X_0\,|\,\widetilde{X}_{1-t}}(x_0\,|\,\widetilde{y}) &= p_{X_0\,|\,\widetilde{X}_{1-t}}(x_0\,|\,\widetilde{y})\frac{\partial}{\partial t}\log p_{X_0\,|\,\widetilde{X}_{1-t}}(x_0\,|\,\widetilde{y})\\ &= p_{X_0\,|\,\widetilde{X}_{1-t}}(x_0\,|\,\widetilde{y})\frac{\partial}{\partial t}\left(\log p_{\widetilde{X}_{1-t}\,|\,X_0}(\widetilde{y}\,|\,x_0)-\log p_{\widetilde{X}_{1-t}}(\widetilde{y})\right)\\ &= -\frac{p_{X_0\,|\,\widetilde{X}_{1-t}}(x_0\,|\,\widetilde{y})}{t^2}\left(\frac{t\|\widetilde{y}-x_0\|_2^2}{2(1-t)}-\int\frac{t\|\widetilde{y}-x_0'\|_2^2}{2(1-t)}p_{X_0\,|\,\widetilde{X}_{1-t}}(x_0'\,|\,\widetilde{y})\mathrm{d}x_0'\right)\\ &= -\frac{p_{X_0\,|\,X_{1-t}}(x_0\,|\,y)}{t^2}\left(\frac{\|y-\sqrt{t}x_0\|_2^2}{2(1-t)}-\int\frac{\|y-\sqrt{t}x_0'\|_2^2}{2(1-t)}p_{X_0\,|\,X_{1-t}}(x_0'\,|\,y)\mathrm{d}x_0'\right)\end{aligned}$$

According to (106), we have

$$\left\|\frac{\partial}{\partial t}p_{X_0\,|\,\widetilde{X}_{1-t}}(x_0\,|\,\widetilde{y})\right\|_2 \lesssim \frac{p_{X_0\,|\,X_{1-t}}(x_0\,|\,y)}{t^2}\left(\frac{\|y-\sqrt{t}x_0\|_2^2}{2(1-t)}+\frac{\|y\|_2^2+tR^2}{1-t}+d^2\right). \tag{117}$$

Substitution into the definition of $\mathcal{I}_2$ leads to

$$\begin{aligned}\|\mathcal{I}_2\|_2 &\le \int r(x_0)\left\|\frac{\partial}{\partial t}p_{X_0\,|\,\widetilde{X}_{1-t}}(x_0\,|\,\widetilde{y})\right\|_2\mathrm{d}x_0\\ &\lesssim \frac{1}{t^2}\int r(x_0)p_{X_0\,|\,X_{1-t}}(x_0\,|\,y)\left(\frac{\|y-\sqrt{t}x_0\|_2^2}{2(1-t)}+\frac{\|y-\sqrt{t}x_0\|_2^2}{2(1-t)}+\frac{\|y\|_2^2+tR^2}{1-t}+d^2\right)\mathrm{d}x_0\\ &= \frac{r_{1-t}(y)}{t^2}\int p_{X_0^{r\text{-wt}}\,|\,X_{1-t}^{r\text{-wt}}}(x_0\,|\,y)\frac{\|y-\sqrt{t}x_0\|_2^2}{2(1-t)}\mathrm{d}x_0+\frac{r_{1-t}(y)}{t^2}\left(\frac{\|y\|_2^2+tR^2}{1-t}+d^2\right)\\ &\lesssim \frac{r_{1-t}(y)}{t^2}\left(\frac{\|y\|_2^2+tR^2}{1-t}+d^2\right).\end{aligned}$$

Combining the above bounds on $\|\mathcal{I}_1\|_2$ and $\|\mathcal{I}_2\|_2$ with (116), we arrive at

$$\left\|\frac{\partial}{\partial t}r_{1-t}(y)\right\|_2 \lesssim \frac{r_{1-t}(y)}{t}\left(\frac{\|y\|_2(\|y\|_2+\sqrt{t}R)}{1-t}+\frac{d\|y\|_2}{\sqrt{1-t}}+\frac{\|y\|_2^2+tR^2}{t(1-t)}+\frac{d^2}{t}\right).$$

It has also been shown in (99) that $r_{1-t}(y)$ is finite, which completes the proof of (53c).

Proof of Claim (53d). It is first seen that

$$\begin{aligned}\frac{\partial\nabla\log p_{X_{1-t}}(y)}{\partial t} &= -\frac{\partial}{\partial t}\int\frac{y-\sqrt{t}x_0}{1-t}p_{X_0\,|\,X_{1-t}}(x_0\,|\,y)\mathrm{d}x_0\\ &= -\frac{\partial}{\partial t}\frac{\sqrt{t}}{1-t}\int(\widetilde{y}-x_0)p_{X_0\,|\,\widetilde{X}_{1-t}}(x_0\,|\,\widetilde{y})\mathrm{d}x_0\\ &= \underbrace{\frac{\sqrt{t}}{1-t}\left(I_d+\int(\widetilde{y}-x)p_{X_0\,|\,\widetilde{X}_{1-t}}(x_0\,|\,\widetilde{y})\nabla\log p_{X_0\,|\,\widetilde{X}_{1-t}}(x_0\,|\,\widetilde{y})\mathrm{d}x_0\right)\frac{y}{2t^{3/2}}}_{=:\mathcal{I}_1}\\ &\quad\underbrace{-\frac{1+t}{2\sqrt{t}(1-t)^2}\int(\widetilde{y}-x_0)p_{X_0\,|\,\widetilde{X}_{1-t}}(x_0\,|\,\widetilde{y})\mathrm{d}x_0}_{=:\mathcal{I}_2}\end{aligned}$$

$$\underbrace{-\frac{\sqrt{t}}{1-t}\int(\widetilde{y}-x_0)\frac{\partial}{\partial t}p_{X_0\,|\,\widetilde{X}_{1-t}}(x_0\,|\,\widetilde{y})\mathrm{d}x_0}_{=:\mathcal{I}_3}, \tag{118}$$

leaving us with three terms to control.

Regarding $\mathcal{I}_1$, invoking (106) and (109), we can demonstrate that

$$\begin{aligned}\|\mathcal{I}_1\|_2 &\le \frac{\|y\|_2}{2t(1-t)}\left(1+\int\|y-\sqrt{t}x\|_2 p_{X_0\,|\,X_{1-t}}(x_0\,|\,y)\left(\frac{\|y-\sqrt{t}x_0\|_2}{1-t}+\|\nabla\log p_{X_{1-t}}(y)\|_2\right)\mathrm{d}x_0\right)\\ &\lesssim \frac{\|y\|_2}{2t(1-t)}\left(1+\frac{\|y\|_2^2+tR^2}{(1-t)}+d^2\right).\end{aligned}$$

With regards to $\mathcal{I}_2$, it follows from (106) that

$$\|\mathcal{I}_2\|_2 \lesssim \frac{1+t}{t(1-t)}\left(\frac{\|y\|_2+\sqrt{t}R}{1-t}+\frac{d}{\sqrt{1-t}}\right).$$

In view of (117), we can bound $\mathcal{I}_3$ as

$$\begin{aligned}\|\mathcal{I}_3\|_2 &\lesssim \frac{\sqrt{t}}{1-t}\int\|\widetilde{y}-x_0\|_2\frac{p_{X_0\,|\,X_{1-t}}(x_0\,|\,y)}{t^2}\left(\frac{\|y-\sqrt{t}x_0\|_2^2}{2(1-t)}+\frac{\|y\|_2^2+tR^2}{1-t}+d^2\right)\mathrm{d}x_0\\ &\lesssim \frac{1}{t^2}\left(\frac{\|y\|_2^3+t^{3/2}R^3}{(1-t)^2}+\frac{d^3}{\sqrt{1-t}}\right).\end{aligned} \tag{119}$$

Taking the above bounds on $\mathcal{I}_1$, $\mathcal{I}_2$ and $\mathcal{I}_3$ together with (118) completes the proof.

C.3 Proof of property (57)

The starting point of this proof is the relation given in (55), which motivates us to focus attention on the score difference $\nabla\log p_{X_{1-u}}(Y_u^g)-\nabla\log p_{X_{1-u}}(Y_u)$.

For notational convenience, we introduce the interpolation between Y_u^g and Y_u as follows:

$$\widetilde{Y}_u(\gamma)=\gamma Y_u^g+(1-\gamma)Y_u \tag{120}$$

for any $\gamma\in[0,1]$. The fundamental theorem of calculus allows us to express the score difference of interest as an integral involving the Jacobian matrix:

$$\nabla\log p_{X_{1-u}}(Y_u^g)-\nabla\log p_{X_{1-u}}(Y_u)=\int_0^1\nabla^2\log p_{X_{1-u}}\big(\widetilde{Y}_u(\gamma)\big)\mathrm{d}\gamma(Y_u^g-Y_u),$$

which in turn implies the following ℓ_2-norm bound:

$$\big\|\nabla\log p_{X_{1-u}}(Y_u^g)-\nabla\log p_{X_{1-u}}(Y_u)\big\|_2\le\max_{0\le\gamma\le1}\big\|\nabla^2\log p_{X_{1-u}}\big(\widetilde{Y}_u(\gamma)\big)\big\|\big\|Y_u^g-Y_u\big\|_2. \tag{121}$$

Note that the spectral norm of the Jacobian matrix has been bounded in (111) as

$$\big\|\nabla^2\log p_{X_{1-t}}(Y)\big\|\le O\left(\frac{\|Y\|_2^2+tR^2}{(1-t)^2}+\frac{d^2}{1-t}\right)=:C_t\|Y\|_2^2+C_{t,R,d}, \tag{122}$$

where $C_t>0$ (resp. $C_{t,R,d}>0$) denotes some quantity depending only on t (resp. t, R (cf. (52)) and d), independent of Δt. Substituting this into (121) and using the definition of (120) of $\widetilde{Y}_u(\gamma)$, we arrive at

$$\big\|\nabla\log p_{X_{1-u}}(Y_u^g)-\nabla\log p_{X_{1-u}}(Y_u)\big\|_2\le\big(C_t(\|Y_u\|_2^2+\|Y_u^g\|_2^2)+C_{t,R,d}\big)\big\|Y_u^g-Y_u\big\|_2. \tag{123}$$

