This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Efficient Weighted Sampling via Score-based Generative Models

Heasung Kim*, Taekyun Lee*, Hyeji Kim, and Gustavo de Veciana
The University of Texas at Austin
Austin, TX 78712
{heasung.kim, taekyun, hyeji, deveciana}@utexas.edu

Abstract

Weighted sampling—sampling from a probability density function (PDF) proportional to the product of a base PDF and a weight function—is a fundamental technique with wide-ranging applications in variance reduction, biased sampling, data augmentation, and more. Leveraging the increasing availability of pretrained score-based generative models (SGMs), we propose a training-free weighted sampling framework that approximates the backward diffusion process of the target distribution by augmenting the pretrained base score function with an auxiliary guidance term, in a principled and computationally efficient manner. Our approach builds on two key components: a lightweight approximation of the guidance that avoids costly higher-order derivatives of both the score and weight functions, and an uncertainty-aware scheduler that dynamically adjusts the guidance strength based on a temporal analysis of approximation error. Together, these components enable accurate and stable sampling without relying on particle-based resampling or Hessian evaluations commonly required by existing methods. We validate the effectiveness of our method from synthetic to large-scale settings such as Stable Diffusion XL, where our framework achieves $1.2\times$ *to* $4.7\times$ *speedups while consistently matching or outperforming state-of-the-art baselines in task performance. These results position our method as a scalable and inference-efficient solution for task-adaptive, time-sensitive sampling in generative applications.*

1. Introduction

Score-based Generative Models (SGMs) have demonstrated the ability to sample from high-dimensional distributions [1–3], with applications spanning image [4, 5], audio [6, 7], and industrial environment modeling [8, 9]. In many practical settings, weighted sampling of the learned distribution of SGMs—achieved by reweighting the density function using a weight function—is essential for a variety of objectives, including variance reduction in estimation, data augmentation, selective feature analysis, reward-based sampling, and addressing bias and fairness considerations [10–13].

Formally, the weighted sampling problem is defined as follows: consider a Probability Density Function (PDF) p over the domain $\mathbb{R}^d$ of the random vector $\mathbf{X}^p$, and weight function $w : \mathbb{R}^d \mapsto \mathbb{R}_+$ where $\mathbb{R}_+$ denotes the set of positive real values. The concept of weighted sampling entails drawing samples from a modified PDF q, whose PDF is proportional to the product of the weight function w and the original PDF p as $q(\mathbf{x}) = \frac{w(\mathbf{x})p(\mathbf{x})}{\int w(\mathbf{x})p(\mathbf{x})\,\mathrm{d}\mathbf{x}}$.

While supporting a wide range of task-specific functions $w(\mathbf{x})$ is desirable, retraining a large-scale generative model for each target distribution q, or fine-tuning it accordingly, is computationally infeasible in practice. To mitigate this challenge, recent efforts have explored leveraging the pretrained score function of an existing SGM to approximate the generation process under $q(\mathbf{x})$—a technique broadly referred to as *guidance* [14–16]. This approach circumvents the need for costly task-specific retraining, making it attractive for real-world deployment.

However, existing guidance-based methods often exhibit limited fidelity in approximating the score function of the reweighted distribution q, resulting in degraded sample quality or unstable generation dynamics [17, 18]. To overcome these limitations, state-of-the-art approaches have introduced advanced inference-time mechanisms such as *time-travel resampling* which involves repeatedly resampling intermediate latent states, typically aligned with semantically meaningful generative phases [14], or *particle-based resampling schemes* [18, 19], which maintain a diverse set of candidate samples and adaptively reweight and resample them during inference. While these techniques can substantially improve sampling fidelity, they introduce additional computational overhead due to increased memory usage and repeated score evaluations, limiting their applicability in computation- and latency-sensitive applica-

*Equal contribution.
Source code: https://github.com/Heasung-Kim/efficient-weighted-sampling-via-score-based-generative-models

tions.

In this work, we address two core challenges in SGM-based weighted sampling: (i) accurately approximating the score function of the target distribution q, and (ii) achieving efficient, real-time generation with minimal computational overhead. To this end, our method offers both principled theoretical foundations and empirically validated improvements over existing techniques. Our contributions are summarized as follows.

Algorithmic Design and Theoretical Development. We propose a two-stage design for training-free weighted sampling using pretrained SGMs. In the first stage, we derive a lightweight approximation of the guidance term that steers the diffusion trajectory of the base SGM toward that of the target sampling density. This approximation avoids second-order derivatives of both the weight function and the base score function, thereby eliminating the need for costly Hessian computations. This is particularly valuable for large pretrained diffusion models, where differentiation is expensive. In the second stage, we analyze the time-dependent approximation error and introduce a principled scheduler that dynamically adjusts the guidance strength to reduce correction errors. Together, these components yield efficient and effective sampling.

Empirical Evaluation and Performance Superiority. We conduct comprehensive empirical studies across a range of challenging settings, including sampling in multimodal base distributions, weighting with neural network-based objectives, and high-dimensional generation tasks. The results demonstrate that our method consistently achieves superior task performance while reducing computational costs. Notably, we observe up to a ***4.7*** $\times$ ***speedup along with superior performance across diverse metrics and tasks*** compared to existing state-of-the-art methods.

2. Background on SGMs

Score-based generative modeling aims to learn the score function $\nabla_{\mathbf{x}} \log q(\mathbf{x})$ of a target data distribution using a parameterized model, typically a neural network. A key innovation in SGM is to learn the score function of noise-perturbed data rather than the original distribution directly. This enables generative modeling via a backward Stochastic Differential Equation (SDE) and improves training and inference stability [3]. Consider the following continuous-time diffusion process.

$$\mathrm{d}\mathbf{X}_t^q = f(\mathbf{X}_t^q, t)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}\mathbf{W}_t, \tag{1}$$

where $\mathbf{X}_0^q \sim q$ is the data variable at diffusion time $t = 0$, and $q_t(\mathbf{x})$ denotes the marginal density of $\mathbf{X}_t^q$ induced by the forward SDE (i.e., $\mathbf{X}_t^q$ is obtained by evolving $\mathbf{X}_0^q$ for time t under equation 1).

When the score function $\nabla_{\mathbf{x}} \log q_t(\mathbf{x})$ is known, this framework enables sampling of $\mathbf{X}_0^q$ that follows PDF $q_0 = q$ via a corresponding backward SDE [20] which is given by

$$\mathrm{d}\mathbf{X}_t^q = b_1(\mathbf{X}_t^q, t)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}\tilde{\mathbf{W}}_t, \quad \text{where} \tag{2}$$

$$b_1(\mathbf{x}, t) = f(\mathbf{x}, t) - \sigma(t)^2 \nabla_{\mathbf{x}} \log q_t(\mathbf{x}) \tag{3}$$

with $\mathbf{X}_t^q \sim q_t$ and $\mathrm{d}\tilde{\mathbf{W}}_t$ denoting the Brownian motion associated with the reverse-time process.

For instance, given an initial sample $\mathbf{X}_T^q \sim q_T(\mathbf{x})$, and running the backward SDE (2) yields a sample $\mathbf{X}_0^q$ from the distribution q_0. Typically, for sufficiently large values of T, a realization $\mathbf{x}_T$ can be initialized as Gaussian noise. Following convention [1, 21–23], we consider the following drift and diffusion coefficients: $f(\mathbf{x}, t) = -\frac{1}{2}\beta(t)\mathbf{x}$ and $\sigma(t) = \sqrt{\beta(t)}$ where $\beta(t)$ is a predefined scalar-valued continuous function such that $0 \leq \beta(t) \leq 1$. Here, in defining $\sigma(t)$ and $f(\mathbf{x}, t)$, we apply a slight abuse of notation: the scalar-vector multiplication in this context represents the elementwise multiplication of the vector by the scalar.

