Inference-Time Alignment for Diffusion Models via Variationally Stable Doob's Matching

Jinyuan Chang[1,2], Chenguang Duan[3], Yuling Jiao[4], Yi Xu[5], and Jerry Zhijian Yang[5]

[1]Joint Laboratory of Data Science and Business Intelligence, Southwestern University of Finance and Economics, Chengdu, Sichuan 611130, China.
changjinyuan@swufe.edu.cn

[2]State Key Laboratory of Mathematical Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

[3]Institut für Geometrie und Praktische Mathematik, RWTH Aachen University, Im Süsterfeld 2, 52072 Aachen,Germany.
duan@igpm.rwth-aachen.de

[4]School of Artificial Intelligence, Wuhan University, Wuhan, Hubei 430072, China.
yulingjiaomath@whu.edu.cn

[5]School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei 430072, China.
{yixu.math,zjyang.math}@whu.edu.cn

February 5, 2026

Abstract. Inference-time alignment for diffusion models aims to adapt a pre-trained reference diffusion model toward a target distribution without retraining the reference score network, thereby preserving the generative capacity of the reference model while enforcing desired properties at the inference time. A central mechanism for achieving such alignment is guidance, which modifies the sampling dynamics through an additional drift term. In this work, we introduce variationally stable Doob's matching, a novel framework for provable guidance estimation grounded in Doob's h-transform. Our approach formulates guidance as the gradient of logarithm of an underlying Doob's h-function and employs gradient-regularized regression to simultaneously estimate both the h-function and its gradient, resulting in a consistent estimator of the guidance. Theoretically, we establish non-asymptotic convergence rates for the estimated guidance. Moreover, we analyze the resulting controllable diffusion processes and prove non-asymptotic convergence guarantees for the generated distributions in the 2-Wasserstein distance. Finally, we show that variationally stable guidance estimators are adaptive to unknown low dimensionality, effectively mitigating the curse of dimensionality under low-dimensional subspace assumptions.

Keywords: Controllable generative learning, inference-time alignment, Doob's h-transform, convergence rate

1 Introduction

Diffusion models (Sohl-Dickstein et al., 2015; Ho et al., 2020; Song and Ermon, 2019; Song et al., 2021) have emerged as powerful generative tools for sampling from data distributions, achieving remarkable success across diverse domains, including text-to-image and text-to-video generation (Ramesh et al., 2021), Bayesian inverse problems (Chung et al., 2023; Song et al., 2023; Chen et al., 2025; Chang et al., 2025b), and scientific applications (Bao et al., 2024; Li et al., 2025; Si and Chen, 2025; Ding et al., 2024; Uehara et al., 2025a). Recent years have witnessed the development of large-scale diffusion models pre-trained on vast datasets. Despite the robust capabilities of these reference models in capturing the training distribution, the target distributions of downstream generative tasks rarely align perfectly with this reference distribution. For example, in conditional generative learning (Dhariwal and Nichol, 2021; Ho and Salimans, 2021), reference diffusion models generate samples from a mixture of distributions, whereas downstream tasks require sampling from specific constituent distributions. Similarly, in posterior sampling contexts (Chung et al., 2023; Song et al., 2023; Chen et al., 2025; Chang et al., 2025b; Purohit et al., 2025; Martin et al., 2025), the reference distribution serves as the prior, while the target is the posterior defined by tilting the prior with a measurement likelihood. Furthermore, downstream tasks frequently impose additional constraints, such as human preferences or safety considerations (Domingo-Enrich et al., 2024, 2025; Uehara et al., 2025b; Kim et al., 2025; Sabour et al., 2025; Denker et al., 2025; Ren et al., 2025), which must be satisfied without compromising the model's generative quality.

To bridge the gap between reference and target distributions, researchers have proposed numerous alignment methods (Xu et al., 2023; Lee et al., 2023b; Fan et al., 2023; Clark et al., 2024; Domingo-Enrich et al., 2024, 2025; Uehara et al., 2024b, 2025b). These strategies generally fall into two categories: fine-tuning and inference-time alignment. Fine-tuning approaches involve retraining the reference score network via supervised learning (Lee et al., 2023b), reinforcement learning (Fan et al., 2023; Black et al., 2024; Clark et al., 2024; Uehara et al., 2024c), or classifier-free fine-tuning (Ho and Salimans, 2021; Zhang et al., 2023; Yuan et al., 2023). Despite its conceptual simplicity, fine-tuning presents significant limitations. First, it often requires a substantial collection of high-quality samples from the target distribution, which may be unavailable in practical scenarios like posterior sampling. Second, the computational cost of retraining score networks can be prohibitive, particularly for large-scale models with billions of parameters (Uehara et al., 2025b). Third, fine-tuning is vulnerable to over-optimization (Gao et al., 2023; Rafailov et al., 2024; Kim et al., 2025), where the network overfits to limited target samples or preferences, causing it to "forget" the valuable prior information encoded in the reference model. This degradation undermines the fundamental advantage of leveraging pre-trained models.

In contrast, inference-time alignment (Uehara et al., 2025b; Kim et al., 2025; Sabour et al., 2025; Denker et al., 2025; Ren et al., 2025; Pachebat et al., 2025) eliminates the need to retrain the reference diffusion model. These methods offer substantial computational advantages and preserve the generative capacity of the underlying model. The core technique is guidance (Jiao et al., 2025), which incorporates target information by introducing an additional drift term to the reference diffusion model. Within the framework of Doob's h-transform (Rogers and Williams, 2000; Särkkä and Solin, 2019; Chewi, 2025), the score function for the target tilted distribution decomposes into the sum of the reference score and a guidance term, where the guidance is defined as the gradient of the log-Doob's h-function (Heng et al., 2024; Tang and Xu, 2024; Denker et al., 2024, 2025). This relationship has also been investigated through the lens of classifier guidance (Dhariwal and Nichol, 2021), stochastic optimal control (Han et al., 2024; Tang and Zhou, 2025), and Bayes' rule (Chung et al., 2023; Song et al., 2023). The primary challenge lies in accurately estimating this guidance.

Guidance estimation methods can be categorized into two main approaches: approximation and learning. The approximation approach, exemplified by diffusion posterior sampling (Chung et al., 2023) and loss-guided diffusion (Song et al., 2023), relies on heuristic approximations that often lead to inconsistencies with the underlying mathematical formulation. The learning approach (Dhariwal and Nichol, 2021; Tang and Xu, 2024; Denker et al., 2024, 2025) attempts to learn the necessary components by deep neural networks. Classifier guidance (Dhariwal and Nichol, 2021) learns Doob's h-function via a classifier, but this is effective primarily for discrete labels. Furthermore, the convergence of the plug-in gradient estimator of the classifier is not guaranteed, potentially undermining guidance reliability (Mou, 2025, Section 3.2.2). To mitigate this, Tang and Xu (2024) estimate Doob's h-function and its gradient using separate neural networks, increasing training complexity. While Denker et al. (2024) directly learn the guidance, their method requires samples from the target tilted distribution. Denker et al. (2025) attempt to address this data requirement via iterative retraining.