Therefore, taking this bound together with (55) gives

$$\begin{aligned}\left\|Y_{t+\Delta t}^g-Y_{t+\Delta t}-g\log\frac{t+\Delta t}{t}\right\|_2 &\le \frac{1}{2}\max_{u:t<u\le t+\Delta t}\big\|Y_u^g-Y_u\big\|_2\log\frac{t+\Delta t}{t}\\ &\quad+\int_t^{t+\Delta t}\big(C_t(\|Y_u\|_2^2+\|Y_u^g\|_2^2)+C_{t,R,d}\big)\frac{\mathrm{d}u}{t}\max_{u:t<u\le t+\Delta t}\big\|Y_u^g-Y_u\big\|_2.\end{aligned} \tag{124}$$

Proof of property (57a). We are now positioned to prove the first claim (57a), concerning what happens on the event $\mathcal{E}$. Taking the ℓ_2 bound in (54) (similar to the arguments for (124)), and maximizing over $t < s \le t + \Delta t$, we derive

$$\begin{aligned}\max_{s:t<s\le t+\Delta t} \|Y_s^g - Y_s\|_2 &\le \frac{1}{2} \max_{s:t<s\le t+\Delta t} \|Y_s^g - Y_s\|_2 \log \frac{t+\Delta t}{t} + \|g\|_2 \log \frac{t+\Delta t}{t} \\ &\quad + \int_t^{t+\Delta t} \big(C_t(\|Y_u\|_2^2 + \|Y_u^g\|_2^2) + C_{t,R,d}\big) \frac{\mathrm{d}u}{t} \max_{s:t<s\le t+\Delta t} \big\|Y_s^g - Y_s\big\|_2 \\ &\le \frac{1}{2} \max_{s:t<s\le t+\Delta t} \|Y_s^g - Y_s\|_2 \frac{\Delta t}{t} + \|g\|_2 \frac{\Delta t}{t} + \big(\widetilde{C} C_t + C_{t,R,d}\big) \max_{s:t<s\le t+\Delta t} \big\|Y_s^g - Y_s\big\|_2 \frac{\sqrt{\Delta t}}{t} \\ &\le \|g\|_2 \frac{\Delta t}{t} + C'_{t,R,d} \max_{s:t<s\le t+\Delta t} \big\|Y_s^g - Y_s\big\|_2 \sqrt{\Delta t}, \end{aligned} \tag{125}$$

where the penultimate inequality follows from the definition (56) of $\mathcal{E}$ and the elementary inequality $\log(1+x) \le x$, and $C'_{t,R,d} > 0$ is a quantity dependent only on t, R, d and the constant $\widetilde{C}$. By choosing Δt small enough so that $C'_{t,R,d}\sqrt{\Delta t} \le 1/2$, we can rearrange terms in (125) to immediately reach

$$\max_{s:t<s\le t+\Delta t} \|Y_s^g - Y_s\|_2 \le 2\|g\|_2 \frac{\Delta t}{t}. \tag{126}$$

Substitution into (124) reveals the following bound when restricted to event $\mathcal{E}$ (cf. (56)):

$$\begin{aligned}\Big\|Y_{t+\Delta t}^g - Y_{t+\Delta t} - g \log \frac{t+\Delta t}{t}\Big\|_2 &\le \frac{1}{2} \max_{u:t<u\le t+\Delta t} \big\|Y_u^g - Y_u\big\|_2 \log \frac{t+\Delta t}{t} + C'_{t,R,d} \max_{u:t<u\le t+\Delta t} \big\|Y_u^g - Y_u\big\|_2 \sqrt{\Delta t} \\ &\le \left(\frac{\sqrt{\Delta t}}{t} + 2C'_{t,R,d}\right) \frac{\|g\|_2}{t} (\Delta t)^{3/2}. \end{aligned} \tag{127}$$

This taken collectively with the basic fact $\log \frac{t+\Delta t}{t} = \frac{\Delta t}{t} + o(\Delta t)$ establishes (57a).

Proof of property (57b). Next, we turn to the proof of (57b). To this end, we first bound $\|Y_s\|$ given $Y_t = y$ by using (10b) and the triangle inequality as follows:

$$\|Y_s\|_2 \le \|y\|_2 + \int_t^s \Big(\frac{1}{2}\|Y_u\|_2 + \|\nabla \log p_{X_{1-u}}(Y_u)\|_2\Big) \frac{\mathrm{d}u}{u} + \Big\| \int_t^s \frac{1}{\sqrt{u}} \mathrm{d}B_u \Big\|_2.$$

Similar to (122), we can establish an upper bound on the size of the score function as

$$\big\|\nabla p_{X_{1-s}}(Y_s)\big\|_2 \lesssim \frac{\|Y_s\|_2 + \sqrt{s}R}{1-s} + \frac{d}{\sqrt{1-s}} =: \widehat{C}_s \|Y_s\|_2 + \widehat{C}_{s,R,d} \tag{128}$$

for some quantity $\widehat{C}_s > 0$ (resp. $\widehat{C}_{s,R,d} > 0$) depending only on s (resp. s, R (cf. (52)) and d) and independent of Δt; see also the corresponding result in (106). Maximizing over $t < s \le t + \Delta t$ gives

$$\begin{aligned}\max_{s:t<s\le t+\Delta t} \|Y_s\|_2 \le \|y\|_2 &+ \Big(\frac{1}{2} \max_{s:t<s\le t+\Delta t} \|Y_s\|_2 + \max_{s:t<s\le t+\Delta t} \widehat{C}_s \max_{s:t<s\le t+\Delta t} \|Y_s\|_2 + \max_{s:t<s\le t+\Delta t} \widehat{C}_{s,R,d}\Big) \frac{\Delta t}{t} \\ &+ \max_{s:t<s\le t+\Delta t} \Big\| \int_t^s \frac{1}{\sqrt{u}} \mathrm{d}B_u \Big\|_2. \end{aligned}$$

Taking Δt to be small enough so that $(1/2 + \max_{s:t<s\le t+\Delta t} \widehat{C}_s)\Delta t/t \le 1/2$, we can rearrange terms in the above display and reach

$$\max_{s:t<s\le t+\Delta t} \|Y_s\|_2 \le 2\|y\|_2 + 2 \max_{s:t<s\le t+\Delta t} \widehat{C}_{s,R,d} \frac{\Delta t}{t} + 2 \max_{s:t<s\le t+\Delta t} \Big\| \int_t^s \frac{1}{\sqrt{u}} \mathrm{d}B_u \Big\|_2.$$

This immediately implies the following bound on the integral of interest:

$$\frac{1}{\Delta t}\int_t^{t+\Delta t}\|Y_s\|_2^2\mathrm{d}s \le 12\|y\|_2^2 + 12\max_{s:t<s\le t+\Delta t}\widehat{C}_{s,R,d}^2\left(\frac{\Delta t}{t}\right)^2 + 12\max_{s:t<s\le t+\Delta t}\left\|\int_t^s\frac{1}{\sqrt{u}}\mathrm{d}B_u\right\|_2^2.$$

As a consequence, by choosing $\widetilde{C}_1$ as

$$\widetilde{C}_1 \coloneqq 12\|y\|_2^2 + \frac{12\max_{s:t<s\le t+\Delta t}\widehat{C}_{s,R,d}^2}{t^2} + \frac{48}{t}\left(2+\frac{1}{\varepsilon}\right)d,$$

which depends on y, t, R, and d, we have

$$\mathbb{P}\left\{\int_t^{t+\Delta t}\|Y_s\|_2^2\mathrm{d}s \ge \widetilde{C}_1\sqrt{\Delta t}\right\} \le \mathbb{P}\left\{\max_{s:t<s\le t+\Delta t}\left\|\int_t^s\frac{1}{\sqrt{u}}\mathrm{d}B_u\right\|_2^2 \ge \frac{4}{t\sqrt{\Delta t}}\left(2+\frac{1}{\varepsilon}\right)d\right\}, \tag{129}$$

which links the integral of interest with the magnitudes of a Brownian motion.

To finish up, note that process $\int_t^s\frac{1}{\sqrt{u}}\mathrm{d}B_u$ shares the same distribution as a standard Brownian motion $W_{t(s)}$ with $t(s) \coloneqq \log(s/t)$. Applying the reflection principle in (129) gives

$$\begin{aligned}
\mathbb{P}\left\{\int_t^{t+\Delta t}\|Y_s\|_2^2\mathrm{d}s \ge \widetilde{C}_1\sqrt{\Delta t}\right\} &\le 2\mathbb{P}\left\{\left\|\int_t^{t+\Delta t}\frac{1}{\sqrt{u}}\mathrm{d}B_u\right\|_2^2 \ge \frac{4}{t\sqrt{\Delta t}}\left(2+\frac{1}{\varepsilon}\right)d\right\}\\
&\le 2\mathbb{P}\left\{\left\|\int_t^{t+\Delta t}\frac{1}{\sqrt{u}}\mathrm{d}B_u\right\|_2^2 \ge \frac{4}{t}\left(2+\frac{1}{\varepsilon}\right)d\Delta t\log\frac{1}{\Delta t}\right\}\\
&\le 4\exp\left(-\left(\left(2+\frac{1}{\varepsilon}\right)\log\frac{1}{\Delta t}\right)\right) \le 4(\Delta t)^{2+1/\varepsilon}.
\end{aligned}$$

Repeating the same argument also yields

$$\mathbb{P}\left\{\int_t^{t+\Delta t}\|Y_s^g\|_2^2\mathrm{d}s \ge \widetilde{C}_1\sqrt{\Delta t}\right\} \le 4(\Delta t)^{2+1/\varepsilon}.$$

Taking $\widetilde{C} = 2\widetilde{C}_1$ readily establishes (57b).

C.4 Proof of inequalities in (63)

In this subsection, we shall focus on proving (63a); the proof of (63b) can be completed in a similar manner and is hence omitted for brevity.