In modern practice, many pretrained SGMs are publicly available and provide accurate realizations of the score function of the base density, $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$. Here $p_t(\mathbf{x})$ is the marginal distribution of $\mathbf{X}_t^p$ which is modeled by the same SDE dynamics as in (1) but with a different initial condition as

$$\mathrm{d}\mathbf{X}_t^p = -\frac{1}{2}\beta(t)\mathbf{X}_t^p\,\mathrm{d}t + \sqrt{\beta(t)}\,\mathrm{d}\mathbf{W}_t, \quad \mathbf{X}_0^p \sim p_0. \tag{4}$$

Throughout this work, we assume access to such a pretrained base SGM, $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$, and focus on efficiently approximating the score function of the weighted sampling density q_t, without any additional training or fine-tuning of the given model.

3. Efficient Weighted Sampling via SGM

The objective of this work is to construct an efficient and practical approximation of the score function $\nabla_{\mathbf{x}} \log q_t(\mathbf{x})$ where $q_0(\mathbf{x}) \propto w(\mathbf{x})p_0(\mathbf{x})$, by leveraging only known quantities based on the following decomposition:

$$\nabla_{\mathbf{x}} \log q_t(\mathbf{x}) = \nabla_{\mathbf{x}} \log p_t(\mathbf{x}) + g(\mathbf{x}, t) \tag{5}$$

$$\approx \nabla_{\mathbf{x}} \log p_t(\mathbf{x}) + \tilde{g}(\mathbf{x}, t) \tag{6}$$

where $g(\mathbf{x}, t)$ denotes the ground-truth *guidance term*, i.e., the direction required to adjust the base score $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ to match the score of the weighted sampling density q_t. Our goal is to approximate $g(\mathbf{x}, t)$ with a computationally lightweight, tractable surrogate $\tilde{g}(\mathbf{x}, t)$, thereby obtaining an efficient estimate of the target score function.

Existing methods [14, 16, 19] can be adapted to approximate the guidance term. However, they often rely on computationally intensive procedures such as differentiation of pretrained score models and/or repeated resampling steps. These approaches, while effective, introduce significant runtime overhead—especially when deployed with large-scale SGMs.

Design Overview. To address the limitations of prior approaches, our method consists of two key components: (i) *Lightweight Approximation.* We derive a first-order approximation of the guidance term g, avoiding second-order derivatives of both the base density p_t and the weight function w. This makes our method well-suited for large-scale pretrained SGMs, where Hessian-based methods are computationally expensive. (ii) *Uncertainty-Adaptive Guidance Scheduling.* We analyze the temporal evolution of the approximation uncertainty, quantifying the confidence in the approximation at each diffusion timestep. Based on this analysis, we propose a dynamic scheduler that modulates the strength of the approximated guidance term during sampling. By downweighting uncertain guidance and emphasizing reliable estimates, the scheduler improves the overall quality of the sampled distribution and enhances robustness.

Together, these components form the basis of our proposed method, Lightweight Approximation with uncertainty-adaptive Guidance Scheduling (LAGS), enabling fast and accurate sampling without the need for retraining or fine-tuning.

3.1. The First-order Approximation of Weighted Sampling Score Function

Connection to Noise-Perturbed Base and Weighted Sampling Densities. We begin by examining the form of the weighted sampling distribution for the noise-perturbed samples, $q_t(\mathbf{x})$, under the SDE defined in (1). Consider the transition probability density function $G : \mathbb{R}^d \times \mathbb{R}^d \times [0, T] \mapsto \mathbb{R}_+$, also referred to as the Green's function or fundamental solution of the SDE [24–26]. This function represents the probability density of the process transitioning from the given initial state $\mathbf{X}_0^q = \mathbf{y}$ to the state $\mathbf{x}$ at time t. Specifically, $G(\mathbf{x}, \mathbf{y}, t)$ denotes the conditional probability density of transitioning from $\mathbf{y}$ to $\mathbf{x}$ over time t. Using this Green's function, we can express $q_t(\mathbf{x})$ as

$$q_t(\mathbf{x}) = \int G(\mathbf{x}, \mathbf{y}, t) q_0(\mathbf{y}) \mathrm{d}\mathbf{y} = \frac{1}{Z_0} \int G(\mathbf{x}, \mathbf{y}, t) w(\mathbf{y}) p_0(\mathbf{y}) \mathrm{d}\mathbf{y} \tag{7}$$

where $q_0 = q$, $p_0 = p$, and Z_0 is the normalization constant, $Z_0 = \int w(\mathbf{x}) p(\mathbf{x}) \, \mathrm{d}\mathbf{x}$.

To further investigate the relationship between q_t and p_t, it should be noted that the corresponding Green's function for SDE in (4) is also G in (7), thereby the distribution of $\mathbf{X}^p$ at time t, denoted by p_t, is given as $p_t(\mathbf{x}) = \int G(\mathbf{x}, \mathbf{y}, t) p_0(\mathbf{y}) \, \mathrm{d}\mathbf{y}$. We then have

$$\begin{aligned} q_t(\mathbf{x}) &= \frac{p_t(\mathbf{x})}{Z_0} \int w(\mathbf{y}) G(\mathbf{x}, \mathbf{y}, t) \frac{p_0(\mathbf{y})}{p_t(\mathbf{x})} \, \mathrm{d}\mathbf{y} \\ &= \frac{p_t(\mathbf{x})}{Z_0} \mathbb{E}_{\mathbf{X}_0^p \sim p_{\mathbf{X}_0^p | \mathbf{X}_t^p}(\cdot|\mathbf{x})} [w(\mathbf{X}_0^p)] \end{aligned} \tag{8}$$

where $p_{\mathbf{X}_0^p | \mathbf{X}_t^p}(\cdot|\mathbf{x})$ is the conditional PDF of the initial state $\mathbf{X}_0^p$ for a given state $\mathbf{X}_t^p = \mathbf{x}$ at t under the corresponding SDE of $\mathbf{X}_t^p$. Based on this, we can observe that q_t can be represented in terms of p_t and w as shown in (8), which implies that the score function of the weighted sampling density can be represented as

$$\begin{aligned} \nabla_{\mathbf{x}} \log q_t(\mathbf{x}) &= \nabla_{\mathbf{x}} \log p_t(\mathbf{x}) + g(\mathbf{x}, t) \quad \text{where} \\ g(\mathbf{x}, t) &\coloneqq \nabla_{\mathbf{x}} \log \mathbb{E}_{\mathbf{X}_0^p \sim p_{\mathbf{X}_0^p | \mathbf{X}_t^p}(\cdot|\mathbf{x})} [w(\mathbf{X}_0^p)]. \end{aligned} \tag{9}$$

Taylor Approximation. While Eq. (9) provides an exact decomposition, the term $g(\mathbf{x}, t)$ is generally intractable. To address this, we apply a first-order Taylor expansion of $w(\mathbf{X}_0^p)$ about its conditional mean, $\bar{\mathbf{x}}_0'|_{\mathbf{x},t} \coloneqq \mathbb{E}[\mathbf{X}_0^p | \mathbf{X}_t^p = \mathbf{x}]$, which gives us $w(\mathbf{X}_0^p) \approx w(\bar{\mathbf{x}}_0'|_{\mathbf{x},t}) + \nabla w(\bar{\mathbf{x}}_0'|_{\mathbf{x},t})^\top (\mathbf{X}_0^p - \bar{\mathbf{x}}_0'|_{\mathbf{x},t})$ and

$$\begin{aligned} g(\mathbf{x}, t) &\coloneqq \nabla_{\mathbf{x}} \log \mathbb{E}_{\mathbf{X}_0^p \sim p_{\mathbf{X}_0^p | \mathbf{X}_t^p}(\cdot|\mathbf{x})} [w(\mathbf{X}_0^p)] \\ &\approx \nabla_{\mathbf{x}} \log w(\bar{\mathbf{x}}_0'|_{\mathbf{x},t}). \end{aligned} \tag{10}$$

The conditional mean of the initial state for a given $\mathbf{x}$ at t can be obtained through the Tweedie's approach ([16, 27, 28]) as $\bar{\mathbf{x}}_0'|_{\mathbf{x},t} = \frac{1}{\sqrt{\bar{\alpha}(t)}}(\mathbf{x} + (1 - \bar{\alpha}(t)) \nabla_{\mathbf{x}} \log p_t(\mathbf{x}))$ where $\bar{\alpha}(t) = \exp(-\int_0^t \beta(s) \, ds)$. Importantly, the representation in (10) can be interpreted from a probabilistic perspective by considering an event E and a random variable Z that satisfy $p(Z \in E \mid \mathbf{x}) \propto w(\mathbf{x})$. Under this formulation, the approximation in (10) is equivalent to applying the posterior sampling and the denoising Tweedie mean technique [16]. See Appendix B.4 for details.