To address these limitations, we introduce variationally stable Doob's h-matching, a novel framework for provable guidance estimation in inference-time alignment. We propose a gradient-regularized regression method that simultaneously estimates Doob's h-function and its gradient, yielding a consistent estimator of the guidance. When combined with the pre-trained reference score, our method enables efficient sampling from the target distribution without the need for computationally expensive fine-tuning or access to target distribution samples.

1.1 Contributions. Our main contributions are summarized as follows:

(i) We introduce *variationally stable Doob's matching*, a novel guidance estimation framework for controllable diffusion models grounded in Doob's h-transform. The Doob h-function is estimated via a least-squares regression approach augmented with a gradient regularization, and the plug-in gradient of the logarithm of the resulting h-function estimator yields an estimator for the Doob's guidance. Additionally, this method is derivative-free, meaning it does not require access to the gradient of the weight function between the target tilted distribution and the reference distribution.

(ii) We establish non-asymptotic convergence rates for variationally stable Doob's matching, showing that the proposed method guarantees convergence of both the h-function estimator and its gradient. These results provide rigorous theoretical guarantees for Doob's guidance estimation (Theorem 5.3). Moreover, we derive non-asymptotic convergence rates for the induced controllable diffusion models, thereby establishing rigorous guarantees for the generated distributions in the 2-Wasserstein distance (Theorem 5.6). Additionally, we obtain convergence rates that depend only on the intrinsic dimension, thereby mitigating the curse of dimensionality under low-dimensional subspace assumptions (Theorem 5.9 and Corollary 5.10).

1.2 Organization. The remainder of this paper is organized as follows. In Section 2, we provide a brief introduction to diffusion models. In Section 3, we propose the stochastic dynamics of controllable diffusion models within the framework of Doob's h-transform, and we present a practical algorithm to simultaneously estimate Doob's h-function and its gradient. Section 5 establishes non-asymptotic error bounds for both the estimation of the h-function and the induced controllable diffusion models. Finally, concluding remarks are provided in Section 6. Detailed proofs of theoretical results are deferred to the appendix.

2 Preliminaries on Diffusion Models

2.1 Forward and time-reversal process. We consider the diffusion model for a reference distribution p_0. The forward process of the reference diffusion model is defined by the Ornstein–Uhlenbeck process:

$$\mathrm{d}\mathbf{X}_t = -\mathbf{X}_t \,\mathrm{d}t + \sqrt{2}\,\mathrm{d}\mathbf{B}_t, \quad t \in (0, T),\ \mathbf{X}_0 \sim p_0, \tag{2.1}$$

where $\mathbf{B}_t$ is a d-dimensional standard Brownian motion, and $T > 0$ is the terminal time. The transition distribution of the forward process can be expressed as:

$$(\mathbf{X}_t|\mathbf{X}_0 = \mathbf{x}_0) \sim \mathcal{N}(\mu_t \mathbf{x}_0, \sigma_t^2 \mathbf{I}_d), \tag{2.2}$$

where the mean and variance coefficients are given, respectively, as $\mu_t = \exp(-t)$ and $\sigma_t^2 = 1 - \exp(-2t)$. The forward process (2.1) is commonly referred to as the variance-preserving (VP) SDE (Song et al., 2021) as $\mu_t^2 + \sigma_t^2 = 1$ for each $t \in (0, T)$. Denote by p_t the marginal density of $\mathbf{X}_t$ for each $t \in (0, T)$, which satisfies

$$p_t(\mathbf{x}_t) = \int \varphi_d(\mathbf{x}_t; \mu_t \mathbf{x}_0, \sigma_t^2 \mathbf{I}_d) p_0(\mathbf{x}_0) \,\mathrm{d}\mathbf{x}_0,$$

where $\varphi_d(\cdot; \mu_t \mathbf{x}_0, \sigma_t^2 \mathbf{I}_d)$ denotes the density function of the Gaussian distribution $\mathcal{N}(\mu_t \mathbf{x}_0, \sigma_t^2 \mathbf{I}_d)$. The corresponding time-reversal process (Anderson, 1982) of (2.1) is characterized by:

$$\begin{aligned} \mathrm{d}\mathbf{X}_t^{\leftarrow} &= \Big(\mathbf{X}_t^{\leftarrow} + 2\overbrace{\nabla \log p_{T-t}(\mathbf{X}_t^{\leftarrow})}^{\text{base score}}\Big)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}\mathbf{B}_t, \quad t \in (0, T), \\ \mathbf{X}_0^{\leftarrow} &\sim p_T. \end{aligned} \tag{2.3}$$

It has been established that the path measure of the time-reversal process $(\mathbf{X}_t^{\leftarrow})_{0 \le t \le T}$ corresponds exactly to the reverse of the forward process $(\mathbf{X}_t)_{0 \le t \le T}$ (Anderson, 1982).

2.2 Path measure and filtration. We formally define the probability space for the time-reversal process (2.3). Let $\Omega := C([0, T], \mathbb{R}^d)$ be the space of continuous functions mapping $[0, T]$ to $\mathbb{R}^d$, equipped with the topology of uniform convergence. Let $\mathcal{F}$ be the Borel σ-algebra on Ω. We define the canonical process $\mathbf{X}^{\leftarrow}$ on Ω via the coordinate mapping $\mathbf{X}_t^{\leftarrow}(\omega) = \omega(t)$ for all $\omega \in \Omega$. The natural filtration is given by $\mathbb{F} := (\mathcal{F}_t)_{0 \le t \le T}$, where $\mathcal{F}_t := \sigma(\mathbf{X}_s^{\leftarrow} \mid 0 \le s \le t)$ is the σ-algebra generated by the path up to time t. We denote by $\mathbb{P}$ the probability measure on $(\Omega, \mathcal{F})$ induced by the law of the solution to the SDE (2.3) with initial distribution $\mathbf{X}_0^{\leftarrow} \sim p_T$. Consequently, the filtered probability space is denoted as $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, and $\mathbf{B}_t$ is an $\mathbb{F}$-Brownian motion under $\mathbb{P}$.

2.3 Training phase: score matching. In generative learning, the exact reference score $\nabla \log p_t$ in the time-reversal process (2.3) is intractable. One can estimate the reference score using samples from the reference density p_0 via standard techniques such as implicit score matching (Hyvärinen, 2005), sliced score matching (Song et al., 2020), and denoising score matching (Vincent, 2011). Let $\widehat{\mathbf{s}} : (0, T) \times \mathbb{R}^d \to \mathbb{R}^d$ denote an estimator for the prior score, that is,

$$\|\widehat{\mathbf{s}}(t, \cdot) - \nabla \log p_{T-t}\|_{L^2(p_{T-t})} \le \varepsilon, \tag{2.4}$$

for a small tolerence $\varepsilon \ll 1$. Considerable research has established theoretical guarantees for this score estimation (Tang and Yang, 2024; Oko et al., 2023; Fu et al., 2024; Ding et al., 2025a), leveraging standard techniques from non-parametric regression with deep neural networks (Bauer and Kohler, 2019; Schmidt-Hieber, 2020; Kohler and Langer, 2021; Jiao et al., 2023).