According to (115), we have

$$\|\nabla r_{1-t-\Delta t}(y')\|_2 \lesssim \mathbb{E}[r(X_0)\,|\,X_{1-t-\Delta t}=y']\left(\frac{\|y'\|_2+\sqrt{t+\Delta t}R}{1-t-\Delta t}+\frac{d}{\sqrt{1-t-\Delta t}}\right).$$

With this relation in mind, we can readily obtain

$$\begin{aligned}
&\mathbb{E}\big[\|\nabla r_{1-t-\Delta t}(Y_{t+\Delta t})\|_2^{1+\varepsilon}\,|\,Y_t=y\big]\\
&= \int\|\nabla r_{1-t-\Delta t}(y')\|_2^{1+\varepsilon}\,p_{X_{1-t-\Delta t}\,|\,X_{1-t}}(y'\,|\,y)\mathrm{d}y'\\
&\overset{\text{(a)}}{\lesssim} \int\mathbb{E}[r(X_0)^{1+\varepsilon}\,|\,X_{1-t-\Delta t}=y']\frac{\|y'\|_2^{1+\varepsilon}}{(1-t-\Delta t)^{1+\varepsilon}}p_{X_{1-t-\Delta t}\,|\,X_{1-t}}(y'\,|\,y)\mathrm{d}y'\\
&\quad+\left(\frac{\sqrt{t+\Delta t}R}{1-t-\Delta t}+\frac{d}{\sqrt{1-t-\Delta t}}\right)^{1+\varepsilon}\int\mathbb{E}[r(X_0)^{1+\varepsilon}\,|\,X_{1-t-\Delta t}=y']p_{X_{1-t-\Delta t}\,|\,X_{1-t}}(y'\,|\,y)\mathrm{d}y'\\
&\overset{\text{(b)}}{=} \int\frac{\mathbb{E}[r(X_0)^{1+\varepsilon}\,|\,X_{1-t-\Delta t}=y']}{(1-t-\Delta t)^{1+\varepsilon}}\|y'\|_2^{1+\varepsilon}\,p_{X_{1-t-\Delta t}\,|\,X_{1-t}}(y'\,|\,y)\mathrm{d}y'
\end{aligned}$$

$$+\left(\frac{\sqrt{t+\Delta t}R}{1-t-\Delta t}+\frac{d}{\sqrt{1-t-\Delta t}}\right)^{1+\varepsilon}\mathbb{E}[r(X_0)^{1+\varepsilon}\,|\,X_{1-t}=y], \tag{130}$$

where (a) uses the cauchy-schwarz inequality that $\mathbb{E}[r(X_0)\,|\,X_{1-t-\Delta t}]^{1+\varepsilon}\le\mathbb{E}[r(X_0)^{1+\varepsilon}\,|\,X_{1-t-\Delta t}]$, and (b) applies Lemma 3. Regarding the first term of (130), we define

$$X_{1-t}^{r\text{-wt},\varepsilon}\,|\,X_0^{r\text{-wt},\varepsilon}\sim\mathcal{N}\left(\sqrt{t}X_0^{r\text{-wt},\varepsilon},(1-t)I_d\right),\qquad p_{X_0^{r\text{-wt},\varepsilon}}(x_0):=\frac{r(x_0)^{1+\varepsilon}p_{X_0}(x_0)}{\mathbb{E}_{X_0\sim p_{\mathsf{data}}}[r(X_0)^{1+\varepsilon}]}\ \text{for any}\ x_0\in\mathbb{R}^d.$$

It is seen that

$$\begin{aligned}
&\int\mathbb{E}[r(X_0)^{1+\varepsilon}\,|\,X_{1-t-\Delta t}=y']\|y'\|_2^{1+\varepsilon}p_{X_{1-t-\Delta t}\,|\,X_{1-t}}(y'\,|\,y)\mathrm{d}y'\\
&=\iint\|y'\|_2^{1+\varepsilon}r(x_0)^{1+\varepsilon}p_{X_0\,|\,X_{1-t-\Delta t}}(x_0\,|\,y')p_{X_{1-t-\Delta t}\,|\,X_{1-t}}(y'\,|\,y)\mathrm{d}x_0\mathrm{d}y'\\
&=\frac{1}{p_{X_{1-t}}(y)}\iint\|y'\|_2^{1+\varepsilon}r(x_0)^{1+\varepsilon}p_{X_0}(x_0)p_{X_{1-t-\Delta t}\,|\,X_0}(y'\,|\,x_0)p_{X_{1-t}\,|\,X_{1-t-\Delta t}}(y\,|\,y')\mathrm{d}x_0\mathrm{d}y'\\
&=\frac{\mathbb{E}[r(X_0)^{1+\varepsilon}]}{p_{X_{1-t}}(y)}\iint\|y'\|_2^{1+\varepsilon}p_{X_0^{r\text{-wt},\varepsilon}}(x_0)p_{X_{1-t-\Delta t}\,|\,X_0}(y'\,|\,x_0)p_{X_{1-t}\,|\,X_{1-t-\Delta t}}(y\,|\,y')\mathrm{d}x_0\mathrm{d}y'\\
&=\frac{\mathbb{E}[r(X_0)^{1+\varepsilon}]p_{X_{1-t}^{r\text{-wt},\varepsilon}}(y)}{p_{X_{1-t}}(y)}\iint\|y'\|_2^{1+\varepsilon}p_{X_0^{r\text{-wt},\varepsilon}\,|\,X_{1-t-\Delta t}^{r\text{-wt},\varepsilon}}(x_0\,|\,y')p_{X_{1-t-\Delta t}^{r\text{-wt},\varepsilon}\,|\,X_{1-t}^{r\text{-wt},\varepsilon}}(y'\,|\,y)\mathrm{d}x_0\mathrm{d}y'\\
&=\mathbb{E}[r(X_0)^{1+\varepsilon}\,|\,X_{1-t}=y]\int\|y'\|_2^{1+\varepsilon}p_{X_{1-t-\Delta t}^{r\text{-wt},\varepsilon}\,|\,X_{1-t}^{r\text{-wt},\varepsilon}}(y'\,|\,y)\mathrm{d}y',
\end{aligned}$$

where the last identity relies on (50). By invoking a similar proof as for (77), we can demonstrate that

$$\begin{aligned}
\int\|y'\|_2^{1+\varepsilon}p_{X_{1-t-\Delta t}^{r\text{-wt},\varepsilon}\,|\,X_{1-t}^{r\text{-wt},\varepsilon}}(y'\,|\,y)\mathrm{d}y'&\lesssim\left(\frac{t+\Delta t}{t}\right)^{\frac{1+\varepsilon}{2}}\int\left\|y-\sqrt{\frac{t}{t+\Delta t}}y'\right\|_2^{1+\varepsilon}p_{X_{1-t-\Delta t}^{r\text{-wt},\varepsilon}\,|\,X_{1-t}^{r\text{-wt},\varepsilon}}(y'\,|\,y)\mathrm{d}y'\\
&\quad+\int\left(\frac{t+\Delta t}{t}\right)^{\frac{1+\varepsilon}{2}}\|y\|_2^{1+\varepsilon}p_{X_{1-t-\Delta t}^{r\text{-wt},\varepsilon}\,|\,X_{1-t}^{r\text{-wt},\varepsilon}}(y'\,|\,y)\mathrm{d}y'\\
&\lesssim\left(\frac{\Delta t}{t}\right)^{\frac{1+\varepsilon}{2}}\left(\frac{\|y\|_2+\sqrt{t}\widetilde{R}}{(1-t)^{(1+\varepsilon)/2}}+d\log\frac{1}{\Delta t}\right)^{1+\varepsilon}+\left(\frac{t+\Delta t}{t}\right)^{\frac{1+\varepsilon}{2}}\|y\|_2^{1+\varepsilon}<\infty,\\
\mathbb{E}[r(X_0)^{1+\varepsilon}\,|\,X_{1-t}=y]&\le 2\mathbb{E}[r(X_0)^{1+\varepsilon}]\exp\left(\frac{(\|y\|_2+\sqrt{t}R)^2}{2(1-t)}\right)<\infty,
\end{aligned}$$

where R is defined in (52), and $\widetilde{R}$ is some quantity such that

$$\mathbb{P}(\|X_0^{r\text{-wt},\varepsilon}\|_2<R)>\frac{1}{2}.$$

The above results taken collectively conclude the proof.

C.5 Proof of inequalities in (77)

Let us focus on establishing (77a); the proof of (77b) is similar and is hence omitted for brevity.

First, we make the observation that, for any $D>0$,

$$\begin{aligned}
&\int\left\|y-\sqrt{\frac{t}{t+\Delta t}}y'\right\|_2p_{X_{1-t-\Delta t}^{r\text{-wt}}\,|\,X_{1-t}^{r\text{-wt}}}(y'\,|\,y)\mathrm{d}y'\\
&=\int_{y':\left\|\frac{y-\sqrt{t/(t+\Delta t)}y'}{\sqrt{\Delta t/(t+\Delta t)}}\right\|_2\le D}p_{X_{1-t-\Delta t}^{r\text{-wt}}\,|\,X_{1-t}^{r\text{-wt}}}(y'\,|\,y)\left\|y-\sqrt{\frac{t}{t+\Delta t}}y'\right\|_2\mathrm{d}y'
\end{aligned}$$

$$+ p_{X_{1-t}^{r\text{-wt}}}(y)^{-1} \int_{y' : \left\| \frac{y - \sqrt{t/(t+\Delta t)}\, y'}{\sqrt{\Delta t/(t+\Delta t)}} \right\|_2 > D} p_{X_{1-t-\Delta t}^{r\text{-wt}}}(y') \left(\frac{2\pi \Delta t}{t + \Delta t} \right)^{-d/2}$$
$$\cdot \exp\Big(- \frac{t + \Delta t}{2\Delta t} \big\| y - \sqrt{\frac{t}{t+\Delta t}} y' \big\|_2^2 \Big) \left\| y - \sqrt{\frac{t}{t+\Delta t}} y' \right\|_2 \mathrm{d}y'. \tag{131}$$

For the first term, we have

$$\int_{y' : \left\| \frac{y - \sqrt{t/(t+\Delta t)}\, y'}{\sqrt{\Delta t/(t+\Delta t)}} \right\|_2 \le D} p_{X_{1-t-\Delta t}^{r\text{-wt}} \,|\, X_{1-t}^{r\text{-wt}}}(y' \,|\, y) \left\| y - \sqrt{\frac{t}{t+\Delta t}} y' \right\|_2 \mathrm{d}y' \le D \sqrt{\frac{\Delta t}{t + \Delta t}}. \tag{132}$$

For the second term, recalling the definition of R in (52) and repeating the proof of (98), we obtain

$$p_{X_{1-t}^{r\text{-wt}}}(y) \ge \frac{1}{2} (2\pi(1-t))^{-d/2} \exp\left(- \frac{(\|y\|_2 + \sqrt{t} R)^2}{2(1-t)} \right).$$