Finite-Difference Approximation for Hessian-Vector Multiplication. Given that we now have a tractable form for $\nabla_{\mathbf{x}} \log w(\bar{\mathbf{x}}_0'|_{\mathbf{x},t})$, we can apply the chain rule to represent $\nabla_{\mathbf{x}} \log w(\bar{\mathbf{x}}_0'|_{\mathbf{x},t})$ as follows.

$$\begin{aligned} &\nabla_{\mathbf{x}} \log w(\bar{\mathbf{x}}_0'|_{\mathbf{x},t}) \\ &= \frac{(\mathbf{I} + (1 - \bar{\alpha}(t)) H_{\log p_t}(\mathbf{x}))^\top}{\sqrt{\bar{\alpha}(t)}} \nabla_{\bar{\mathbf{x}}_0'|_{\mathbf{x},t}} \log w(\bar{\mathbf{x}}_0'|_{\mathbf{x},t}) \end{aligned} \tag{11}$$

where $H_{\log p_t}(\mathbf{x})$ is the Hessian matrix of $\log p_t(\mathbf{x})$, which can be computationally demanding in practice, particularly when the pre-trained score function is implemented as a

high-capacity model. Since $H_{\log p_t}(\mathbf{x}) = \nabla^2_{\mathbf{x}} \log p_t(\mathbf{x})$ is symmetric, the transpose can be dropped. To enhance the practicality of our approach, we propose approximating the Hessian matrix-vector multiplication in (11) by directional derivative approximation using finite differences with a sufficiently small $\epsilon > 0$ as

$$\begin{aligned}&\nabla_{\mathbf{x}} \log\ p_t(\mathbf{x}) + \epsilon H_{\log p_t}(\mathbf{x}) \nabla_{\bar{\mathbf{x}}'_0|\mathbf{x},t} \log w(\bar{\mathbf{x}}'_0|_{\mathbf{x},t}) \\ &\approx \nabla_{\mathbf{x}} \log p_t(\mathbf{x} + \epsilon \nabla_{\bar{\mathbf{x}}'_0|\mathbf{x},t} \log w(\bar{\mathbf{x}}'_0|_{\mathbf{x},t})). \end{aligned} \tag{12}$$

Substituting (12) into the score decomposition (10) yields the following first-order approximation of the guidance term $g(\mathbf{x}, t) \approx \tilde{g}^{(1)}(\mathbf{x}, t)$:

$$\tilde{g}^{(1)}(\mathbf{x}, t) := \frac{\nabla_{\bar{\mathbf{x}}'_0|\mathbf{x},t} \log w(\bar{\mathbf{x}}'_0|_{\mathbf{x},t})}{\sqrt{\bar{\alpha}(t)}} + \frac{\nabla_{\mathbf{x}} \log p_t(\mathbf{x} + \epsilon \nabla_{\bar{\mathbf{x}}'_0|\mathbf{x},t} \log w(\bar{\mathbf{x}}'_0|_{\mathbf{x},t})) - \nabla_{\mathbf{x}} \log p_t(\mathbf{x})}{\epsilon (1 - \bar{\alpha}(t))^{-1} \sqrt{\bar{\alpha}(t)}}. \tag{13}$$

Note that the score approximation in (13) does not involve any second-order derivatives of the base density p or the weight function w. This property is particularly valuable when performing weighted sampling with large-scale diffusion models, where computing second-order derivatives of the score model is often prohibitive.

3.2. Approximation Accuracy

We now analyze the accuracy of the approximation $\tilde{g}^{(1)}$, which later will drive the design of a new approximation $\tilde{g}$, presented in Sec. 3.3, that incorporates an uncertainty-aware, time-dependent scheduling policy. In particular, we establish Theorem 1 providing an upper bound on the Euclidean norm discrepancy between the approximation and the true guidance, with a detailed proof provided in Appendix B.

Assumption 1. (Bounded w and its derivatives) For all $\mathbf{x} \in \mathbb{R}^d$, we have $w(\mathbf{x}) \geq m$ for some $m > 0$ and $\|\nabla_{\mathbf{x}} \log w(\mathbf{x})\| \leq \eta$. Moreover, w is twice differentiable and its Hessian is uniformly bounded as $\|H_w(\mathbf{z})\| \leq \eta_2$ for all $\mathbf{z} \in \mathbb{R}^d$, where $H_w(\mathbf{x})$ is the Hessian matrix of $w(\mathbf{x})$.

Assumption 1 ensures that w is bounded away from zero and its norm of log-gradient and Hessian are uniformly bounded.

Assumption 2. (Bounded log PDF derivatives) Suppose that $\text{supp}(p_0)$ is contained in a bounded set, i.e., $\|\mathbf{y}\| \leq B$ for all $\mathbf{y} \in \text{supp}(p_0)$. For all $\mathbf{x} \in \mathbb{R}^d$, $\mathbf{y} \in \text{supp}(p_0)$, and $t \in [0, T]$, we have $\|\nabla_{\mathbf{x}} \log p_{\mathbf{X}^p_0|\mathbf{X}^p_t}(\mathbf{y}|\mathbf{x})\| \leq \gamma_t$, $\|H_{\log p_t}(\mathbf{x})\| \leq \zeta_t$, and $\|H_{\log p_t}(\mathbf{x}) - H_{\log p_t}(\mathbf{y})\| \leq L_t \|\mathbf{x} - \mathbf{y}\|$.

The bounded norm of the score function and Hessian of the log-likelihood are standard assumptions in the analysis of SGMs [29–32]. The compact support condition is mild in practice, as data distributions such as images are inherently bounded (e.g., pixel values in $[0, 1]^d$). Under the variance preserving (VP) SDE, the conditional score is bounded whenever $\mathbf{y}$ lies in a compact set. We additionally assume Lipschitz continuity of the Hessian.

Theorem 1. (Approximation gap) *Suppose Assumptions 1–2 hold. Then, for all $\mathbf{x} \in \mathbb{R}^d$ and $t \in [0, T]$, and for any $\epsilon > 0$, the deviation between the true guidance term $g(\mathbf{x}, t)$ and the approximation $\tilde{g}^{(1)}(\mathbf{x}, t)$, denoted by $\Delta^{(1)}(\mathbf{x}, t)$, is upper-bounded as follows.*

$$\Delta^{(1)}(\mathbf{x}, t) = g(\mathbf{x}, t) - \tilde{g}^{(1)}(\mathbf{x}, t) \ \textit{and}\ \|\Delta^{(1)}(\mathbf{x}, t)\| \leq u(\mathbf{x}, t),$$
$$\textit{where} \quad u(\mathbf{x}, t) := \frac{(1 - \bar{\alpha}(t)) L_t \eta^2 \epsilon}{2\sqrt{\bar{\alpha}(t)}} + \Big(\frac{\gamma_t \eta_2}{2m} + \frac{\big(1 + (1 - \bar{\alpha}(t))\zeta_t\big)\eta\eta_2}{2m\sqrt{\bar{\alpha}(t)}}\Big) \mathbb{E}\big[\|\mathbf{X}^p_0 - \bar{\mathbf{x}}'_0|_{\mathbf{x},t}\|^2 | \mathbf{X}^p_t = \mathbf{x}\big]. \tag{14}$$

Remark 1 (Bounded approximation gap by variance)**.** The norm discrepancy between the true guidance term $g(\mathbf{x}, t)$ and its approximation $\tilde{g}^{(1)}(\mathbf{x}, t)$ is bounded in terms of $\bar{\alpha}(t)$ and $\mathbb{E}\left[\|\mathbf{X}^p_0 - \bar{\mathbf{x}}'_0|_{\mathbf{x},t}\|^2 | \mathbf{X}^p_t = \mathbf{x}\right]$, which is the *trace of the covariance matrix of $\mathbf{X}^p_0$ whose PDF is conditioned on the observation $\mathbf{x}$ at t*. As $t \to 0$, this conditional variance converges to zero, resulting in the second term of $u(\mathbf{x}, t)$ vanishing, making the gap negligible. Additionally, the first term also vanishes as $\bar{\alpha}(t) \to 1$, provided L_t, γ_t, ζ_t remain bounded as $t \to 0$—a mild condition satisfied by common noise schedules.