2.4 Inference phase: sampling. Given a reference score estimator $\widehat{\mathbf{s}}$ in (2.4), the inference phase of diffusion models aims to generate samples by simulating the time-reversal process with estimated score. Since the explicit solution of the time-reversal process is intractable, we employ an exponential integrator (Hochbruck and Ostermann, 2005, 2010; Lu et al., 2022a; Zhang and Chen, 2023). This approach is well-suited for solving the time-reversal process due to the semi-linearity of the drift term of the SDE in (2.3).

Let $K \in \mathbb{N}$ denote the number of discretization steps, and let $T_0 > 0$ be an early-stopping time. We define a sequence of uniform time points $t_k := kh$ for $k = 0, \dots, K$, where the step size is $h := (T - T_0)/K$. In each time sub-interval, the exponential integrator approximates the score function by its value at the left endpoint:

$$\begin{aligned} \mathrm{d}\widehat{\mathbf{X}}_t^{\leftarrow} &= (\widehat{\mathbf{X}}_t^{\leftarrow} + 2\widehat{\mathbf{s}}(kh, \widehat{\mathbf{X}}_{kh}^{\leftarrow}))\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}\mathbf{B}_t, \quad t \in [kh, (k+1)h), \\ \widehat{\mathbf{X}}_0^{\leftarrow} &\sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d), \end{aligned} \tag{2.5}$$

where $0 \le k \le K - 1$. The resulting linear approximation to the original semi-linear SDE has the following explicit solution:

$$\widehat{\mathbf{X}}_{(k+1)h}^{\leftarrow} = \exp(h)\widehat{\mathbf{X}}_{kh}^{\leftarrow} + 2\phi^2(h)\widehat{\mathbf{s}}(kh, \widehat{\mathbf{X}}_{kh}^{\leftarrow}) + \phi(2h)\boldsymbol{\xi}_k, \quad 0 \le k \le K - 1,$$

where $\phi(z) := \sqrt{\exp(z) - 1}$, and $\boldsymbol{\xi}_0, \dots, \boldsymbol{\xi}_{K-1}$ are i.i.d. standard Gaussian random variables.

Remark 2.1 (Initialization). Note that the true initial distribution of the time-reversal process (2.3) is p_T, rather than the $\mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ used in (2.5). We adopt the standard normal distribution because sampling from $\mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ is significantly more computationally tractable. This approximation is justified by the fact that p_T converges to $\mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ exponentially in KL-divergence as $T \to \infty$ (Bakry et al., 2014; Vempala and Wibisono, 2019); thus, the Gaussian initialization is valid for a sufficiently large terminal time T.

3 Controllable Diffusion Models and Doob's Transform

In this section, we propose controllable diffusion models for sampling from a target distribution, defined as the reference distribution tilted by a weight function. We utilize the theory of measure change for diffusion processes on the filtered space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$.

3.1 Problem setup. We assume access to a pre-trained reference diffusion model (2.5) that generates samples approximately from the reference distribution p_0. We aim to sample from a tilted distribution q_0, defined by reweighting the reference distribution with a known weight function $w : \mathbb{R}^d \to \mathbb{R}_{\ge 0}$:

$$q_0(\mathbf{x}) := \frac{w(\mathbf{x})p_0(\mathbf{x})}{Z}, \quad \text{where } Z := \int w(\mathbf{x})p_0(\mathbf{x})\,\mathrm{d}\mathbf{x} < \infty. \tag{3.1}$$

Our goal is to derive a new diffusion process that generates samples from the tilted distribution q_0 directly by introducing a drift correction to the pre-trained reference diffusion model (2.5).

This problem encompasses a wide range of application scenarios.

Example 1 (Bayesian inverse problems). Bayesian inverse problems play a critical role in scientific computing (Stuart, 2010; Kantas et al., 2014; Ding et al., 2024), image science (Chung et al., 2023; Mardani et al., 2024; Purohit et al., 2025; Chang et al., 2025b), and medical imaging (Song et al., 2022). In Bayesian inverse problems, we aim to recover an unknown signal $\mathbf{X}_0 \in \mathbb{R}^d$ from noisy measurements $\mathbf{Y} \in \mathbb{R}^m$, which are linked by the following measurement model:

$$\mathbf{Y} = \mathcal{A}(\mathbf{X}_0) + \mathbf{n}, \tag{3.2}$$

where $\mathcal{A} : \mathbb{R}^d \to \mathbb{R}^m$ is a known measurement operator, and $\mathbf{n} \in \mathbb{R}^m$ represents a measurement noise with a known distribution. The Bayesian approach incorporates prior knowledge about $\mathbf{X}_0$ in the form of a prior distribution p_0. Given observed measurements $\mathbf{Y} = \mathbf{y}$, the goal of Bayesian inverse problems is to sample from the posterior distribution:

$$q_0(\mathbf{x}) := p_{\mathbf{X}_0|\mathbf{Y}}(\mathbf{x}|\mathbf{y}) = \frac{w(\mathbf{x})p_0(\mathbf{x})}{Z}, \tag{3.3}$$

where $w(\mathbf{x}) \coloneqq p_{\mathbf{Y}|\mathbf{X}_0}(\mathbf{y}|\mathbf{x})$ is a likelihood determined by the measurement model (3.2), and Z is a partition function to ensure q_0 is a valid probability density. For example, for a Gaussian noise $\mathbf{n} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_d)$, it holds that

$$w(\mathbf{x}) = (2\pi\sigma^2)^{-\frac{d}{2}} \exp\Big(- \frac{1}{2\sigma^2} \|\mathbf{y} - \mathcal{A}(\mathbf{x})\|_2^2 \Big).$$

In Bayesian inverse problems, one typically has a reference diffusion model pre-trained on the prior distribution, and aims to sample from the posterior distribution (3.3) without retraining the reference model.

Example 2 (Reward-guided generation)*.* In the reward-guided generation (Domingo-Enrich et al., 2025; Uehara et al., 2025b; Kim et al., 2025; Sabour et al., 2025; Denker et al., 2025; Ren et al., 2025), human preferences and constraints can be encoded into a reward function $r : \mathbb{R}^d \to \mathbb{R}$. For instance, in text-to-image generation, the reward function r quantifies how well the generated data aligns with the input prompt. In practice, such reward function can be learned from the human feedback or preference data (Stiennon et al., 2020; Ouyang et al., 2022; Lee et al., 2023b). For the sake of simplicity, we assume throughout this work that the reward function has already been given. A naive approach to reward-guided generation is to maximize the expected reward $\max_{\pi \in \mathcal{P}} \mathbb{E}_\pi[r]$, where $\mathcal{P}$ is the set of probability measures on $\mathbb{R}^d$. However, solely maximizing the expected reward may lead to over-optimization and degenerate solutions (Kim et al., 2025). To mitigate this, entropy regularization (Uehara et al., 2024a; Tang and Zhou, 2025) is incorporated into the objective functional, yielding the following optimization problem:

$$q_0 \coloneqq \arg\max_{\pi \in \mathcal{P}} \mathbb{E}_\pi[r] - \alpha \mathrm{KL}(\pi \| p_0), \tag{3.4}$$

where $\alpha > 0$ is a regularization parameter, and p_0 is the density of a reference distribution, i.e., the distribution corresponding to the pre-trained reference diffusion model. This objective comprises two components: the expected reward, which captures human preferences, and the KL-divergence term, which prevents the distribution from deviating excessively from the reference diffusion model. The closed-form solution to this optimization problem (3.4) is given by (Rafailov et al., 2023):

$$q_0(\mathbf{x}) = \frac{w(\mathbf{x})p_0(\mathbf{x})}{Z}, \quad w(\mathbf{x}) := \exp\Big(\frac{r(\mathbf{x})}{\alpha}\Big),$$

where Z is the partition function to ensure q_0 is a valid probability density. In reward-guided generation, the central objective is to incorporate preferences exclusively during the inference phase, thereby avoiding the substantial computational cost of retraining the large-scale reference diffusion model.

Example 3 (Transfer learning for diffusion models)*.* Diffusion models have achieved remarkable success in image generation. However, their performance critically depends on the availability of large-scale training data. In the scenarios of few-shot generation, training diffusion models solely on limited samples typically results in poor generative performance. To retain strong generative capability under data scarcity, a common strategy is to transfer expressive diffusion models pre-trained on large datasets to the target domain (Ouyang et al., 2024; Wang et al., 2024; Zhong et al., 2025; Bahram et al., 2026). Formally, transfer learning aims to adapt a model pre-trained on a large-scale source distribution to a target distribution of much smaller size and diversity of samples. However, directly fine-tuning a large pre-trained diffusion model using only limited target samples often leads to severe overfitting (Wang et al., 2024). To address this issue, transfer learning approaches for diffusion models typically train a lightweight guidance network on the limited target data and combine it with the pre-trained reference score network, yielding a modified diffusion model capable of sampling from the target distribution (Ouyang et al., 2024; Zhong et al., 2025; Bahram et al., 2026). Concretely, Ouyang et al. (2024) estimates the density ratio between the target and source distributions,

$$w(\mathbf{x}) := \frac{q_0(\mathbf{x})}{p_0(\mathbf{x})},$$

using limited samples from the target dsitribution. For simplicity, we assume throughout this work that the density ratio is known. Under this assumption, the objective of transfer learning for diffusion models is to sample from the target distribution q_0 by leveraging the pre-trained reference diffusion model together with the estimated density ratio.

3.2 Controllable diffusion models with Doob's h-transform. In this subsection, we achieve inference-time alignment by incorporating guidance into the reference diffusion models via Doob's h-transform (Rogers and Williams, 2000; Särkkä and Solin, 2019; Heng et al., 2024; Chewi, 2025). We begin by constructing a target path measure $\mathbb{Q}$ on the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F})$. We assume that $\mathbb{Q}$ is absolutely continuous with respect to the reference path measure $\mathbb{P}$ defined in Section (2.2). By the Radon-Nikodym theorem, this relationship is characterized by the existence of an $\mathbb{F}$-adapted process $(L_t)_{0\le t\le T}$ with $L_t \ge 0$, such that for every $t \in [0, T]$:

$$L_t := \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\Big|_{\mathcal{F}_t}. \tag{3.5}$$

The process L_t is known as the Radon-Nikodym derivative process, representing the likelihood ratio between the target and reference measures conditioned on the filtration $\mathcal{F}_t$.

We impose two boundary conditions on the target path measure $\mathbb{Q}$: (i) the marginal distribution of the initial state $\mathbf{X}_0^{\leftarrow}$ under $\mathbb{Q}$ must match that of $\mathbb{P}$; and (ii) the marginal distribution of the terminal state $\mathbf{X}_T^{\leftarrow}$ under $\mathbb{Q}$ must coincide with the tilted distribution q_0 defined in (3.1). To enforce these constraints, we specify the Radon-Nikodym derivatives of $\mathbb{Q}$ with respect to $\mathbb{P}$ at the initial and terminal times as follows:

$$L_0 = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\Big|_{\mathcal{F}_0} \equiv 1, \quad \text{and} \quad L_T = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\Big|_{\mathcal{F}_T} = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = \frac{w(\mathbf{X}_T^{\leftarrow})}{\mathbb{E}^{\mathbb{P}}[w(\mathbf{X}_T^{\leftarrow})]}, \tag{3.6}$$

where the denominator serves as the normalizing constant ensuring that $\mathbb{Q}$ is a valid probability measure, i.e., $\mathbb{E}^{\mathbb{P}}[L_T] = 1$.

The rest of our derivation proceeds in two steps. First, we characterize the stochastic dynamics of $\mathbf{X}_t^{\leftarrow}$ under the target measure $\mathbb{Q}$. Second, by invoking the weak uniqueness of solutions to stochastic differential equations (Øksendal, 2003, Lemma 5.3.1), we construct a controllable diffusion process under the reference path measure $\mathbb{P}$ whose terminal distribution coincides with the target tilted distribution q_0.

Dynamics under the target path measure. Girsanov's theorem (Øksendal, 2003, Theorem 8.6.8) establishes a fundamental correspondence between a drift shift in the driving Brownian motion and the dynamics of the associated Radon-Nikodym derivative process. Accordingly, we first characterize the evolution of the time-dependent likelihood ratio L_t.

Proposition 3.1 (Doob's h-function)**.** *The Radon-Nikodym derivative process L_t, as defined in (3.5), admits the following representation:*

$$L_t = \mathbb{E}^{\mathbb{P}}\Big[\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\Big|\mathcal{F}_t\Big] = \frac{h^*(t, \mathbf{X}_t^{\leftarrow})}{\mathbb{E}^{\mathbb{P}}[w(\mathbf{X}_T^{\leftarrow})]},$$

where $h^* : [0, T] \times \mathbb{R}^d \to \mathbb{R}$, referred to as Doob's h-function, is defined as the conditional expectation of the terminal weight:

$$h^*(t, \mathbf{x}) := \mathbb{E}^{\mathbb{P}}[w(\mathbf{X}_T^{\leftarrow}) \mid \mathbf{X}_t^{\leftarrow} = \mathbf{x}]. \tag{3.7}$$

Furthermore, the log-likelihood ratio satisfies the following SDE:

$$\mathrm{d}(\log L_t) = \nabla \log h^*(t, \mathbf{X}_t^{\leftarrow})^\top \sqrt{2}\, \mathrm{d}\mathbf{B}_t - \|\nabla \log h^*(t, \mathbf{X}_t^{\leftarrow})\|_2^2\, \mathrm{d}t.$$

The proof of Proposition 3.1 is provided in Appendix A. With the dynamics of $\log L_t$ established, we derive the stochastic dynamics of $\mathbf{X}_t^{\leftarrow}$ under the target measure $\mathbb{Q}$ via Girsanov's theorem (Øksendal, 2003, Theorem 8.6.6).