Substituting this into the second term on the right-hand side of (131) gives

$$\begin{aligned}
& p_{X_{1-t}^{r\text{-wt}}}(y)^{-1} \int_{y' : \left\| \frac{y - \sqrt{t/(t+\Delta t)}\, y'}{\sqrt{\Delta t/(t+\Delta t)}} \right\|_2 > D} p_{X_{1-t-\Delta t}^{r\text{-wt}}}(y') \left(\frac{2\pi \Delta t}{t + \Delta t} \right)^{-d/2} \cdot \\
& \qquad \exp\Big(- \frac{t + \Delta t}{2\Delta t} \big\| y - \sqrt{\frac{t}{t+\Delta t}} y' \big\|_2^2 \Big) \left\| y - \sqrt{\frac{t}{t+\Delta t}} y' \right\|_2 \mathrm{d}y' \\
& \le 2 \left(\frac{(1-t)(t+\Delta t)}{\Delta t} \right)^{d/2} \exp\Big(\frac{(\|y\|_2 + \sqrt{t} R)^2}{2(1-t)} \Big) \cdot \\
& \qquad \int_{y' : \left\| \frac{y - \sqrt{t/(t+\Delta t)}\, y'}{\sqrt{\Delta t/(t+\Delta t)}} \right\|_2 > D} p_{X_{1-t-\Delta t}^{r\text{-wt}}}(y') \exp\left(- \frac{D^2}{2} + \log\left(D^2\right) \right) \sqrt{\frac{\Delta t}{t + \Delta t}} \mathrm{d}y' \\
& \le 2 \sqrt{\frac{\Delta t}{t + \Delta t}} \exp\left(\frac{(\|y\|_2 + \sqrt{t} R)^2}{2(1-t)} - \frac{D}{2} + \log D + \frac{d}{2} \log \frac{(1-t)(t+\Delta t)}{\Delta t} \right).
\end{aligned} \tag{133}$$

By taking

$$D = C' \left(\frac{\|y\|_2 + \sqrt{t} R}{\sqrt{1-t}} + \sqrt{d} \log^{1/2} \frac{(1-t)(t+\Delta t)}{\Delta t} \right)$$

for some constant $C' > 0$ large enough, we arrive at

$$\begin{aligned}
\int \left\| y - \sqrt{\frac{t}{t+\Delta t}} y' \right\|_2 p_{X_{1-t-\Delta t}^{r\text{-wt}} \,|\, X_{1-t}^{r\text{-wt}}}(y' \,|\, y) \mathrm{d}y' & \lesssim D \sqrt{\frac{\Delta t}{t + \Delta t}} \\
& \lesssim \sqrt{\frac{\Delta t}{t + \Delta t}} \left(\frac{\|y\|_2 + \sqrt{t} R}{\sqrt{1-t}} + \sqrt{d \log \frac{(1-t)(t+\Delta t)}{\Delta t}} \right).
\end{aligned} \tag{134}$$

C.6 Proof of Claim (85)

We start by decomposing the expectation of interest as follows:

$$\begin{aligned}
& \mathbb{E}\big[r_{1-t-\delta}(Y_{t+\delta}) - r_{1-t}(Y_t) \,|\, Y_t = y_t \big] \\
& \quad = \mathbb{E}\big[r_{1-t}(Y_{t+\delta}) - r_{1-t}(Y_t) \,|\, Y_t = y_t \big] + \mathbb{E}\big[r_{1-t-\delta}(Y_{t+\delta}) - r_{1-t}(Y_{t+\delta}) \,|\, Y_t = y_t \big].
\end{aligned} \tag{135}$$

In the following, we shall analyze these two terms separately.

Analysis of the first term in (135). Recall the SDE governing (Y_t) in (10b). Ito's formula tells us that

$$r_{1-t}(Y_{t+\delta}) - r_{1-t}(Y_t) = \int_t^{t+\delta} \Big\{ \frac{1}{2s} \mathsf{Tr}\Big(\nabla^2 r_{1-t}(Y_s)\Big) \mathrm{d}s + \nabla r_{1-t}(Y_s)^\top \Big(\frac{1}{2} Y_s + \nabla \log p_{X_{1-s}}(Y_s) \Big) \frac{\mathrm{d}s}{s} + \frac{1}{\sqrt{s}} \mathrm{d}B_s \Big) \Big\}. \tag{136}$$

Regarding the first term above, we can invoke Ito's formula again to show that

$$\begin{aligned} &\mathsf{Tr}\Big(\nabla^2 r_{1-t}(Y_s)\Big) - \mathsf{Tr}\Big(\nabla^2 r_{1-t}(Y_t)\Big) \\ &= \int_t^s \Big\{ \frac{1}{2r} \mathsf{Tr}\Big(\nabla^2 \mathsf{Tr}\Big(\nabla^2 r_{1-t}(Y_r)\Big)\Big) \mathrm{d}r + \nabla \mathsf{Tr}\Big(\nabla^2 r_{1-t}(Y_r)\Big)^\top \Big(\frac{1}{2} Y_r + \nabla \log p_{X_{1-r}}(Y_r) \Big) \frac{\mathrm{d}r}{r} + \frac{1}{\sqrt{r}} \mathrm{d}B_r \Big) \Big\}. \end{aligned} \tag{137}$$

According to the bound (53b), we have, for $t \le r$,

$$\mathbb{E}\Big[\mathsf{Tr}\Big(\nabla^2 \mathsf{Tr}\Big(\nabla^2 r_{1-t}(Y_r)\Big)\Big) \,|\, Y_t = y_t \Big] \le \mathbb{E}\Big[\exp(C_{r,d,4,R,\mathbb{E}[r(X_0)]} + C_{r,d,4,R,\mathbb{E}[r(X_0)]} \|Y_r\|_2^2) \,|\, Y_t = y_t \Big] < \infty \tag{138}$$

$$\text{and} \qquad \mathbb{E}\Big[\nabla \mathsf{Tr}\Big(\nabla^2 r_{1-t}(Y_r)\Big)^\top \Big(\frac{1}{2} Y_r + \nabla \log p_{X_{1-r}}(Y_r) \Big) \,|\, Y_t = y_t \Big] < \infty. \tag{139}$$

Inserting (138) and (139) into (137) reveals that: for $t \le s \le t+\delta$,

$$\mathsf{Tr}\Big(\nabla^2 r_{1-t}(Y_s)\Big) = \mathsf{Tr}\Big(\nabla^2 r_{1-t}(Y_t)\Big) + O(\delta). \tag{140}$$

Applying similar analysis arguments reveals that for $t \le s \le t+\delta$,

$$\mathbb{E}\Big[\nabla r_{1-t}(Y_s)^\top \Big(\frac{1}{2} Y_s + \nabla \log p_{X_{1-s}}(Y_s) \Big) \,|\, Y_t = y_t \Big] = \nabla r_{1-t}(y_t)^\top \Big(\frac{1}{2} y_t + \nabla \log p_{X_{1-t}}(y_t) \Big) + O(\delta), \tag{141}$$

which makes use of the facts that

$$\begin{aligned} &\mathbb{E}\Big[\mathsf{Tr}\left(\nabla^2 \left(\nabla r_{1-t}(Y_r)^\top \Big(\frac{1}{2} Y_r + \nabla \log p_{X_{1-r}}(Y_r) \Big) \right) \right) \,|\, Y_t = y_t \Big] \\ &\le d\mathbb{E}\Big[\big\| \nabla^3 r_{1-t}(Y_r) \big\|_F^2 \Big\| \frac{1}{2} Y_r + \nabla \log p_{X_{1-r}}(Y_r) \Big\|_2^2 \,|\, Y_t = y_t \Big] + d\mathbb{E}\Big[\|\nabla^3 \log p_{1-r}(Y_r)\|_F^2 \|\nabla r_{1-t}(Y_r)\|_F^2 \,|\, Y_t = y_t \Big] \\ &\quad + \mathbb{E}\Big[\|\nabla^2 r_{1-t}(Y_r)\|_F^2 \|\frac{1}{2} I_d + \nabla^2 \log p_{X_{1-r}}(Y_r)\|_F^2 \,|\, Y_t = y_t \Big] < \infty, \\ &\mathbb{E}\Big[\Big(\frac{1}{2} Y_r + \nabla \log p_{X_{1-r}}(Y_r) \Big)^\top \nabla \left(\nabla r_{1-t}(Y_r)^\top \left(\frac{1}{2} Y_r + \nabla \log p_{X_{1-r}}(Y_r) \right) \right) \,|\, Y_t = y_t \Big] \\ &\le \mathbb{E}\Big[\Big(\frac{1}{2} Y_r + \nabla \log p_{X_{1-r}}(Y_r) \Big)^\top \nabla^2 r_{1-t}(Y_r) \Big(\frac{1}{2} Y_r + \nabla \log p_{X_{1-r}}(Y_r) \Big) \,|\, Y_t = y_t \Big] \\ &\quad + \mathbb{E}\Big[\Big(\frac{1}{2} Y_r + \nabla \log p_{X_{1-r}}(Y_r) \Big)^\top \Big(\frac{1}{2} I_d + \nabla^2 \log p_{X_{1-r}}(Y_r) \Big) \nabla r_{1-t}(Y_r) \,|\, Y_t = y_t \Big] < \infty \end{aligned}$$

and

$$\mathbb{E}\Big[\nabla r_{1-t}(Y_r)^\top \frac{\partial}{\partial r} \nabla p_{X_{1-r}}(Y_r) \,|\, Y_t = y_t \Big] < \infty.$$

Taking (140) and (141) together with (136) establishes that

$$\begin{aligned} &\frac{1}{\delta} \mathbb{E}\left[r_{1-t}(Y_{t+\delta}) - r_{1-t}(Y_t) \,|\, Y_t = y_t \right] \\ &\quad = \frac{1}{2t} \mathsf{Tr}\big(\nabla^2 r_{1-t}(y_t)\big) + \nabla r_{1-t}(y_t)^\top \Big(\frac{1}{2} y_t + \nabla \log p_{X_{1-t}}(y_t) \Big) \frac{1}{t} + O(\delta). \end{aligned}$$

Analysis of the second term in (135). The second term can be expressed as:

$$\mathbb{E}\big[r_{1-t-\delta}(Y_{t+\delta}) - r_{1-t}(Y_{t+\delta}) \,|\, Y_t = y_t\big] = \mathbb{E}\bigg[\int_t^{t+\delta} \frac{\partial}{\partial s} r_{1-s}(Y_{t+\delta}) \mathrm{d}s \,|\, Y_t = y_t\bigg].$$

Similar to the above analysis of the first term in (135), we observe that

$$\frac{\partial}{\partial s} r_{1-s}(y) - \frac{\partial}{\partial t} r_{1-t}(y) = \int_t^s \frac{\partial^2}{\partial l^2} r_{1-l}(y) \mathrm{d}l.$$

According to (53c), one has

$$\mathbb{E}\bigg[\frac{\partial^2}{\partial l^2} r_{1-l}(Y_{t+\delta}) \,|\, Y_t = y_t\bigg] < \infty.$$

Taken collectively these allow us to demonstrate that

$$\frac{1}{\delta}\mathbb{E}\big[r_{1-t-\delta}(Y_{t+\delta}) - r_{1-t}(Y_{t+\delta}) \,|\, Y_t = y_t\big] = \frac{\partial}{\partial t} r_{1-t}(y_t) + O(\delta).$$

Putting all this together. Combining the above two results with (135) completes the proof.