3.3. Uncertainty-Adaptive Scheduling

Theorem 1 and Remark 1 show that the approximation error decreases as $t \to 0$, i.e., toward the end of the reverse-time sampling trajectory, which explains why early reverse steps are more unstable. This behavior aligns with prior observations in various existing methods for image generation, where early-stage instability has led to resampling heuristics or manually tuned guidance strengths [14, 19]. However, such approaches may incur significant computational cost, particularly in high-dimensional settings.

To address this issue systematically, we propose a time-dependent function $\tau : [0, T] \to [0, 1]$ that modulates the contribution of the guidance term $\tilde{g}^{(1)}$ in the approximated score function as $\nabla_{\mathbf{x}} \log q_t(\mathbf{x}) \approx \nabla_{\mathbf{x}} \log p_t(\mathbf{x}) + \tau(t)\tilde{g}^{(1)}(\mathbf{x}, t)$. That is, we consider the form of $\tilde{g}(\mathbf{x}, t) = \tau(t)\tilde{g}^{(1)}(\mathbf{x}, t)$. The magnitude of $\tau(t)$ is selected to minimize the mean squared error between the true score function

and the approximation

$$r(t;\tau) = \mathbb{E}[\|\nabla_{\mathbf{X}_t^q} \log q_t(\mathbf{X}_t^q) - \nabla_{\mathbf{X}_t^q} \log p_t(\mathbf{X}_t^q) - \tau(t)\tilde{g}^{(1)}(\mathbf{X}_t^q, t)\|^2], \quad (15)$$

where the expectation is taken over $\mathbf{X}_t^q \sim q_t$, the state at time t under the weighted sampling diffusion process. By Theorem 1, the approximation error $\Delta^{(1)}(\mathbf{x}, t)$ is controlled by the conditional variance term $\mathbb{E}\big[\|\mathbf{X}_0^p - \mathbb{E}[\mathbf{X}_0^p \mid \mathbf{X}_t^p]\|^2 \mid \mathbf{X}_t^p = \mathbf{x}\big]$, which admits a deterministic upper bound expressible as a finite sum of $\frac{(1-\bar{\alpha}(t))^i}{\bar{\alpha}(t)^j}$ terms under the VP SDE (see Appendix B.2). We then have the following Proposition 1 with a detailed proof provided in Appendix B.2.

Assumption 3. (Simplifying orthogonality condition) The approximation error $\Delta^{(1)}(\mathbf{x}, t)$ satisfies $\mathbb{E}[\Delta^{(1)}(\mathbf{X}_t^q, t)] = \mathbf{0}$ and is uncorrelated with g, i.e., $\mathbb{E}[g(\mathbf{X}_t^q, t)^\top \Delta^{(1)}(\mathbf{X}_t^q, t)] = 0$. This is adopted for analytical tractability; see Remark 2 for an alternative without this assumption.

Proposition 1 (Uncertainty-Adaptive Scheduling)**.** *Suppose that Assumptions 1–3 hold. Then, there exists a deterministic upper envelope $r'(t;\tau)$ such that $r(t;\tau) \le r'(t;\tau)$, where $r'(t;\tau) = (c_1 + \sum_{(i,j)\in\mathcal{I}} c_{(i,j)} \cdot \frac{(1-\bar{\alpha}(t))^i}{\bar{\alpha}(t)^j})\tau(t)^2 - 2c_1\tau(t) + c_1$ with constants $c_1 > 0$, $c_{(i,j)} \ge 0$ and a finite index set $\mathcal{I} \subset \mathbb{R}_{>0} \times \mathbb{R}_{\ge 0}$. When $c_1 > 0$, the minimizer $\tau^*(t)$ of $r'(t;\tau)$ is given as*

$$\tau^*(t) = \arg\min_{\tau} r'(t;\tau) = \left(1 + \sum_{(i,j)\in\mathcal{I}} \frac{c_{(i,j)}}{c_1} \cdot \frac{(1-\bar{\alpha}(t))^i}{\bar{\alpha}(t)^j}\right)^{-1}. \quad (16)$$

Remark 2 (Increasing Confidence and Practical Scheduler Design)**.** To minimize $r'(t;\tau)$, which is the upper bound of the score function gap (15), the corresponding optimal scheduling function $\tau^*(t)$ is given by (16). Note that as $t \to 0$, $\bar{\alpha}(t) \to 1$ thereby $\tau^*(t)$ converges to 1. This behavior is expected since the uncertainty $u(\mathbf{x}, t)$ in Theorem 1 vanishes as $t \to 0$, reflecting that the guidance approximation error becomes negligible towards the end of the diffusion process and thus justifying full reliance on the approximation. Although computing $\tau(t)$ exactly requires access to unknown problem-specific quantities (see Appendix B.2), each term of the form $\frac{(1-\bar{\alpha}(t))^i}{\bar{\alpha}(t)^j}$ vanishes at $t \to 0$ and increases monotonically with t. This motivates a practical single-parameter approximation $\tau^*(t) \approx \big(1 + c(1-\bar{\alpha}(t))^2/\bar{\alpha}(t)^2\big)^{-1}$, where $c > 0$ is a tunable constant. This approximation implies that $\tau^*(t)$ is monotone *decreasing* in the forward diffusion time t (and hence *increasing along the reverse-time sampling direction*), for any positive variance schedule $\beta(t)$, offering a simple and effective control mechanism.

Without Assumption 3, minimizing a tractable upper bound of equation (15) yields a hard gating scheduler $\tau^*(t) \in \{0, 1\}$ (a step function in t; see Appendix B.2). For empirical evaluation of different choices of c and the approximation form τ^*, see our ablation study in Appendix C.2.

Relation to Existing Methods. Our scheduler parallels heuristic strategies such as guidance learning rate control [14] and tempering [19], but derives directly from uncertainty analysis with a single tunable parameter, generalizing well across tasks and SGMs.

Finally, combining $\tilde{g}^{(1)}(\mathbf{x}, t)$ in (13) and the parameterized $\tau(t)$ in Remark 2, we suggest using the following approximated score function for weighted sampling backward diffusion process with only two predefined hyperparameters ϵ and c as follows.

$$\nabla_{\mathbf{x}} \log q_t(\mathbf{x}) \approx \nabla_{\mathbf{x}} \log p_t(\mathbf{x}) + \tilde{g}(\mathbf{x}, t) \text{ where } \tilde{g}(\mathbf{x}, t) = \frac{\bar{\alpha}(t)^2}{\bar{\alpha}(t)^2 + c(1-\bar{\alpha}(t))^2}\tilde{g}^{(1)}(\mathbf{x}, t). \quad (17)$$

Remark 3 (Training-Free Weighted Sampling)**.** The proposed approximation form in (17) demonstrates that, for a given score function $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ and weight function $w(\mathbf{x})$, *weighted sampling can be approximated without additional training*. This presents a substantial advantage over existing diffusion model fine-tuning methods or cross-entropy methods for weighted sampling, which require learning the density of q through training for each specific w.