Proposition 3.2. *Let the reference process $\mathbf{X}_t^{\leftarrow}$ satisfy the SDE (2.3) under the path measure $\mathbb{P}$, and let $\mathbb{Q}$ be the target path measure defined by (3.6). Assume that the Novikov condition holds:*

$$\mathbb{E}^{\mathbb{P}}\Big[\exp\Big(\int_0^T \|\nabla \log h^*(s, \mathbf{X}_s^{\leftarrow})\|_2^2\, \mathrm{d}s\Big)\Big] < \infty. \tag{3.8}$$

Define a process $(\widetilde{\mathbf{B}}_t)_{1 \le t \le T}$ by

$$\mathrm{d}\widetilde{\mathbf{B}}_t = \mathrm{d}\mathbf{B}_t - \sqrt{2}\nabla \log h^*(t, \mathbf{X}_t^{\leftarrow})\, \mathrm{d}t, \tag{3.9}$$

where $\mathbf{B}_t$ is a Brownian motion under $\mathbb{P}$. Then $\widetilde{\mathbf{B}}_t$ is a standard Brownian motion under $\mathbb{Q}$. Further, under the path measure $\mathbb{Q}$, the reference process $\mathbf{X}_t^{\leftarrow}$ in (2.3) evolves according to:

$$\mathrm{d}\mathbf{X}_t^{\leftarrow} = (\mathbf{X}_t^{\leftarrow} + 2\nabla \log p_{T-t}(\mathbf{X}_t^{\leftarrow}) + 2\nabla \log h^*(t, \mathbf{X}_t^{\leftarrow}))\, \mathrm{d}t + \sqrt{2}\, \mathrm{d}\widetilde{\mathbf{B}}_t. \tag{3.10}$$

The proof is deferred to Appendix A. By the construction in (3.6), the law of $\mathbf{X}_T^{\leftarrow}$ under $\mathbb{Q}$ coincides with the target tilted distribution q_0.

Remark 3.3 (Novikov condition)*.* The condition in (3.8) ensures that the exponential local martingale defined by the drift shift is a true martingale (Karatzas and Shreve, 1998, Corollary 5.13), which is sufficient for the Radon-Nikodym derivative to be well-defined and for Girsanov's theorem to apply.

Controllable diffusion process under the reference measure. Although Proposition 3.2 constructs stochastic dynamics driven by Brownian motion under $\mathbb{Q}$ that achieve the desired terminal distribution q_0, practical implementation necessitates an SDE driven by Brownian motion under the reference measure $\mathbb{P}$.

To address this, we construct a surrogate process $\mathbf{Z}_t^{\leftarrow}$ on the reference probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ that adopts the drift derived for $\mathbb{Q}$:

$$(3.11) \qquad \mathrm{d}\mathbf{Z}_t^{\leftarrow} = \Big(\mathbf{Z}_t^{\leftarrow} + 2\underbrace{\nabla \log p_{T-t}(\mathbf{Z}_t^{\leftarrow})}_{\text{base score}} + 2\underbrace{\nabla \log h^*(t, \mathbf{Z}_t^{\leftarrow})}_{\text{Doob's guidance}}\Big)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}\mathbf{B}_t.$$

Since the process $\mathbf{Z}_t^{\leftarrow}$ driven by the $\mathbb{P}$-Brownian motion (3.11) satisfies the same SDE as $\mathbf{X}_t^{\leftarrow}$ driven by the $\mathbb{Q}$-Brownian motion (3.10), the weak uniqueness property of SDE solutions guarantees (Øksendal, 2003, Lemma 5.3.1) that the law of $\mathbf{Z}_t^{\leftarrow}$ under $\mathbb{P}$ is identical to the law of $\mathbf{X}_t^{\leftarrow}$ under $\mathbb{Q}$ for all $t \in [0, T]$. Thus, the law of $\mathbf{Z}_T^{\leftarrow}$ under $\mathbb{P}$ coincides the target tilted distribution q_0. For a detailed formal statement, see Øksendal (2003, Theorem 8.6.8).

4 Variationally Stable Doob's Matching

We have thus far established a controllable diffusion process (3.11) capable of generating samples from the target tilted distribution q_0. However, the Doob's h-guidance, required by (3.11), remains intractable. This subsection proposes a variationally stable Doob's matching method to address the estimation of the Doob's guidance.

4.1 Vanilla least-squares regression for Doob's matching. For any $t \in (0, T)$, the Doob's h-function $h_t^* := h^*(t, \cdot)$ defined as (3.7) is the unique minimizer of the following implicit Doob's matching objective:

$$(4.1) \qquad \begin{aligned} \mathcal{J}_t(h_t) &= \mathbb{E}^{\mathbb{P}}\big[\|h_t(\mathbf{X}_t^{\leftarrow}) - w(\mathbf{X}_T^{\leftarrow})\|_2^2\big] \\ &= \mathbb{E}_{\mathbf{X}_0 \sim p_0}\mathbb{E}_{\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)}\big[\|h_t(\mu_{T-t}\mathbf{X}_0 + \sigma_{T-t}\boldsymbol{\epsilon}) - w(\mathbf{X}_0)\|_2^2\big], \end{aligned}$$

where $w : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ is the known weight function in (3.1). The following proposition justifies the use of $\mathcal{J}_t$ as a surrogate for the explicit L^2-distance.

Proposition 4.1. *For every $t \in (0, T)$, the Doob's h-function h_t^* in (3.7) minimizes the implicit Doob's matching objective (4.1). Further,*

$$\mathcal{J}_t(h_t) = \mathbb{E}^{\mathbb{P}}\big[\|h_t(\mathbf{X}_t^{\leftarrow}) - h_t^*(\mathbf{X}_t^{\leftarrow})\|_2^2\big] + V_t^2,$$

where $V_t^2 := \mathbb{E}^{\mathbb{P}}[\mathrm{Var}(w(\mathbf{X}_T^{\leftarrow})|\mathbf{X}_t^{\leftarrow})]$ is a constant independent of h_t.

The proof of Proposition 4.1 is provided in Appendix B.

4.2 Limitations of vanilla regression. Crucially, computing the Doob's guidance $\nabla \log h_t^*$ in (3.11) requires estimating not only the function h_t^* itself but also its gradient ∇h_t^*, as the guidance is given by:

$$\nabla \log h_t^*(\mathbf{x}) = \frac{\nabla h_t^*(\mathbf{x})}{h_t^*(\mathbf{x})}, \quad \mathbf{x} \in \mathbb{R}^d.$$

According to Proposition 4.1, the objective $\mathcal{J}_t$ is only coercive with respect to the $L^2(p_{T-t})$-norm. Specifically, for any $h_t \in L^2(p_{T-t})$,

$$(4.2) \qquad \mathcal{J}_t(h_t) - \mathcal{J}_t(h_t^*) = \|h_t - h_t^*\|_{L^2(p_{T-t})}^2.$$

However, even if h_t is close to h_t^* in the L^2 sense, the gradient ∇h_t may remain highly oscillatory. This leads to an unstable plug-in estimator for the Doob's guidance, a difficulty noted in prior works such as Tang and Xu (2024, Section 3.2.1) and Mou (2025, Section 3.2.2). Similar issues arise in related contexts, including classifier guidance (Dhariwal and Nichol, 2021), Monte Carlo regression (Uehara et al., 2025b, Section 2.2), and Ouyang et al. (2024).