D Time discretization and stability

In practice, the diffusion-based sampler operates in discrete time and often rely on imperfect score estimates. To account for such practical considerations, we provide in this section a stability analysis that incorporates both time discretization and errors in score estimation. In particular, we will show that the discrete-time sampling process (27) is able to approximate its continuous-time counterpart (26) as the number of iterations N becomes sufficiently large, thereby justifying the applicability of Theorem 1 in more practical settings. Note, however, that establishing the sharpest possible convergence guarantees is beyond the scope of this work and not our primary focus.

D.1 Assumptions and theoretical analysis

To avoid notational ambiguity, we shall denote the continuous process (26) by $Y_t^{w,\mathsf{cont}}$, and the discrete process (27) by $Y_t^{w,\mathsf{disc}}$, throughout this section. We adopt the stepsize schedule β_n given in (49), under which the score function $s_n(\cdot)$ corresponds with the gradient $\nabla \log p_{X_{1-\overline{\alpha}_n}}(\cdot)$.

We begin by introducing the assumptions required for the stability analysis. We assume access to score function estimates $s_n(\cdot)$, $s_n(\cdot \,|\, c)$, and $s_n^{r\text{-}\mathsf{wt}}(\cdot)$, whose time-averaged mean squared estimation errors are bounded as follows:

Assumption 1. *Assume that the score estimation errors of $s_n(\cdot)$, $s_n(\cdot \,|\, c)$, and $s_n^{r\text{-}\mathsf{wt}}(\cdot)$ are bounded by*

$$\frac{1}{N}\sum_{n=1}^{N} \mathbb{E}_{y\sim p_{Y_{\overline{\alpha}_n}^{w,\mathsf{cont}}}}\Big[\big\|s_n(y) - \nabla \log p_{X_{1-\overline{\alpha}_n}}(y)\big\|_2^2\Big] \le \varepsilon_{\mathsf{score}}^2,$$

$$\frac{1}{N}\sum_{n=1}^{N} \mathbb{E}_{y\sim p_{Y_{\overline{\alpha}_n}^{w,\mathsf{cont}}}}\Big[\big\|s_n(y \,|\, c) - \nabla \log p_{X_{1-\overline{\alpha}_n} \,|\, c}(y \,|\, c)\big\|_2^2\Big] \le \varepsilon_{\mathsf{score}}^2, \tag{142a}$$

$$\frac{1}{N}\sum_{n=1}^{N} \mathbb{E}_{y\sim p_{Y_{\overline{\alpha}_n}^{w,\mathsf{cont}}}}\Big[\big\|s_n^{r\text{-}\mathsf{wt}}(y) - \nabla \log p_{X_{1-\overline{\alpha}_n}^{r\text{-}\mathsf{wt}}}(y)\big\|_2^2\Big] \le \varepsilon_{\mathsf{score}}^2, \tag{142b}$$

where the sequence $Y_{\overline{\alpha}_n}^{w,\mathsf{cont}}$ is defined in (26).

Additionally, we assume the following bounds concerning second-order moments.

Assumption 2. *There exists some quantity $R > 0$ such that, for any $0 < t < 1$, the following holds:*

$$\mathbb{E}_{Y\sim Y_t^{w,\mathsf{cont}}}\Big[\big\|Y\big\|_2^2 + \big\|\nabla \log p_{X_{1-t}}(Y)\big\|_2^2 + \big\|\nabla \log p_{X_{1-t}\,|\,c}(Y\,|\,c)\big\|_2^2 + \big\|\nabla \log p_{X_{1-t}^{r\text{-}\mathsf{wt}}}(Y)\big\|_2^2\Big] \le R^2,$$
$$\mathbb{E}_{Y\sim Y_t^{w,\mathsf{cont}}}\left[\left\|\frac{\partial \nabla \log p_{X_{1-t}}(Y)}{\partial t}\right\|_2^2 + \left\|\frac{\partial \nabla \log p_{X_{1-t}\,|\,c}(Y|c)}{\partial t}\right\|_2^2 + \left\|\frac{\partial \nabla \log p_{X_{1-t}^{r\text{-}\mathsf{wt}}}(Y)}{\partial t}\right\|_2^2\right] \le R^2, \tag{143}$$

where $Y_t^{w,\mathsf{cont}}$ is defined in (26).

Finally, it is assumed that the score functions are Lipschitz-continuous as follows.

Assumption 3. *For all $0 < t < 1$, the score functions $\nabla \log p_{X_t}(x)$, $\nabla \log p_{X_t\,|\,c}(x)$, and $\nabla \log p_{X_{1-t}^{r\text{-}\mathsf{wt}}}(x)$ are Lipschitz-continuous with Lipschitz constant L, i.e., for all $x_1, x_2 \in \mathbb{R}^d$,*

$$\begin{aligned}
\big\|\nabla \log p_{X_t}(x_1) - \nabla \log p_{X_t}(x_2)\big\|_2 &\le L\|x_1 - x_2\|_2,\\
\big\|\nabla \log p_{X_t\,|\,c}(x_1\,|\,c) - \nabla \log p_{X_t\,|\,c}(x_2\,|\,c)\big\|_2 &\le L\|x_1 - x_2\|_2,\\
\big\|\nabla \log p_{X_{1-t}^{r\text{-}\mathsf{wt}}}(x_1) - \nabla \log p_{X_{1-t}^{r\text{-}\mathsf{wt}}}(x_2)\big\|_2 &\le L\|x_1 - x_2\|_2.
\end{aligned} \tag{144}$$

Armed with the above assumptions, we are now ready to state our performance guarantees for the discrete-time sampler. The proof is provided in Appendix D.3.

Theorem 3. *Suppose that Assumptions 1, 2 and 3 hold. Then the KL divergence between the endpoint distributions of the discrete-time sampler (27) (with stepsizes given in (49)) and the continuous-time process (26) is bounded above by*

$$\begin{aligned}
\big(\mathsf{TV}(Y_{\overline{\alpha}_1}^{w,\mathsf{cont}}, Y_1^{w,\mathsf{disc}})\big)^2 &\le \frac{1}{2}\mathsf{KL}(Y_{\overline{\alpha}_1}^{w,\mathsf{cont}} \parallel Y_1^{w,\mathsf{disc}})\\
&\le C\Big(\frac{(1+w^2)L^2 d \log^3 N}{N} + \frac{(1+w^4)L^2R^2\log^4 N}{N^2} + (1+w^2)\varepsilon_{\mathsf{score}}^2 \log N\Big)
\end{aligned} \tag{145}$$

for some sufficiently large constant $C > 0$. Here $Y_{\overline{\alpha}_1}^{w,\mathsf{cont}}$ and $Y_1^{w,\mathsf{disc}}$ are defined in (26) and (27), respectively, and $\mathsf{KL}(X \parallel Y)$ (resp. $\mathsf{TV}(X,Y)$) denotes the KL divergence (resp. total-variation (TV) distance) between the distributions of X and Y.

In words, this result shows that for sufficiently large N and reasonably accurate score estimates, the distribution of the generated sample $Y_1^{w,\mathsf{disc}}$ can well approximate that of the continuous-time limit $Y_{\overline{\alpha}_1}^{w,\mathsf{cont}}$.

For a positive-valued function r_δ, the reward difference between the discrete-time and continuous-time processes can be bounded as follows:

$$\begin{aligned}
\mathbb{E}[r_\delta(Y_1^{w,\mathsf{disc}})] &= \int r_\delta(y) p_{Y_1^{w,\mathsf{disc}}}(y)\mathrm{d}y \ge \int r_\delta(y) p_{Y_{1-\delta}^{w,\mathsf{cont}}}(y)\mathrm{d}y + \int_{\mathcal{R}} r_\delta(y)\big(p_{Y_1^{w,\mathsf{disc}}}(y) - p_{Y_{1-\delta}^{w,\mathsf{cont}}}(y)\big)\mathrm{d}y\\
&= \int r_\delta(y) p_{Y_{1-\delta}^{w,\mathsf{cont}}}(y)\mathrm{d}y + \int_{\mathcal{R}} r_\delta(y)\widetilde{p}(y)\mathrm{d}y\\
&\ge \mathbb{E}[r_\delta(Y_{1-\delta}^{w,\mathsf{cont}})] - \mathbb{E}\big[r_\delta(Y_{1-\delta}^{w,\mathsf{cont}})\,\mathbb{1}\left(r_\delta(Y_{1-\delta}^{w,\mathsf{cont}}) > \tau^{\mathsf{r}}(Y_{1-\delta}^{w,\mathsf{cont}})\right)\big],
\end{aligned}$$

where $\mathcal{R} \coloneqq \{y : p_{Y_1^{w,\mathsf{disc}}}(y) < p_{Y_{1-\delta}^{w,\mathsf{cont}}}(y)\}$. Here, we choose

$$\widetilde{p}(y)\,\mathbb{1}(y \in \mathcal{R}) = p_{Y_1^{w,\mathsf{disc}}}(y)\,\mathbb{1}(y \in \mathcal{R}) - p_{Y_{1-\delta}^{w,\mathsf{cont}}}(y)\,\mathbb{1}(y \in \mathcal{R}) \tag{146}$$

when restricted to the region $\mathcal{R}$, which satisfies

$$\begin{aligned}
\widetilde{p}(y) &\le p_{Y_1^{w,\mathsf{disc}}}(y) \qquad \text{for } y \in \mathcal{R},\\
\int_{\mathcal{R}} \widetilde{p}(y)\mathrm{d}x &\le \mathsf{TV}(p_{Y_1^{w,\mathsf{disc}}}(y), p_{Y_{1-\delta}^{w,\mathsf{cont}}}(y));
\end{aligned}$$

and $\tau^{\mathsf{r}}(X)$ is the threshold such that the probability $\mathbb{P}(r_\delta(X) > \tau)$ does not exceed the TV distance:

$$\tau^{\mathsf{r}}(X) \coloneqq \arg\max_{\tau'} \big\{ \mathsf{TV}(Y_{1-\delta}^{w,\mathsf{cont}}, Y_1^{w,\mathsf{disc}}) \le \mathbb{P}(r_\delta(X) > \tau') \big\}. \tag{147}$$