Remark 4 (High Scalability and Computational Efficiency)**.** The additional computational cost introduced by our approach, compared to the base score-based sampling, is minimal. Calculating $\bar{\mathbf{x}}_0'|_{\mathbf{x},t}$ involves only a linear transformation of $\mathbf{x}$ and the precomputed score function. Aside from this negligible cost, the potential overhead is the computation of $\nabla_{\bar{\mathbf{x}}_0'|_{\mathbf{x},t}} \log w(\bar{\mathbf{x}}_0'|_{\mathbf{x},t})$ in (13). This is efficiently achieved with a single gradient backpropagation step on w with respect to the realization $\bar{\mathbf{x}}_0'|_{\mathbf{x},t}$ and one inference step of the score function, i.e., evaluating $\nabla_{\mathbf{x}} \log p_t(\mathbf{x} + \epsilon\Delta)$ in addition to $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$. Overall, this additional cost is *significantly less demanding* than methods requiring high-order derivatives of densities or resampling.

4. Experiments

We conducted comprehensive experiments to evaluate LAGS across a range of tasks. Specifically, we aimed to: *(i) assess its ability to approximate complex target distributions, especially multimodal structures and exponential behavior weights; (ii) demonstrate its effectiveness on*

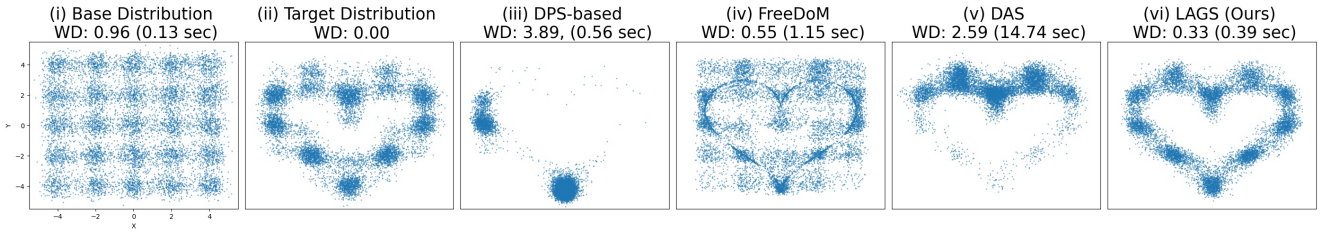

Figure 1. ***Left* to *Right***: Base distribution, target distribution, and estimated sample densities. Wasserstein Distance and runtime (in seconds) are reported for DPS, FreeDoM, DAS, and LAGS (ours). Our method achieves the ***lowest Wasserstein Distance (WD) and fastest runtime***.

high-dimensional weighted sampling tasks; (iii) quantify the computational efficiency gains enabled by our method.

Across various settings, our method consistently achieves the lowest Wasserstein Distance to the target distribution or yields the highest average weights. Especially in high-dimensional scenarios, it delivers a roughly 4.7× speedup over prior state-of-the-art methods while maintaining or improving task performance.

Baselines. We compare our approach against the following representative training-free SGM guidance methods. DPS [16], originally designed for inverse problems, can be adapted for weighted sampling by leveraging its denoising mean estimator. FreeDoM [14], which stabilizes the reverse-time SDE sampling via a time-travel strategy; DAS [19], which uses sequential Monte Carlo with adaptive particle weighting, and has shown good performance on tasks involving high-modality base distributions. We also tested a single-particle version of DAS requiring less computational power for comparison. Pretrained SGMs such as [33, 34] are also utilized as baselines.

4.1. Training-Free Weighted Sampling over Multimodal Densities

Setup. We evaluate the proposed method on a 2D weighted sampling task in which the base distribution is a mixture of 25 Gaussians (Fig. 1 (i)), and the weight concentrates along a heart-shaped manifold, defining a structured target distribution (Fig. 1 (ii)). Implementation details are provided in Appendix C.1. The task is analytically well-defined, enabling quantitative evaluation via the Wasserstein Distance (WD) between the generated samples and the ground-truth target distribution. We report both WD and wall-clock time required to generate 10^4 samples along with the generated samples by each algorithm in Fig. 1 (iii)–(vi).

Results. As shown in Fig. 1, our method (LAGS) achieves both the lowest WD and fastest runtime. It generates all samples in 0.33 seconds without computing any second-order derivatives or resampling. In contrast, baselines such as FreeDoM and DAS incur higher computational cost due to iterative refinement or generating particles for sequential Monte Carlo methods.

FreeDoM provides broad coverage but often produces samples outside the support of the target distribution (where $q(\mathbf{x}) \approx 0$). DAS improves precision by concentrating on high-density regions, but this comes at substantial computational expense. Our method achieves both high coverage and high precision, accurately approximating the support of $q(\mathbf{x})$ with markedly improved efficiency. We provide additional examples in Appendix C.1.

4.2. Application to Training-Free Text-to-Image Alignment with Human Preference Scores

We further evaluated our method on high-dimensional generative tasks, focusing on text-to-image synthesis guided by human preference scores. In this setting, the base SGM is a prompt-conditioned image generator, and the weight function $w(\mathbf{x})$ corresponds to a predefined human preference score for a given text-image pair. The objective is to maximize the expected preference score—i.e., average weight—by prioritizing samples that are better aligned with human preference. Such scenarios demand both high alignment accuracy and computational efficiency, as they directly impact user experience and system scalability in real-world deployments.

Limitations of the experimental scope. We focus on training-free guidance with a target weight function, distinct from approaches based on text embedding manipulation [35], model fine-tuning [36], or Large Language Models [37]. For fairness, we restrict comparisons to training-free guidance methods.

Setup, Metrics, and Fairness. We follow the evaluation protocol where $w(\mathbf{x})$ is defined via PickScore [38], a learned alignment metric trained from human preferences. All methods use the same pretrained SGMs—Stable Diffusion (SD) [33] and Stable Diffusion XL (SDXL) [34]—with a fixed number of backward diffusion steps (T).

For SD, we adopt a prompt set from [19]. For SDXL, we combine prompts from [39], [19], and additional manually curated samples to assess generalization. Evaluation in-

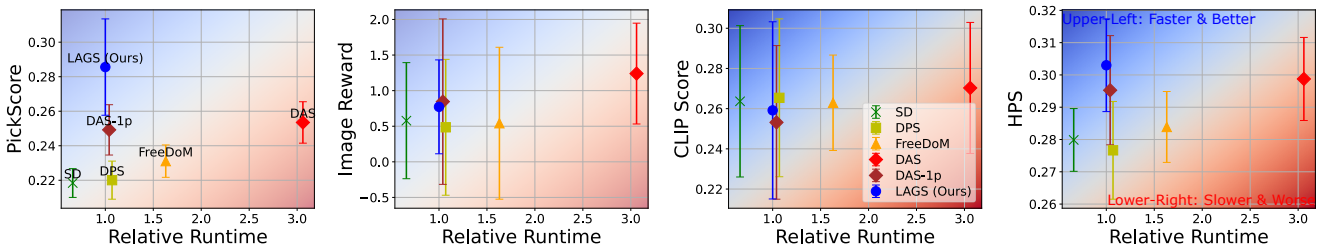

Figure 2. **Results on SD**: Comparative performance across PickScore, ImageReward, CLIP Score, and HPS (left to right), with runtime normalized to LAGS. Our method achieves the best performance in PickScore and HPS with the lowest runtime among the guidance methods. We adopt the benchmark set of prompts from [19].

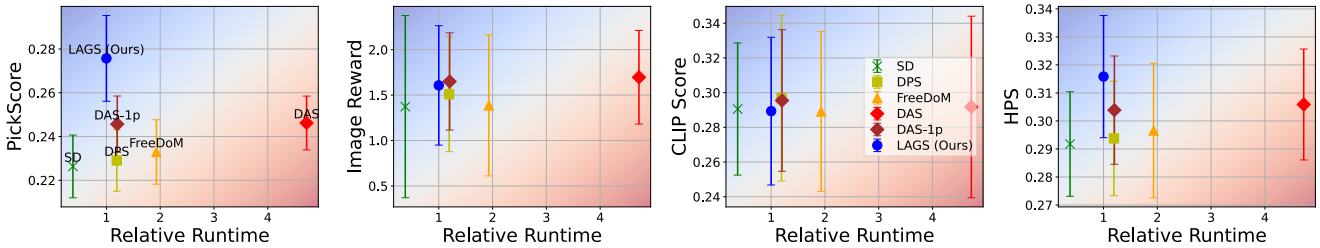

Figure 3. **Results on SDXL**: Evaluation on the same metrics as in Fig. 2. LAGS consistently achieves the best performance in PickScore (target metric) and HPS with the lowest runtime among the guidance methods, breaking the typical trade-off between high score and computation complexity.

cludes expected weight $\mathbb{E}[w(\mathbf{X}^q)]$, PickScore, and runtime. We also report secondary metrics: CLIP Score [40], ImageReward [41], and Human Preference Score (HPS) [42].