To illustrate this fundamental limitation, consider the sequence of functions $f_n : \mathcal{I} \to \mathbb{R}$ defined by $x \mapsto n^{-1}\sin(nx)$, alongside the zero function $f_0 \equiv 0$, where $\mathcal{I} := [0, 2\pi]$. While $\|f_n - f_0\|_{L^2(\mathcal{I})} \to 0$ as $n \to \infty$, the distance between their derivatives does not vanish, i.e., $\lim_{n\to\infty} \|f_n' - f_0'\|_{L^2(\mathcal{I})} \neq 0$. As a result, the convergence of the plug-in Doob's guidance estimator derived from vanilla regression (4.1), as utilized in Uehara et al. (2025b); Ouyang et al. (2024), cannot be guaranteed in general.

To mitigate this, Tang and Xu (2024) estimate Doob's h-function and its gradient separately via a martingale approach. In contrast, in the remainder of this work, we propose an approach to simultaneously estimate both the function and its gradient.

4.3 Variationally stable Doob's matching. To simultaneously estimate the Doob's h-function and its gradient, we adopt a gradient-regularized regression (Drucker and Le Cun, 1991, 1992; Ding et al., 2025b). The population risk is defined by incorporating an additional Sobolev regularization to (4.1):

$$
\begin{aligned}
(4.3) \qquad h_t^\lambda = \operatorname*{arg\,min}_{h_t:\mathbb{R}^d\to\mathbb{R}^d} \mathcal{J}_t^\lambda(h_t) &:= \mathcal{J}_t(h_t) + \overbrace{\lambda \mathbb{E}^{\mathbb{P}}\left[\|\nabla h_t(\mathbf{X}_t^{\leftarrow})\|_2^2\right]}^{\text{gradient regularization}} \\
&= \mathcal{J}_t(h_t) + \lambda \mathbb{E}_{\mathbf{X}_0\sim p_0}\mathbb{E}_{\boldsymbol{\varepsilon}\sim\mathcal{N}(\mathbf{0},\mathbf{I}_d)}\left[\|\nabla h_t(\mu_{T-t}\mathbf{X}_0 + \sigma_{T-t}\boldsymbol{\varepsilon})\|_2^2\right],
\end{aligned}
$$

where $\lambda > 0$ is a regularization parameter. The following results characterize the regularization gap and the variational stability of this formulation.

Proposition 4.2 (Regularization gap)**.** *Let $\lambda > 0$, h_t^* be the Doob's h-function defined as (3.7), and h_t^λ be the minimizer of $\mathcal{J}_t^\lambda$ defined as (4.3). Then $h_t^* \in H^2(p_{T-t})$, and*

$$
\begin{aligned}
\|h_t^\lambda - h_t^*\|_{L^2(p_{T-t})}^2 &\le \lambda^2 \|\Delta h_t^* + \nabla h_t^* \cdot \nabla \log p_{T-t}\|_{L^2(p_{T-t})}^2, \\
\|\nabla h_t^\lambda - \nabla h_t^*\|_{L^2(p_{T-t})}^2 &\le \lambda \|\Delta h_t^* + \nabla h_t^* \cdot \nabla \log p_{T-t}\|_{L^2(p_{T-t})}^2.
\end{aligned}
$$

The proof of Proposition 4.2 is provided in Appendix B. This proposition demonstrates that as $\lambda \to 0$, the minimizer h_t^λ of the regularized objective (4.3) converges to h_t^* in H^1-norm. More importantly, the objective is variationally stable in the H^1 sense:

Proposition 4.3 (Variational stability)**.** *Let $\lambda > 0$, and h_t^λ be the minimizer of $\mathcal{J}_t^\lambda$ defined as (4.3). Then for any $h_t \in H^1(p_{T-t})$, we have*

$$
\frac{1}{\max\{\lambda, 1\}}\{\mathcal{J}_t^\lambda(h_t) - \mathcal{J}_t^\lambda(h_t^\lambda)\} \le \|h_t - h_t^\lambda\|_{H^1(p_{T-t})}^2 \le \frac{1}{\min\{\lambda, 1\}}\{\mathcal{J}_t^\lambda(h_t) - \mathcal{J}_t^\lambda(h_t^\lambda)\}.
$$

The proof of Proposition 4.3 is provided in Appendix B. We refer to the regularized objective $\mathcal{J}_t^\lambda$ as variationally stable because the convergence in this objective functional necessitates simultaneous convergence in both the function values and their gradients, i.e., stability in the H^1 sense. Such variational stability ensures that any candidate function h_t achieving a low objective value $\mathcal{J}_t^\lambda(h_t)$ is guaranteed to be an approximation of the ground-truth Doob's h-function in both value and gradient.

Comparison between vanilla and gradient-regularized Doob's matching. The distinction between vanilla Doob's matching (4.1) and the proposed gradient-regularized Doob's matching (4.3) is fundamental for stable diffusion guidance estimation. While vanilla regression guarantees convergence in the L^2-norm; see (4.2), it can be unstable in the H^1 sense. In contrast, the proposed objective $\mathcal{J}_t^\lambda$ is coercive with respect to the H^1 norm; that is, for a fixed $\lambda > 0$,

$$\mathcal{J}_t^\lambda(h_t) - \mathcal{J}_t^\lambda(h_t^\lambda) \simeq \|h_t - h_t^\lambda\|_{H^1(p_{T-t})}^2.$$

This property ensures the simultaneous estimation of Doob's h-function and its gradient, thereby yielding a plug-in estimator for Doob's guidance with mathematical guarantees.

4.4 Doob's guidance estimation. Since the expectation in the population risk (4.3) is computationally intractable, we approximate it by empirical risk using independent and identically distributed samples:

$$\widehat{\mathcal{J}}_t^\lambda(h_t) \coloneqq \widehat{\mathcal{J}}_t(h_t) + \frac{\lambda}{n}\sum_{i=1}^n \|\nabla h_t(\mu_{T-t}\mathbf{X}_0^i + \sigma_{T-t}\boldsymbol{\varepsilon}^i)\|_2^2, \tag{4.4}$$

where the empirical least-squares risk is defined as

$$\widehat{\mathcal{J}}_t(h_t) \coloneqq \frac{1}{n}\sum_{i=1}^n \|h_t(\mu_{T-t}\mathbf{X}_0^i + \sigma_{T-t}\boldsymbol{\varepsilon}^i) - w(\mathbf{X}_0^i)\|_2^2. \tag{4.5}$$

Here $\mathbf{X}_0^1, \ldots, \mathbf{X}_0^n$ are independent and identically distributed random variables drawn from the reference distribution p_0, and $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n$ are independent standard Gaussian random variables. Then one has a gradient-regularized empirical risk minimizer:

$$\widehat{h}_t^\lambda \in \operatorname*{arg\,min}_{h_t \in \mathscr{H}_t} \widehat{\mathcal{J}}_t^\lambda(h_t), \tag{4.6}$$

where $\mathscr{H}_t$ is a hypothesis class, which is chosen as a neural network class in this work.