By setting $1-\delta \coloneqq \overline{\alpha}_1$, we can further derive a bound on the relative error between the discrete and continuous processes:

$$\begin{aligned}\frac{\mathbb{E}[r_{1-\overline{\alpha}_1}(Y_1^{w,\mathsf{disc}})] - \mathbb{E}[r_{1-\overline{\alpha}_1}(Y_{\overline{\alpha}_1}^{0,\mathsf{cont}})]}{\mathbb{E}[r_{1-\overline{\alpha}_1}(Y_{\overline{\alpha}_1}^{w,\mathsf{cont}})] - \mathbb{E}[r_{1-\overline{\alpha}_1}(Y_{\overline{\alpha}_1}^{0,\mathsf{cont}})]} &= 1 - \frac{\mathbb{E}[r_{1-\overline{\alpha}_1}(Y_{\overline{\alpha}_1}^{w,\mathsf{cont}})] - \mathbb{E}[r_{1-\overline{\alpha}_1}(Y_1^{w,\mathsf{disc}})]}{\mathbb{E}[r_{1-\overline{\alpha}_1}(Y_{\overline{\alpha}_1}^{w,\mathsf{cont}})] - \mathbb{E}[r_{1-\overline{\alpha}_1}(Y_{\overline{\alpha}_1}^{0,\mathsf{cont}})]} \\ &\ge 1 - \frac{\mathbb{E}[r_{1-\overline{\alpha}_1}(Y_{\overline{\alpha}_1}^{w,\mathsf{cont}})\,\mathbb{1}(r_{1-\overline{\alpha}_1}(Y_{\overline{\alpha}_1}^{w,\mathsf{cont}}) > \tau(Y_{\overline{\alpha}_1}^{w,\mathsf{cont}}))]}{\mathbb{E}[r_{1-\overline{\alpha}_1}(Y_{\overline{\alpha}_1}^{w,\mathsf{cont}})] - \mathbb{E}[r_{1-\overline{\alpha}_1}(Y_{\overline{\alpha}_1}^{0,\mathsf{cont}})]}.\end{aligned} \tag{148}$$

Similarly, in the case of cost reduction, we have

$$\begin{aligned}\mathbb{E}[J_\delta(Y_1^{w,\mathsf{disc}})] &\le \int J_\delta(y) p_{Y_{1-\delta}^{w,\mathsf{cont}}}(y)\mathrm{d}y + \int_{\mathcal{R}} J_\delta(y)\big(p_{Y_1^{w,\mathsf{disc}}}(y) - p_{Y_{1-\delta}^{w,\mathsf{cont}}}(y)\big)\mathrm{d}y \\ &\le \mathbb{E}[J_\delta(Y_{1-\delta}^{w,\mathsf{cont}})] + \mathbb{E}\big[J_\delta(Y_{1-\delta}^{w,\mathsf{disc}})\,\mathbb{1}\,\big(J_\delta(Y_{1-\delta}^{w,\mathsf{disc}}) > \tau(Y_{1-\delta}^{w,\mathsf{disc}})\big)\big],\end{aligned}$$

where $\mathcal{R} \coloneqq \{y : p_{Y_1^{w,\mathsf{disc}}}(y) > p_{Y_{1-\delta}^{w,\mathsf{cont}}}(y)\}$, and $\tau^{\mathsf{J}}(X)$ is defined in a similar way as (147):

$$\tau^{\mathsf{J}}(X) \coloneqq \arg\max_{\tau'} \big\{ \mathsf{TV}(Y_{1-\delta}^{w,\mathsf{cont}}, Y_1^{w,\mathsf{disc}}) \le \mathbb{P}(J_\delta(X) > \tau') \big\}.$$

Accordingly, we have

$$\frac{\mathbb{E}[J_{1-\overline{\alpha}_1}(Y_1^{0,\mathsf{cont}})] - \mathbb{E}[J_{1-\overline{\alpha}_1}(Y_{\overline{\alpha}_1}^{w,\mathsf{disc}})]}{\mathbb{E}[J_{1-\overline{\alpha}_1}(Y_{\overline{\alpha}_1}^{0,\mathsf{cont}})] - \mathbb{E}[J_{1-\overline{\alpha}_1}(Y_{\overline{\alpha}_1}^{w,\mathsf{cont}})]} \ge 1 - \frac{\mathbb{E}[J_{1-\overline{\alpha}_1}(Y_{\overline{\alpha}_1}^{w,\mathsf{disc}})\,\mathbb{1}(J_{1-\overline{\alpha}_1}(Y_{\overline{\alpha}_1}^{w,\mathsf{disc}}) > \tau(Y_{\overline{\alpha}_1}^{w,\mathsf{disc}}))]}{\mathbb{E}[J_{1-\overline{\alpha}_1}(Y_{\overline{\alpha}_1}^{0,\mathsf{cont}})] - \mathbb{E}[J_{1-\overline{\alpha}_1}(Y_{\overline{\alpha}_1}^{w,\mathsf{cont}})]}. \tag{149}$$

D.2 Numerical validation

For different values of the TV distance, we evaluate the relative error (148) on the ImageNet dataset and the guidance task (cf. (17)). Specifically, for each value of w, we generate 2×10^4 samples $Y_1^{w,\mathsf{disc}}$ and their unguided counterparts $Y_1^{0,\mathsf{disc}}$ by using a pretrained diffusion model (Rombach et al., 2021). Then we calculate the classifier probability $p_{c\,|\,X_0}(c\,|\,Y_1^{w,\mathsf{disc}})$ and $p_{c\,|\,X_0}(c\,|\,Y_1^{0,\mathsf{disc}})$ by using the Inception v3 classifier (Szegedy et al., 2016). Finally, we compute the relative error

$$\frac{\mathbb{E}[p_{c\,|\,X_0}(c\,|\,Y_1^{w,\mathsf{disc}})^{-1}\mathbb{1}(p_{c\,|\,X_0}(c\,|\,Y_1^{w,\mathsf{disc}})^{-1} > \tau)]}{\mathbb{E}[p_{c\,|\,X_0}(c\,|\,Y_1^{0,\mathsf{disc}})^{-1}] - \mathbb{E}[p_{c\,|\,X_0}(c\,|\,Y_1^{w,\mathsf{disc}})^{-1}]}. \tag{150}$$

In light of (149), we take the cost function $J(\cdot) = p_{c|X_0}(c|\cdot)^{-1}$, use $\mathbb{E}[p_{c\,|\,X_0}(c\,|\,Y_1^{0,\mathsf{disc}})^{-1}] - \mathbb{E}[p_{c\,|\,X_0}(c\,|\,Y_1^{w,\mathsf{disc}})^{-1}]$ to estimate $\mathbb{E}[p_{c\,|\,X_{1-\overline{\alpha}_1}}(c\,|\,Y_{\overline{\alpha}_1}^{0,\mathsf{cont}})^{-1}] - \mathbb{E}[p_{c\,|\,X_{1-\overline{\alpha}_1}}(c\,|\,Y_{\overline{\alpha}_1}^{w,\mathsf{cont}})^{-1}]$. The numerical results are reported in Table 1. As can be seen, the relative error remains small, particularly for practical values of $w \ge 1$, thus corroborating the stability of the discrete-time sampler.

Table 1: Empirical values of (150) under varying choices of w and the TV distance.

TV	$w = 0.2$	0.4	0.6	0.8	1	2	3	4
0.30	0.447	0.196	0.115	0.085	0.029	0.006	0.006	0.002
0.10	0.440	0.194	0.114	0.085	0.029	0.006	0.005	0.002

D.3 Proof of Theorem 3

Repeating similar analysis as in Chen et al. (2022, Section 5), we can show that

$$\mathsf{KL}(Y_{\overline{\alpha}_1}^{w,\mathsf{cont}} \parallel Y_1^{w,\mathsf{disc}}) \le \sum_{n=2}^{N} \mathbb{E}\bigg[\int_{\overline{\alpha}_n}^{\overline{\alpha}_{n-1}} \Big\| s_n(Y_{\overline{\alpha}_n}^{w,\mathsf{cont}}) - \nabla \log p_{X_{1-t}}(Y_t^{w,\mathsf{cont}}) + w[g_n(Y_{\overline{\alpha}_n}^{w,\mathsf{cont}}) - g_t^\star(Y_t^{w,\mathsf{cont}})] \Big\|_2^2 \frac{\mathrm{d}t}{t} \bigg] + \mathsf{KL}(Y_{\overline{\alpha}_N}^{w,\mathsf{cont}}, Y_N^{w,\mathsf{disc}}), \tag{151}$$

where $g_n(\cdot) \coloneqq s_n^{r\text{-}\mathsf{wt}}(\cdot) - s_n(\cdot)$ and $g_n^\star(\cdot) \coloneqq s_n^{r\text{-}\mathsf{wt},\star}(\cdot) - s_n^\star(\cdot)$. Separating the discretization error and score estimation error, we obtain

$$\begin{aligned}
\mathsf{KL}(Y_{\overline{\alpha}_1}^{w,\mathsf{cont}} \parallel Y_1^{w,\mathsf{disc}}) &\le 2\sum_{n=2}^{N} \mathbb{E}\bigg[\int_{\overline{\alpha}_n}^{\overline{\alpha}_{n-1}} \Big\| s_n(Y_{\overline{\alpha}_n}^{w,\mathsf{cont}}) - \nabla \log p_{X_{1-\overline{\alpha}_n}}(Y_{\overline{\alpha}_n}^{w,\mathsf{cont}}) + w[g_n(Y_{\overline{\alpha}_n}^{w,\mathsf{cont}}) - g_{\overline{\alpha}_n}^\star(Y_{\overline{\alpha}_n}^{w,\mathsf{cont}})] \Big\|_2^2 \frac{\mathrm{d}t}{t} \bigg] \\
&\quad + 4\sum_{n=2}^{N} \mathbb{E} \int_{\overline{\alpha}_n}^{\overline{\alpha}_{n-1}} \bigg[\Big\| \nabla \log p_{X_{1-\overline{\alpha}_n}}(Y_{\overline{\alpha}_n}^{w,\mathsf{cont}}) - \nabla \log p_{X_{1-t}}(Y_t^{w,\mathsf{cont}}) \Big\|_2^2 \\
&\qquad + w^2 \Big\| g_{\overline{\alpha}_n}^\star(Y_{\overline{\alpha}_n}^{w,\mathsf{cont}}) - g_t^\star(Y_t^{w,\mathsf{cont}}) \Big\|_2^2 \bigg] \frac{\mathrm{d}t}{t} + \mathsf{KL}(Y_{\overline{\alpha}_N}^{w,\mathsf{cont}}, Y_N^{w,\mathsf{disc}}) \\
&=: \mathcal{E}_1 + \mathcal{E}_2 + \mathcal{E}_3,
\end{aligned} \tag{152}$$

leaving us with three terms to control.