Figures 2 and 3 show the results on SD and SDXL, respectively. Each subplot presents one metric (PickScore, ImageReward, CLIP Score, HPS from left to right), with runtime shown on the horizontal axis relative to our method (normalized to $1\times$). For SDXL, our method completes sampling in 85 seconds per image, while DAS requires approximately $4.72\times$ more time.

Results. The proposed method, LAGS, consistently achieves the *highest PickScore* and *lowest runtime* across both SD and SDXL settings, as shown in the leftmost columns of Figures 2 and 3. This represents a substantial improvement over existing baselines, which typically exhibit a trade-off between target performance and computational cost. In contrast, LAGS breaks this trend by achieving both higher weight, i.e., PickScore, and faster sampling.

Among baselines, DAS delivers strong performance but incurs a higher computational cost. FreeDoM and DPS improve upon the raw SD and SDXL outputs in terms of PickScore, but remain notably less effective than LAGS in both performance and efficiency.

To qualitatively illustrate the impact of PickScore on the generated outputs, we present visual examples in Fig. 4. The top row depicts samples with the *lowest* PickScores for a given prompt, while the bottom row shows those with the *highest* PickScores among outputs generated by each method. Samples with low PickScores often display semantic or stylistic failures, such as omission of positional cues (e.g., "next to a window"), incorrect object counts (e.g., four apples instead of three), or reduced representational fidelity (the associated text prompts are shown in Fig. 4). Consistent with the trends shown in Figures 2-3, the highest PickScores for the selected prompts are attained by our method. Additional qualitative examples are provided in Appendix C.2.2.

Cross-Metric Generalization. To assess whether improvements in PickScore extend to other alignment metrics, we evaluate ImageReward [41], CLIP Score [40], and HPS [42]. As shown in Fig. 2, LAGS achieves the highest HPS on SD, while DAS performs better on ImageReward and CLIP Score. On SDXL (Fig. 3), LAGS still outperforms all baselines on PickScore and HPS, where the margin against DAS on ImageReward and CLIP score is minor (less than 0.1 and 0.06, respectively), despite being faster. These results indicate that LAGS's reward-aligned sampling generalizes well to diverse human-preference proxies with high efficiency.

We further report image quality metrics in Table 1 using the widely adopted no-reference measure BRISQUE [43] and MANIQA metrics [44]. While DAS achieves the best overall image quality, our method attains comparable scores with only a modest gap—within the variability across guidance methods—despite its substantially lower runtime.

Hyperparameter Sensitivity. Our method introduces only a small number of hyperparameters. In all experiments, we fix the confidence parameter to $c = 10$ across models and tasks, yet LAGS consistently outperforms baseline methods. By construction, the scheduler in (16)

Figure 4. ***Sampling results with the proposed lightweight guidance*** $\tilde{g}(\mathbf{x}, t)$ ***on a high-dimensional domain.*** *Top row:* Samples with the **lowest PickScore**; *Bottom row:* those with the **highest PickScore**, selected from all generations. Prompts (left to right): *"A cat sitting inside a cardboard box next to a window", "Three apples on a table", "A majestic, photorealistic alien spaceship drifting before a vast galaxy", "A photo-realistic image of flying lion with blue butterfly wings", "A red sports car in the style of Vincent van Gogh"*. Notably, for all prompts, the highest PickScore samples are achieved by our method. Additional qualitative results are provided in Appendix C.2.2.

Method	Runtime [× 85s] ↓	BRISQUE ↓	MANIQA ↑
SDXL (default)	0.377	21.48 ± 12.45	0.733 ± 0.139
DPS	1.193	24.52 ± 13.03	0.720 ± 0.139
FreeDoM	1.930	34.49 ± 15.20	0.711 ± 0.144
DAS	4.722	**19.55** ± **12.75**	0.720 ± 0.145
DAS-1P	1.202	22.56 ± 11.97	**0.728** ± **0.130**
LAGS (ours)	**1.000**	26.40 ± 12.99	0.720 ± 0.134

Table 1. Runtime and no-reference image quality measures (BRISQUE, MANIQA). Values are mean ± standard deviation. Bold indicates best among guidance methods.

smoothly modulates the guidance strength between 0 and 1 (with $\tau(0) = 1$), and we observe that performance remains stable over a wide range of c values (approximately $0.1 \leq c \leq 20$). Appendix C.2.3 provides a detailed ablation study on the effect of varying c and further confirms the robustness of the proposed scheduler.

Computational Efficiency. A key advantage of LAGS lies in its computational efficiency achieved by avoiding second-order derivative computations and eliminating the need for resampling techniques. Notably, the efficiency gain becomes more pronounced as model size increases. On SD, LAGS achieves a 6% speedup over DAS-1p and DPS. With the larger SDXL model, the improvement grows to 16.7%, demonstrating the method's scalability with model complexity. Compared to resampling-based approaches, the gap is even more substantial, for example, LAGS is 1.9× and 4.7× faster than FreeDoM and DAS, respectively, on SDXL.

4.3. Additional Applications

Owing to its efficiency and scalability, the proposed method extends naturally to a range of practical scenarios with diverse weight functions $w(\mathbf{x})$. Below, we summarize additional applications: **Sampling for Fairness.** Using classifier outputs as weights enables our method to target underrepresented or misclassified groups, serving as a training-free tool for mitigating class bias. We demonstrate this in Appendix C.3 and C.4 across multiple datasets; **Sampling Control Beyond Prompts.** Beyond prompt conditioning, LAGS supports rapid fine-grained control via custom $w(\mathbf{x})$. In Appendix C.5, we showcase frequency- and color-sensitive sampling by emphasizing spectral features through $w(\mathbf{x})$ *without prompt control*, highlighting LAGS's versatility for controllable generation.

5. Conclusion

We first propose a lightweight first-order approximation that avoids second-order gradient computations of the generative model. To mitigate potential approximation gaps, we introduce an uncertainty-adaptive scaling mechanism. Together, these components enable LAGS to achieve both strong empirical performance and excellent scalability with a fraction of the computational cost of state-of-the-art baselines. As larger SGMs continue to be developed and released, we anticipate that the proposed Hessian-free, lightweight formulation will become increasingly relevant for efficient and stable weighted sampling.

Acknowledgements

This work was supported in part by the National Science Foundation under Grant 2148224 and Grant 2443857; in part by the Army Research Office (ARO) under Award W911NF2310062; in part by the Office of Naval Research (ONR) under Award N000142412542; CNS-2212202; in part by Office of the Under Secretary of Defense for Research and Engineering (OUSD R&E), National Institute of Standards and Technology (NIST), and industry partners as specified in the Resilient and Intelligent NextG Systems (RINGS) Program; and in part by Wireless Networking and Communications Group (WNCG)/6G@UT.