The Doob's matching with gradient regularization (4.6) yields a valid plug-in estimator of the Doob's guidance:

$$\widehat{\mathbf{g}}_t^\lambda(\mathbf{z}) \coloneqq \nabla \log \widehat{h}_t^\lambda(\mathbf{z}) = \frac{\nabla \widehat{h}_t^\lambda(\mathbf{z})}{\widehat{h}_t^\lambda(\mathbf{z})} \approx \frac{\nabla h_t^*(\mathbf{z})}{h_t^*(\mathbf{z})} = \nabla \log h_t^*(\mathbf{z}), \quad \mathbf{z} \in \mathbb{R}^d. \tag{4.7}$$

4.5 A summary of computing procedure. By a similar argument as Section 2.4, we have the exponential integrator for the controllable diffusion model:

$$\begin{aligned} &\mathrm{d}\widehat{\mathbf{Z}}_t^{\leftarrow} = (\widehat{\mathbf{Z}}_t^{\leftarrow} + 2\widehat{\mathbf{s}}(kh, \widehat{\mathbf{Z}}_{kh}^{\leftarrow}) + 2\widehat{\mathbf{g}}^\lambda(kh, \widehat{\mathbf{Z}}_{kh}^{\leftarrow}))\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}\mathbf{B}_t, \quad t \in (kh, (k+1)h), \\ &\widehat{\mathbf{Z}}_0^{\leftarrow} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d), \end{aligned} \tag{4.8}$$

where $0 \leq k \leq K-1$, the pre-trained reference score estimator $\widehat{\mathbf{s}}$ is defined as (2.4), and the Doob's guidance estimator $\widehat{\mathbf{g}}^\lambda$ is defined as (4.7).

We apply post-processing to the generated particle $\widehat{\mathbf{Z}}_{T-T_0}^{\leftarrow}$ to ensure numerical stability and facilitate the theoretical analysis presented in Theorem 5.6 (Lee et al., 2023a; Chen et al., 2023a). First, we assume the target distribution q_0 is concentrated on a domain centered at the origin, such as a distribution with compact support (Assumption 1) or with light tails. Consequently, we introduce a truncation operator to the particles obtained from the controllable diffusion model (4.8). Second, because the controllable diffusion process is terminated at an early-stopping time T_0, there exists a mean shift between the target distribution q_0 and the early-stopping distribution $q_{T_0} \approx \widehat{q}_{T-T_0}$, as indicated by (2.2). To mitigate this drift, we employ a scaling operator. Specifically, for $R > 0$, we define a truncation operator $\mathcal{T}_R : \mathbf{z} \mapsto \mathbf{z}\mathbb{1}_{B(\mathbf{0},R)}(\mathbf{z})$ and a scaling operator $\mathcal{M} : \mathbf{z} \mapsto \mu_{T_0}^{-1}\mathbf{z}$. The final processed particle is defined as

$$\mathcal{M} \circ \mathcal{T}_R(\widehat{\mathbf{Z}}^{\leftarrow}_{T-T_0}) = \mu_{T_0}^{-1}\widehat{\mathbf{Z}}^{\leftarrow}_{T-T_0}\mathbb{1}_{B(\mathbf{0},R)}(\widehat{\mathbf{Z}}^{\leftarrow}_{T-T_0}), \tag{4.9}$$

and we denote its density by $(\mathcal{M} \circ \mathcal{T}_R)_\sharp \widehat{q}_{T-T_0}$.

A complete procedure is summarized in Algorithm 1.

Algorithm 1: Inference-time alignment via variationally stable Doob's matching

Input: Reference score estimator $\widehat{\mathbf{s}}$, the weight function w, the regularization parameter λ, the step size h, and the number of steps K.

Output: Particle $\mathcal{M} \circ \mathcal{T}_R(\widehat{\mathbf{Z}}^{\leftarrow}_{T-T_0})$ follows the tilted distribution q_0 approximately.

```
# Doob's matching
```
Estimate Doob's h-function by $\widehat{h}_t^\lambda$ via gradient regularized Doob's matching (4.6).
```
# Controllable generation
```
Generate the initial particle $\widehat{\mathbf{Z}}^{\leftarrow}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$.

for $k = 0, \ldots, K-1$ **do**

  Evaluate the reference score: $\widehat{\mathbf{s}}_k \leftarrow \widehat{\mathbf{s}}(kh, \widehat{\mathbf{Z}}^{\leftarrow}_{kh})$.

  Evaluate the Doob's guidance: $\widehat{\mathbf{g}}_k \leftarrow \nabla \log \widehat{h}^\lambda(kh, \widehat{\mathbf{Z}}^{\leftarrow}_{kh})$.

  Exponential integrator: $\widehat{\mathbf{Z}}^{\leftarrow}_{(k+1)h} \sim \mathcal{N}(\exp(h)\widehat{\mathbf{Z}}^{\leftarrow}_{kh} + 2\phi^2(h)(\widehat{\mathbf{s}}_k + \widehat{\mathbf{g}}_k), \phi^2(2h)\mathbf{I}_d)$.

end
```
# Truncation and scaling
```
$\mathcal{M} \circ \mathcal{T}_R(\widehat{\mathbf{Z}}^{\leftarrow}_{T-T_0}) \leftarrow \mu_{T_0}^{-1}\widehat{\mathbf{Z}}^{\leftarrow}_{T-T_0}\mathbb{1}_{B(\mathbf{0},R)}(\widehat{\mathbf{Z}}^{\leftarrow}_{T-T_0})$.

return $\mathcal{M} \circ \mathcal{T}_R(\widehat{\mathbf{Z}}^{\leftarrow}_{T-T_0})$

---

5 Convergence Analysis

In this section, we derive a non-asymptotic convergence rate for the variationally stable Doob's matching (4.6) and the induced controllable diffusion model (4.8). Furthermore, we demonstrate that this convergence rate mitigates the curse of dimensionality under mild assumptions.

5.1 Assumptions. We begin by outlining the essential technical assumptions required for our theoretical results.

Assumption 1 (Bounded support)**.** The support of the target distribution q_0 is a compact set contained within the hypercube $\{\mathbf{x}_0 \in \mathbb{R}^d : \|\mathbf{x}_0\|_\infty \leq 1\}$.

Assumption 1 is a standard condition imposed on the data distribution (Lee et al., 2023a; Oko et al., 2023; Chang et al., 2025a; Beyler and Bach, 2025). This constraint is well-motivated by practical applications; for instance, image and video data consist of bounded pixel values, thereby satisfying this requirement.