According to Assumption 1, the first term $\mathcal{E}_1$ in (152) can be bounded by the estimation error as

$$\begin{aligned}
\mathcal{E}_1 &= 2\sum_{n=2}^{N} \frac{\overline{\alpha}_{n-1} - \overline{\alpha}_n}{\overline{\alpha}_n} \mathbb{E}\Big[\Big\| s_n(Y_{\overline{\alpha}_n}^{w,\mathsf{cont}}) - \nabla \log p_{X_{1-\overline{\alpha}_n}}(Y_{\overline{\alpha}_n}^{w,\mathsf{cont}}) + w[g_n(Y_{\overline{\alpha}_n}^{w,\mathsf{cont}}) - g_{\overline{\alpha}_n}^\star(Y_{\overline{\alpha}_n}^{w,\mathsf{cont}})] \Big\|_2^2 \Big] \\
&\le \frac{2c\log N}{N} \sum_{n=2}^{N} \mathbb{E}\Big[\Big\| s_n(Y_{\overline{\alpha}_n}^{w,\mathsf{cont}}) - \nabla \log p_{X_{1-\overline{\alpha}_n}}(Y_{\overline{\alpha}_n}^{w,\mathsf{cont}}) + w[g_n(Y_{\overline{\alpha}_n}^{w,\mathsf{cont}}) - g_{\overline{\alpha}_n}^\star(Y_{\overline{\alpha}_n}^{w,\mathsf{cont}})] \Big\|_2^2 \Big] \\
&\lesssim (1+w^2)\varepsilon_{\mathsf{score}}^2 \log N.
\end{aligned}$$

Moreover, observing that the distribution of $Y_{\overline{\alpha}_N}^{w,\mathsf{cont}}$ is initialized as $p_{X_{1-\overline{\alpha}_N}}$, we can bound the last term $\mathcal{E}_3$ in (152) as

$$\mathcal{E}_3 \lesssim \overline{\alpha}_N \mathbb{E}[\|X_0\|_2^2] + d\frac{\overline{\alpha}_N^2}{1-\overline{\alpha}_N} \lesssim \frac{1}{N^2},$$

as long as $\overline{\alpha}_N \lesssim \min\{\mathbb{E}[\|X_0\|_2^2]^{-1}, d^{-1}\}/N^2$.

When it comes to the second term $\mathcal{E}_3$ in (152), note that it can be bounded using the Lipschitz property of $\nabla \log p_{X_{1-t}}$ and $g_t^\star$ as follows:

$$\begin{aligned}
&\mathbb{E}\bigg[\int_{\overline{\alpha}_n}^{\overline{\alpha}_{n-1}} \Big\| \nabla \log p_{X_{1-\overline{\alpha}_n}}(Y_{\overline{\alpha}_n}^{w,\mathsf{cont}}) - \nabla \log p_{X_{1-\overline{\alpha}_n}}(Y_t^{w,\mathsf{cont}}) \Big\|_2^2 \frac{\mathrm{d}t}{t} \bigg] + w^2 \mathbb{E}\bigg[\int_{\overline{\alpha}_n}^{\overline{\alpha}_{n-1}} \Big\| g_{\overline{\alpha}_n}^\star(Y_{\overline{\alpha}_n}^{w,\mathsf{cont}}) - g_{\overline{\alpha}_n}^\star(Y_t^{w,\mathsf{cont}}) \Big\|_2^2 \frac{\mathrm{d}t}{t} \bigg] \\
&\le L^2(1+w^2)\mathbb{E}\bigg[\int_{\overline{\alpha}_n}^{\overline{\alpha}_{n-1}} \Big\| Y_{\overline{\alpha}_n}^{w,\mathsf{cont}} - Y_t^{w,\mathsf{cont}} \Big\|_2^2 \frac{\mathrm{d}t}{t} \bigg] \\
&\le L^2(1+w^2)\mathbb{E}\bigg[\int_{\overline{\alpha}_n}^{\overline{\alpha}_{n-1}} \bigg\| \int_{\overline{\alpha}_n}^{t} \Big\{ \Big(\frac{Y_\tau^{w,\mathsf{cont}}}{2} + \nabla \log p_{X_{1-\tau}}(Y_\tau^{w,\mathsf{cont}}) + w g_\tau^\star(Y_\tau^{w,\mathsf{cont}}) \Big) \frac{\mathrm{d}\tau}{\tau} + \frac{\mathrm{d}B_\tau}{\sqrt{\tau}} \Big\} \bigg\|_2^2 \frac{\mathrm{d}t}{t} \bigg] \\
&\overset{\text{(i)}}{\lesssim} L^2(1+w^2)^2 R^2 \left(\frac{\overline{\alpha}_{n-1} - \overline{\alpha}_n}{\overline{\alpha}_n} \right)^3 + L^2(1+w^2) d \left(\frac{\overline{\alpha}_{n-1} - \overline{\alpha}_n}{\overline{\alpha}_n} \right)^2 \\
&\lesssim L^2(1+w^2) \left(\frac{R^2(1+w^2)\log^3 N}{N^3} + \frac{d\log^2 N}{N^2} \right),
\end{aligned}$$

where (i) invokes Assumption 2. Moreover, we make the observation that

$$\begin{aligned}
&\mathbb{E}\bigg[\int_{\overline{\alpha}_n}^{\overline{\alpha}_{n-1}}\Big[\big\|\nabla\log p_{X_{1-\overline{\alpha}_n}}(Y_t^{w,\mathsf{cont}})-\nabla\log p_{X_{1-t}}(Y_t^{w,\mathsf{cont}})\big\|_2^2+w^2\big\|g^\star_{\overline{\alpha}_n}(Y_t^{w,\mathsf{cont}})-g_t^\star(Y_t^{w,\mathsf{cont}})\big\|_2^2\Big]\frac{\mathrm{d}t}{t}\bigg]\\
&\leq\frac{(\overline{\alpha}_{n-1}-\overline{\alpha}_n)^2}{\overline{\alpha}_n}\mathbb{E}\bigg[\int_{\overline{\alpha}_n}^{\overline{\alpha}_{n-1}}\bigg\{\bigg\|\frac{\partial\nabla\log p_{X_{1-t}}(Y_t^{w,\mathsf{cont}})}{\partial t}\bigg\|_2^2+w^2\bigg\|\frac{\partial g_t^\star(Y_t^{w,\mathsf{cont}})}{\partial t}\bigg\|_2^2\bigg\}\mathrm{d}t\bigg]\\
&\lesssim\frac{(1+w^2)R^2\log^3 N}{N^3},
\end{aligned}\tag{153}$$

where the last inequality arises from Assumption 2. Combine the above two inequalities to yield

$$\mathcal{E}_2\lesssim L^2(1+w^2)\sum_{n=2}^{N}\left(\frac{R^2(1+w^2)\log^3 N}{N^3}+\frac{d\log^2 N}{N^2}\right)\lesssim L^2(1+w^2)\left(\frac{R^2(1+w^2)\log^3 N}{N^2}+\frac{d\log^2 N}{N}\right).$$

Inserting the above bounds on $\mathcal{E}_1$, $\mathcal{E}_2$ and $\mathcal{E}_3$ into (151), we establish the upper bound on the KL divergence. The proof is thus complete by invoking the Pinsker inequality to bound the TV distance.

References

Anderson, B. D. (1982). Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326.

Azangulov, I., Deligiannidis, G., and Rousseau, J. (2024). Convergence of diffusion models under the manifold hypothesis in high-dimensions. *arXiv preprint arXiv:2409.18804*.

Benton, J., De Bortoli, V., Doucet, A., and Deligiannidis, G. (2023). Nearly d-linear convergence bounds for diffusion models via stochastic localization. In *The Twelfth International Conference on Learning Representations*.

Black, K., Janner, M., Du, Y., Kostrikov, I., and Levine, S. (2023). Training diffusion models with reinforcement learning. *arXiv preprint arXiv:2305.13301*.

Bradley, A. and Nakkiran, P. (2024). Classifier-free guidance is a predictor-corrector. *arXiv preprint arXiv:2408.09000*.

Cai, C. and Li, G. (2025). Minimax optimality of the probability flow ode for diffusion models. *arXiv preprint arXiv:2503.09583*.

Chen, H., Lee, H., and Lu, J. (2023). Improved analysis of score-based generative modeling: User-friendly bounds under minimal smoothness assumptions. In *International Conference on Machine Learning*, pages 4735–4763. PMLR.

Chen, M., Mei, S., Fan, J., and Wang, M. (2024a). Opportunities and challenges of diffusion models for generative ai. *National Science Review*, 11(12):nwae348.

Chen, S., Chewi, S., Lee, H., Li, Y., Lu, J., and Salim, A. (2024b). The probability flow ode is provably fast. *Advances in Neural Information Processing Systems*, 36.

Chen, S., Chewi, S., Li, J., Li, Y., Salim, A., and Zhang, A. R. (2022). Sampling is as easy as learning the score: theory for diffusion models with minimal data assumptions. *arXiv preprint arXiv:2209.11215*.

Chidambaram, M., Gatmiry, K., Chen, S., Lee, H., and Lu, J. (2024). What does guidance do? a fine-grained analysis in a simple setting. *Advances in Neural Information Processing Systems*, 37:84968–85005.

Chung, H., Kim, J., Mccann, M. T., Klasky, M. L., and Ye, J. C. (2022). Diffusion posterior sampling for general noisy inverse problems. *arXiv preprint arXiv:2209.14687*.

Clark, K., Vicol, P., Swersky, K., and Fleet, D. J. (2023). Directly fine-tuning diffusion models on differentiable rewards. *arXiv preprint arXiv:2309.17400*.

Croitoru, F.-A., Hondru, V., Ionescu, R. T., and Shah, M. (2023). Diffusion models in vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee.

Dhariwal, P. and Nichol, A. (2021). Diffusion models beat GANs on image synthesis. *Advances in Neural Information Processing Systems*, 34:8780–8794.

Fan, Y., Watkins, O., Du, Y., Liu, H., Ryu, M., Boutilier, C., Abbeel, P., Ghavamzadeh, M., Lee, K., and Lee, K. (2023). DPOK: Reinforcement learning for fine-tuning text-to-image diffusion models. *Advances in Neural Information Processing Systems*, 36:79858–79885.