References

[1] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Advances in Neural Information Processing Systems*, 2020. 1, 2, 14, 21, 28

[2] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," in *International Conference on Learning Representations*, 2021. 22, 23, 32

[3] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," in *International Conference on Learning Representations*, 2021. 1, 2, 14

[4] P. Dhariwal and A. Nichol, "Diffusion models beat GANs on image synthesis," in *Advances in Neural Information Processing Systems*, 2021. 1, 14

[5] A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, and I. Sutskever, "Zero-shot text-to-image generation," in *International Conference on Machine Learning*, 2021. 1

[6] Z. Kong, W. Ping, J. Huang, K. Zhao, and B. Catanzaro, "DiffWave: A versatile diffusion model for audio synthesis," in *International Conference on Learning Representations*, 2021. 1

[7] N. Chen, Y. Zhang, H. Zen, R. J. Weiss, M. Norouzi, and W. Chan, "WaveGrad: Estimating gradients for waveform generation," in *International Conference on Learning Representations*, 2021. 1

[8] T. Lee, J. Park, H. Kim, and J. G. Andrews, "Generating high dimensional user-specific wireless channels using diffusion models," *IEEE Transactions on Wireless Communications*, vol. 25, p. 2907–2921, 2025. 1

[9] J. Zhang, S. Xu, Z. Zhang, C. Li, and L. Yang, "A denoising diffusion probabilistic model-based digital twinning of ISAC MIMO channel," *IEEE Internet of Things Journal*, vol. 12, no. 15, pp. 29121 – 29134, 2024. 1

[10] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. Springer Texts in Statistics, New York: Springer, 2nd ed., 2004. 1

[11] J. Byrd and Z. C. Lipton, "What is the effect of importance weighting in deep learning?," in *International Conference on Machine Learning*, June 2019.

[12] C. Cortes, Y. Mansour, and M. Mohri, "Learning bounds for importance weighting," in *Advances in Neural Information Processing Systems*, 2010.

[13] F. Kamiran and T. Calders, "Data preprocessing techniques for classification without discrimination," *Knowledge and Information Systems*, vol. 33, pp. 1–33, 2012. 1

[14] J. Yu, Y. Wang, C. Zhao, B. Ghanem, and J. Zhang, "FreeDoM: Training-free energy-guided conditional diffusion model," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023. 1, 3, 4, 5, 6, 15, 22, 24

[15] J. Song, A. Vahdat, M. Mardani, and J. Kautz, "Pseudoinverse-guided diffusion models for inverse problems," in *International Conference on Learning Representations*, 2023.

[16] H. Chung, J. Kim, M. T. Mccann, M. L. Klasky, and J. C. Ye, "Diffusion posterior sampling for general noisy inverse problems," in *International Conference on Learning Representations*, 2022. 1, 3, 6, 15, 21, 22, 24

[17] Y. Guo, H. Yuan, Y. Yang, M. Chen, and M. Wang, "Gradient guidance for diffusion models: An optimization perspective," *arXiv preprint arXiv:2404.14743*, 2024. 1

[18] A. Phillips, H.-D. Dau, M. J. Hutchinson, V. De Bortoli, G. Deligiannidis, and A. Doucet, "Particle denoising diffusion sampler," in *International Conference on Machine Learning*, 2024. 1

[19] S. Kim, M. Kim, and D. Park, "Test-time alignment of diffusion models without reward over-optimization," in *International Conference on Learning Representations*, 2025. 1, 3, 4, 5, 6, 7, 14, 15, 21, 22, 24

[20] B. D. O. Anderson, "Reverse-time diffusion equation models," *Stochastic Processes and their Applications*, vol. 12, no. 3, pp. 313–326, 1982. 2

[21] A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte, and L. Van Gool, "RePaint: Inpainting using denoising diffusion probabilistic models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 2

[22] A. Q. Nichol and P. Dhariwal, "Improved denoising diffusion probabilistic models," in *International Conference on Machine Learning*, 2021.

[23] J. Choi, S. Kim, Y. Jeong, Y. Gwon, and S. Yoon, "ILVR: Conditioning method for denoising diffusion probabilistic models," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021. 2

[24] J. V. Beck, K. D. Cole, A. Haji-Sheikh, and B. Litkouhl, *Heat conduction using Green's function*. Taylor & Francis, 2015. 3

[25] D. G. Duffy, *Green's functions with applications*. Chapman and Hall/CRC, 2015.

[26] E. N. Economou, *Green's functions in quantum physics*, vol. 7. Springer Science & Business Media, 2006. 3

[27] B. Efron, "Tweedie's formula and selection bias," *Journal of the American Statistical Association*, vol. 106, no. 496, pp. 1602–1614, 2011. 3

[28] K. Kim and J. C. Ye, "Noise2Score: Tweedie's approach to self-supervised image denoising without clean images," *Advances in Neural Information Processing Systems*, 2021. 3

[29] H. Chen, H. Lee, and J. Lu, "Improved analysis of score-based generative modeling: User-friendly bounds under min-

imal smoothness assumptions," in *International Conference on Machine Learning*, 2023. 4

[30] V. De Bortoli, "Convergence of denoising diffusion models under the manifold hypothesis," *Transactions on Machine Learning Research*, 2022.

[31] S. Chen, S. Chewi, J. Li, Y. Li, A. Salim, and A. Zhang, "Sampling is as easy as learning the score: theory for diffusion models with minimal data assumptions," in *International Conference on Learning Representations*, 2023.

[32] V. De Bortoli, J. Thornton, J. Heng, and A. Doucet, "Diffusion Schrödinger bridge with applications to score-based generative modeling," *Advances in Neural Information Processing Systems*, 2021. 4

[33] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10684–10695, June 2022. 6

[34] D. Podell, Z. English, K. Lacey, A. Blattmann, T. Dockhorn, J. Müller, J. Penna, and R. Rombach, "SDXL: Improving latent diffusion models for high-resolution image synthesis," in *International Conference on Learning Representations*, 2024. 6, 14

[35] S. Kang, W. Han, D. Ju, and S. J. Hwang, "Rare text semantics were always there in your diffusion transformer," in *Neural Information Processing Systems*, 2025. 6

[36] K. Black, M. Janner, Y. Du, I. Kostrikov, and S. Levine, "Training diffusion models with reinforcement learning," in *International Conference on Learning Representations*, 2024. 6

[37] S. Chen, M. Xu, J. Ren, Y. Cong, S. He, Y. Xie, A. Sinha, P. Luo, T. Xiang, and J.-M. Perez-Rua, "GenTron: Diffusion transformers for image and video generation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. 6

[38] Y. Kirstain, A. Polyak, U. Singer, S. Matiana, J. Penna, and O. Levy, "Pick-a-Pic: An open dataset of user preferences for text-to-image generation," *Advances in Neural Information Processing Systems*, 2023. 6

[39] X. Wu, K. Sun, F. Zhu, R. Zhao, and H. Li, "Better aligning text-to-image models with human preference," 2023. 6, 24

[40] J. Hessel, A. Holtzman, M. Forbes, R. L. Bras, and Y. Choi, "CLIPScore: A reference-free evaluation metric for image captioning," *arXiv preprint arXiv:2104.08718*, 2021. 7

[41] J. Xu, X. Liu, Y. Wu, Y. Tong, Q. Li, M. Ding, J. Tang, and Y. Dong, "ImageReward: Learning and evaluating human preferences for text-to-image generation," *Advances in Neural Information Processing Systems*, 2023. 7

[42] X. Wu, Y. Hao, K. Sun, Y. Chen, F. Zhu, R. Zhao, and H. Li, "Human preference score v2: A solid benchmark for evaluating human preferences of text-to-image synthesis," *arXiv preprint arXiv:2306.09341*, 2023. 7, 27

[43] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012. 7

[44] S. Yang, T. Wu, S. Shi, S. Lao, Y. Gong, M. Cao, J. Wang, and Y. Yang, "MANIQA: Multi-dimension attention network for no-reference image quality assessment," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 7

[45] R. Y. Rubinstein and D. P. Kroese, *Simulation and the Monte Carlo method.* John Wiley & Sons, 2016. 14

[46] A. Grover, J. Song, A. Agarwal, K. Tran, E. J. Horvitz, and S. Ermon, "Bias correction of learned generative models using likelihood-free importance weighting," *Advances in Neural Information Processing Systems*, 2019. 14

[47] A. Antoniou, A. Storkey, and H. Edwards, "Data augmentation generative adversarial networks," *arXiv preprint arXiv:1711.04340*, 2017.