Fu, H., Yang, Z., Wang, M., and Chen, M. (2024). Unveil conditional diffusion models with classifier-free guidance: A sharp statistical theory. *arXiv preprint arXiv:2403.11968*.

Galashov, A., Pokle, A., Doucet, A., Gretton, A., Delbracio, M., and De Bortoli, V. (2025). Learn to guide your diffusion model. *arXiv preprint arXiv:2510.00815*.

Gao, X., Zha, J., and Zhou, X. Y. (2024). Reward-directed score-based diffusion models via q-learning. *arXiv preprint arXiv:2409.04832*.

Guo, Y., Yuan, H., Yang, Y., Chen, M., and Wang, M. (2024). Gradient guidance for diffusion models: An optimization perspective. *Advances in Neural Information Processing Systems*, 37:90736–90770.

Gupta, S., Cai, L., and Chen, S. (2024). Faster diffusion-based sampling with randomized midpoints: Sequential and parallel. *arXiv preprint arXiv:2406.00924*.

Haussmann, U. G. and Pardoux, E. (1986). Time reversal of diffusions. *The Annals of Probability*, pages 1188–1205.

Ho, J., Jain, A., and Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851.

Ho, J. and Salimans, T. (2021). Classifier-free diffusion guidance. In *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*.

Huang, Z., Wei, Y., and Chen, Y. (2024). Denoising diffusion probabilistic models are optimally adaptive to unknown low dimensionality. *arXiv preprint arXiv:2410.18784*.

Huh, D. and Mohapatra, P. (2025). Maximize your diffusion: A study into reward maximization and alignment for diffusion-based control. *arXiv preprint arXiv:2502.12198*.

Hyvärinen, A. (2005). Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(4).

Jiao, Y., Chen, Y., and Li, G. (2025a). Connections between reinforcement learning with feedback, test-time scaling, and diffusion guidance: An anthology. *arXiv preprint arXiv:2509.04372*.

Jiao, Y., Zhou, Y., and Li, G. (2025b). Optimal convergence analysis of DDPM for general distributions. *arXiv preprint arXiv:2510.27562*.

Jin, C., Shi, Q., and Gu, Y. (2025). Stage-wise dynamics of classifier-free guidance in diffusion models. *arXiv preprint arXiv:2509.22007*.

Karras, T., Aittala, M., Kynkäänniemi, T., Lehtinen, J., Aila, T., and Laine, S. (2024). Guiding a diffusion model with a bad version of itself. *Advances in Neural Information Processing Systems*, 37:52996–53021.

Keramati, H., Kirchen, P., Hannan, M., and Jaiman, R. K. (2025). A reward-directed diffusion framework for generative design optimization. *arXiv preprint arXiv:2508.01509*.

Lai, C.-H., Song, Y., Kim, D., Mitsufuji, Y., and Ermon, S. (2025). The principles of diffusion models. *arXiv preprint arXiv:2510.21890*.

Lee, H., Lu, J., and Tan, Y. (2022). Convergence for score-based generative modeling with polynomial complexity. In *Advances in Neural Information Processing Systems*.

Lee, H., Lu, J., and Tan, Y. (2023). Convergence of score-based generative modeling for general data distributions. In *International Conference on Algorithmic Learning Theory*, pages 946–985. PMLR.

Li, G. and Cai, C. (2024). Provable acceleration for diffusion models under minimal assumptions. *arXiv preprint arXiv:2410.23285*.

Li, G., Cai, C., and Wei, Y. (2025a). Dimension-free convergence of diffusion models for approximate gaussian mixtures. *arXiv preprint arXiv:2504.05300*.

Li, G., Huang, Y., Efimov, T., Wei, Y., Chi, Y., and Chen, Y. (2024a). Accelerating convergence of score-based diffusion models, provably. In *Forty-first International Conference on Machine Learning*.

Li, G. and Jiao, Y. (2024). Improved convergence rate for diffusion probabilistic models. *arXiv preprint arXiv:2410.13738*.

Li, G. and Jiao, Y. (2025). Provable efficiency of guidance in diffusion models for general data distribution. *arXiv preprint arXiv:2505.01382*.

Li, G., Wei, Y., Chi, Y., and Chen, Y. (2024b). A sharp convergence theory for the probability flow odes of diffusion models. *arXiv preprint arXiv:2408.02320*.

Li, G. and Yan, Y. (2024). Adapting to unknown low-dimensional structures in score-based diffusion models. *arXiv preprint arXiv:2405.14861*.

Li, G., Zhou, Y., Wei, Y., and Chen, Y. (2025b). Faster diffusion models via higher-order approximation. *arXiv preprint arXiv:2506.24042*.

Li, X., Wang, R., and Qu, Q. (2025c). Towards understanding the mechanisms of classifier-free guidance. *arXiv preprint arXiv:2505.19210*.

Liang, J., Huang, Z., and Chen, Y. (2025). Low-dimensional adaptation of diffusion models: Convergence in total variation. *arXiv preprint arXiv:2501.12982*.

Liang, Y., Ju, P., Liang, Y., and Shroff, N. (2024). Theory on score-mismatched diffusion models and zero-shot conditional samplers. *arXiv preprint arXiv:2410.13746*.

Nisonoff, H., Xiong, J., Allenspach, S., and Listgarten, J. (2024). Unlocking guidance for discrete state-space diffusion and flow models. *arXiv preprint arXiv:2406.01572*.

Pavasovic, K. L., Verbeek, J., Biroli, G., and Mezard, M. (2025). Understanding classifier-free guidance: High-dimensional theory and non-linear generalizations. *arXiv preprint arXiv:2502.07849*.

Potaptchik, P., Azangulov, I., and Deligiannidis, G. (2024). Linear convergence of diffusion models under the manifold hypothesis. *arXiv preprint arXiv:2410.09046*.

Rafailov, R., Sharma, A., Mitchell, E., Manning, C. D., Ermon, S., and Finn, C. (2023). Direct preference optimization: Your language model is secretly a reward model. *Advances in neural information processing systems*, 36:53728–53741.

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., and Chen, M. (2022). Hierarchical text-conditional image generation with CLIP latents. *arXiv preprint arXiv:2204.06125*.

Rojas, K., He, Y., Lai, C.-H., Takida, Y., Mitsufuji, Y., and Tao, M. (2025). Theory-informed improvements to classifier-free guidance for discrete diffusion models. *arXiv preprint arXiv:2507.08965*.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. (2021). High-resolution image synthesis with latent diffusion models.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695.

Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E. L., Ghasemipour, K., Gontijo Lopes, R., Karagol Ayan, B., Salimans, T., et al. (2022). Photorealistic text-to-image diffusion models with deep language understanding. *Advances in Neural Information Processing Systems*, 35:36479–36494.

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., and Chen, X. (2016). Improved techniques for training gans. *Advances in neural information processing systems*, 29.

Schiff, Y., Sahoo, S. S., Phung, H., Wang, G., Boshar, S., Dalla-torre, H., de Almeida, B. P., Rush, A., Pierrot, T., and Kuleshov, V. (2024). Simple guidance mechanisms for discrete diffusion models. *arXiv preprint arXiv:2412.10193*.

Shang, E., Wei, Y., and Roeder, K. (2025). Predicting the unseen: a diffusion-based debiasing framework for transcriptional response prediction at single-cell resolution. *accepted to Proceedings of the National Academy of Sciences*.

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., and Ganguli, S. (2015). Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pages 2256–2265.

Song, Y., Durkan, C., Murray, I., and Ermon, S. (2021a). Maximum likelihood training of score-based diffusion models. *Advances in Neural Information Processing Systems*, 34:1415–1428.

Song, Y. and Ermon, S. (2019). Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. (2021b). Score-based generative modeling through stochastic differential equations. *International Conference on Learning Representations*.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826.

Tang, W. and Xu, R. (2024). A stochastic analysis approach to conditional diffusion guidance. *Columbia University Preprint*.

Tang, W. and Zhao, H. (2024). Score-based diffusion models via stochastic differential equations–a technical tutorial. *arXiv preprint arXiv:2402.07487*.

Uehara, M., Zhao, Y., Biancalani, T., and Levine, S. (2024a). Understanding reinforcement learning-based fine-tuning of diffusion models: A tutorial and review. *arXiv preprint arXiv:2407.13734*.

Uehara, M., Zhao, Y., Black, K., Hajiramezanali, E., Scalia, G., Diamant, N. L., Tseng, A. M., Biancalani, T., and Levine, S. (2024b). Fine-tuning of continuous-time diffusion models as entropy-regularized control. *arXiv preprint arXiv:2402.15194*.

Wallace, B., Dang, M., Rafailov, R., Zhou, L., Lou, A., Purushwalkam, S., Ermon, S., Xiong, C., Joty, S., and Naik, N. (2024). Diffusion model alignment using direct preference optimization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8228–8238.

Wu, Y., Chen, M., Li, Z., Wang, M., and Wei, Y. (2024a). Theoretical insights for diffusion guidance: A case study for gaussian mixture models. *arXiv preprint arXiv:2403.01639*.

Wu, Y., Chen, Y., and Wei, Y. (2024b). Stochastic Runge-Kutta methods: Provable acceleration of diffusion models. *arXiv preprint arXiv:2410.04760*.

Ye, H., Kevin, R., and Molei, T. (2025). What exactly does guidance do in masked discrete diffusion models. *arXiv preprint arXiv:2506.10971*.

Yu, J., Wang, Y., Zhao, C., Ghanem, B., and Zhang, J. (2023). Freedom: Training-free energy-guided conditional diffusion model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 23174–23184.

Yuan, H., Huang, K., Ni, C., Chen, M., and Wang, M. (2023). Reward-directed conditional diffusion: Provable distribution estimation and reward improvement. *Advances in Neural Information Processing Systems*, 36:60599–60635.

Zhang, M. S., Huan, S., Huang, J., Boffi, N. M., Chen, S., and Chewi, S. (2025). Sublinear iterations can suffice even for DDPMs. *arXiv preprint arXiv:2511.04844*.

Zhao, H., Chen, H., Zhang, J., Yao, D. D., and Tang, W. (2024). Scores as actions: a framework of fine-tuning diffusion models by continuous-time reinforcement learning. *arXiv preprint arXiv:2409.08400*.

Ziegler, D. M., Stiennon, N., Wu, J., Brown, T. B., Radford, A., Amodei, D., Christiano, P., and Irving, G. (2019). Fine-tuning language models from human preferences. *arXiv preprint arXiv:1909.08593*.