[48] G. Li, W. Cai, L. Tian, L. Yang, Z. Weihua, and Z. Wu, "Efficient augmented radial basis function for reliability analysis with dynamic importance sampling," *Engineering Optimization*, vol. 57, pp. 1486–1505, 2024. 14

[49] R. Rubinstein, "The cross-entropy method for combinatorial and continuous optimization," *Methodology and Computing in Applied Probability*, vol. 1, pp. 127–190, 1999. 14

[50] R. Rubinstein, "Combinatorial optimization, cross-entropy, ants and rare events," *Stochastic Optimization: Algorithms and Applications*, pp. 303–363, 2001. 14

[51] D. Rezende and S. Mohamed, "Variational inference with normalizing flows," in *International Conference on Machine Learning*, 2015. 14

[52] I. Kobyzev, S. J. Prince, and M. A. Brubaker, "Normalizing flows: An introduction and review of current methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 11, pp. 3964–3979, 2020. 14

[53] T. Müller, B. McWilliams, F. Rousselle, M. Gross, and J. Novák, "Neural importance sampling," *ACM Transactions on Graphics (ToG)*, vol. 38, no. 5, pp. 1–19, 2019. 14

[54] Z. Xiang, X. He, Y. Zou, and H. Jing, "An importance sampling method for structural reliability analysis based on interpretable deep generative network," *Engineering with Computers*, vol. 40, no. 1, pp. 367–380, 2024. 14

[55] Q. Zhang and Y. Chen, "Diffusion normalizing flow," *Advances in Neural Information Processing Systems*, 2021. 14

[56] R. Cornish, A. Caterini, G. Deligiannidis, and A. Doucet, "Relaxing bijectivity constraints with continuously indexed normalising flows," in *International Conference on Machine Learning*, 2020. 14

[57] P. Vincent, "A connection between score matching and denoising autoencoders," *Neural Computation*, vol. 23, no. 7, pp. 1661–1674, 2011. 14

[58] A. Hyvärinen, "Estimation of non-normalized statistical models by score matching," *Journal of Machine Learning Research*, vol. 6, no. 4, 2005. 14

[59] F.-A. Croitoru, V. Hondru, R. T. Ionescu, and M. Shah, "Diffusion models in vision: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 9, pp. 10850–10869, 2023. 14

[60] E. Mansimov, E. Parisotto, J. L. Ba, and R. Salakhutdinov, "Generating images from captions with attention," in *International Conference on Learning Representations*, 2016. 14

[61] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, "Generative adversarial text to image synthesis," in *International Conference on Machine Learning*, 2016.

[62] B. Li, X. Qi, T. Lukasiewicz, and P. Torr, "Controllable text-to-image generation," *Advances in Neural Information Processing Systems*, 2019.

[63] J. Yu, Y. Xu, J. Y. Koh, T. Luong, G. Baid, Z. Wang, V. Vasudevan, A. Ku, Y. Yang, B. K. Ayan, *et al.*, "Scaling autoregressive models for content-rich text-to-image generation," *Transactions on Machine Learning Research*, 2022.

[64] A. Q. Nichol, P. Dhariwal, A. Ramesh, P. Shyam, P. Mishkin, B. Mcgrew, I. Sutskever, and M. Chen, "GLIDE: Towards photorealistic image generation and editing with text-guided diffusion models," in *International Conference on Machine Learning*, 2022. 14

[65] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2017. 14

[66] E. Richardson, Y. Alaluf, O. Patashnik, Y. Nitzan, Y. Azar, S. Shapiro, and D. Cohen-Or, "Encoding in style: a StyleGAN encoder for image-to-image translation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.

[67] M.-Y. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," in *Advances in Neural Information Processing Systems*, 2017. 14

[68] H. Liu, Z. Chen, Y. Yuan, X. Mei, X. Liu, D. Mandic, W. Wang, and M. D. Plumbley, "AudioLDM: text-to-audio generation with latent diffusion models," in *International Conference on Machine Learning*, 2023. 14

[69] C. Meng, Y. He, Y. Song, J. Song, J. Wu, J.-Y. Zhu, and S. Ermon, "SDEdit: Guided image synthesis and editing with stochastic differential equations," in *International Conference on Learning Representations*, 2022. 14

[70] B. L. Trippe, J. Yim, D. Tischer, D. Baker, T. Broderick, R. Barzilay, and T. S. Jaakkola, "Diffusion probabilistic modeling of protein backbones in 3D for the motif-scaffolding problem," in *International Conference on Learning Representations*, 2023.

[71] L. Wu, B. Trippe, C. Naesseth, D. Blei, and J. P. Cunningham, "Practical and asymptotically exact conditional sampling in diffusion models," in *Advances in Neural Information Processing Systems*, 2024. 14

[72] L. Rout, Y. Chen, A. Kumar, C. Caramanis, S. Shakkottai, and W.-S. Chu, "Beyond first-order Tweedie: Solving inverse problems using latent diffusion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. 14, 18

[73] L. Rout, N. Raoof, G. Daras, C. Caramanis, A. Dimakis, and S. Shakkottai, "Solving linear inverse problems provably via posterior sampling with latent diffusion models," *Advances in Neural Information Processing Systems*, 2024. 14

[74] D. Kim, Y. Kim, S. J. Kwon, W. Kang, and I.-C. Moon, "Refining generative process with discriminator guidance in score-based diffusion models," in *International Conference on Machine Learning*, 2023. 15

[75] A. Doucet, W. Grathwohl, A. G. Matthews, and H. Strathmann, "Score-based diffusion meets annealed importance sampling," in *Advances in Neural Information Processing Systems*, 2022. 15

[76] W. Deng, G. Lin, and F. Liang, "A contour stochastic gradient langevin dynamics algorithm for simulations of multimodal distributions," in *Advances in Neural Information Processing Systems*, 2020. 15

[77] W. Deng, G. Lin, and F. Liang, "An adaptively weighted stochastic gradient MCMC algorithm for Monte Carlo simulation and global optimization," *Statistics and Computing*, vol. 32, no. 4, p. 58, 2022.

[78] W. Deng, S. Liang, B. Hao, G. Lin, and F. Liang, "Interacting contour stochastic gradient Langevin dynamics," *arXiv preprint arXiv:2202.09867*, 2022. 15

[79] A. Bansal, H.-M. Chu, A. Schwarzschild, S. Sengupta, M. Goldblum, J. Geiping, and T. Goldstein, "Universal guidance for diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. 22

[80] J. Song, Q. Zhang, H. Yin, M. Mardani, M.-Y. Liu, J. Kautz, Y. Chen, and A. Vahdat, "Loss-guided diffusion models for plug-and-play controllable generation," in *International Conference on Machine Learning*, 2023. 22

[81] S. Sadat, O. Hilliges, and R. M. Weber, "Eliminating oversaturation and artifacts of high guidance scales in diffusion models," in *International Conference on Learning Representations*, 2025. 24

[82] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proceedings of IEEE/CVF International Conference on Computer Vision*, December 2015. 28

[83] Y. Kim, B. Na, M. Park, J. Jang, D. Kim, W. Kang, and I.-C. Moon, "Training unbiased diffusion models from biased dataset," in *International Conference on Learning Representations*, 2024. 29

[84] R. K. Bellamy, K. Dey, M. Hind, S. C. Hoffman, S. Houde, K. Kannan, P. Lohia, J. Martino, S. Mehta, A. Mojsilović, *et al.*, "AI fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias," *IBM Journal of Research and Development*, vol. 63, no. 4/5, pp. 1–20, 2019. 29

[85] S. Trewin, S. Basson, M. Muller, S. Branham, J. Treviranus, D. Gruen, D. Hebert, N. Lyckowski, and E. Manser, "Considerations for AI fairness for people with disabilities," *AI Matters*, vol. 5, no. 3, pp. 40–63, 2019.

[86] J.-M. John-Mathews, D. Cardon, and C. Balagué, "From reality to world. a critical perspective on AI fairness," *Journal of Business Ethics*, vol. 178, no. 4, pp. 945–959, 2022. 29

[87] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. 29

[88] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016. 29

[89] P. Pernias, D. Rampas, M. L. Richter, C. Pal, and M. Aubreville, "Würstchen: An efficient architecture for large-scale

text-to-image diffusion models," in *International Conference on Learning Representations*, 2024. 32