Assumption 2 (Bounded weight function)**.** The weight function w defined in (3.1) is bounded from above and bounded away from zero. Specifically, there exist constants $0 < \underline{B} < 1 < \overline{B} < \infty$ such that

$$\underline{B} \leq w(\mathbf{x}) \leq \bar{B}, \quad \text{for all } \mathbf{x} \in \operatorname{supp}(q_0).$$

Assumption 2 implies that the reference distribution p_0 and the tilted distribution q_0 satisfy mutual absolute continuity, ensuring that their supports coincide (i.e., $\mathrm{supp}(p_0) = \mathrm{supp}(q_0)$). This condition is crucial for establishing the regularity of Doob's h-function. Furthermore, the ratio $\kappa := \bar{B}/\underline{B}$ serves as a condition number that characterizes the difficulty of the controllable diffusion task, as discussed in the context of posterior sampling by Purohit et al. (2025); Ding et al. (2024); Chang et al. (2025b).

Under Assumptions 1 and 2, we establish the regularity properties of Doob's h-function defined in (3.7).

Proposition 5.1. *Suppose Assumptions 1 and 2 hold. Then for all $t \in (0, T)$ and $\mathbf{x} \in \mathbb{R}^d$, the following bounds hold:*

(i) $\underline{B} \leq h_t^*(\mathbf{x}) \leq \bar{B}$*;*

(ii) $\max_{1 \leq k \leq d} |D_k h_t^*(\mathbf{x})| \leq 2\sigma_{T-t}^{-2} \bar{B}$*; and*

(iii) $\max_{1 \leq k, \ell \leq d} |D_{k\ell}^2 h_t^*(\mathbf{x})| \leq 6\sigma_{T-t}^{-4} \bar{B}$*,*

where D_k and $D_{k\ell}^2$ denote the first-order and second-order partial derivatives with respect to the input coordinates, respectively.

The proof of Proposition 5.1 is deferred to Appendix C. It is worth noting that Proposition 5.1 relies solely on the boundedness of the weight function w, without requiring the existence or smoothness of its gradients. Nevertheless, we establish that Doob's h-function admits bounded derivatives. This result stems from the definition of Doob's h-function as a posterior expectation under a Gaussian likelihood; the inherent smoothness of the Gaussian kernel endows the posterior expectation with strong regularity properties.

Assumption 3 (Reference score estimation error)**.** The reference score estimator $\widehat{\mathbf{s}}$ defined in (2.4) satisfies the following error bound:

$$\frac{1}{T} \sum_{k=0}^{K-1} h \mathbb{E}^{\mathbb{P}} \Big[\| \widehat{\mathbf{s}}(kh, \mathbf{X}_{kh}^{\leftarrow}) - \nabla \log p_{T-kh}(\mathbf{X}_{kh}^{\leftarrow}) \|_2^2 \Big] \leq \varepsilon_{\mathrm{ref}}^2.$$

Assumption 3 requires the L^2-error of the reference score estimator $\widehat{\mathbf{s}}$ to be bounded with respect to the reference path measure $\mathbb{P}$. In our setting, where numerous samples from the reference distribution p_0 are available, estimators satisfying this bound can be obtained via implicit score matching (Hyvärinen, 2005), sliced score matching (Song et al., 2020), or denoising score matching (Vincent, 2011). While one can derive explicit bounds of reference score matching as Tang and Yang (2024); Oko et al. (2023); Fu et al. (2024); Ding et al. (2025a); Yakovlev and Puchkin (2025a); Yakovlev et al. (2025) using non-parametric regression theory for deep neural networks (Bauer and Kohler, 2019; Schmidt-Hieber, 2020; Kohler and Langer, 2021; Jiao et al., 2023), we adopt this condition to maintain clarity of presentation, following the convention of Lee et al. (2023a); Chen et al. (2023a); Beyler and Bach (2025); Kremling et al. (2025).

5.2 Error bounds for the Doob's guidance estimator. We begin by introducing the concept of Vapnik-Chervonenkis (VC) dimension (Vapnik and Chervonenkis, 1971; Anthony et al., 1999; Bartlett et al., 2019), which measures the complexity of a function class.

Definition 1 (VC-dimension)**.** Let $\mathscr{H}$ be a class of functions mapping from $\mathcal{X}$ to $\mathbb{R}$. For any num-negative integer m, the growth function of $\mathscr{H}$ is defined as

$$\Pi_{\mathscr{H}}(m) := \max_{x_1, \ldots, x_m \in \mathcal{X}} |\{(\mathrm{sgn}\, h(x_1), \ldots, \mathrm{sgn}\, h(x_m)) : h \in \mathscr{H}\}|.$$

We say $\mathscr{H}$ shatters the set $\{x_1, \dots, x_m\}$, if

$$|\{(\operatorname{sgn} h(x_1), \dots, \operatorname{sgn} h(x_m)) : h \in \mathscr{H}\}| = 2^m.$$

The Vapnik-Chervonenkis dimension of $\mathscr{H}$, denoted by $\mathrm{VCdim}(\mathscr{H})$, is the size of the largest shattered set, i.e., the largest m such that $\Pi_{\mathscr{H}}(m) = 2^m$.

To simplify notation, we define the gradient classes and their associated VC-dimensions. For a differentiable hypothesis class $\mathscr{H}$ consisting of functions mapping from $\mathbb{R}^d$ to $\mathbb{R}$, the VC-dimension of the gradient hypothesis class is defined as

$$\mathrm{VCdim}(\nabla\mathscr{H}) := \max_{1\le k\le d} \mathrm{VCdim}(D_k\mathscr{H}), \quad D_k\mathscr{H} := \{D_k h : h \in \mathscr{H}\},$$

where D_k represents the derivative with respect to the k-th entry of the input.

The following lemma provides an oracle inequality for the variationally stable Doob's matching (4.6).

Lemma 5.2 (Oracle inequality)**.** *Suppose Assumptions 1 and 2 hold. Let $t \in (0, T)$ and let $\mathscr{H}_t$ be a hypothesis class. Let $\widehat{h}_t^\lambda$ be the gradient-regularized empirical risk minimizer defined as (4.6), and let h_t^* be the Doob's h-function defined as (3.7). Then the following inequalities hold:*

$$\begin{aligned}\mathbb{E}\Big[\|\widehat{h}_t^\lambda - h_t^*\|_{L^2(p_{T-t})}^2\Big] \lesssim\ & \underbrace{\inf_{h_t\in\mathscr{H}_t}\Big\{\|h_t - h_t^*\|_{L^2(p_{T-t})}^2 + \lambda\|\nabla h_t - \nabla h_t^*\|_{L^2(p_{T-t})}^2\Big\}}_{\text{(I)}} \\ & + \underbrace{\bar{B}^2\Big(\frac{\mathrm{VCdim}(\mathscr{H}_t)}{n\log^{-1} n}\Big)^{\frac12} + \frac{\lambda d\bar{B}^2}{\sigma_{T-t}^4}\Big(\frac{\mathrm{VCdim}(\nabla\mathscr{H}_t)}{n\log^{-1} n}\Big)^{\frac12}}_{\text{(II)}} + \underbrace{\frac{\lambda^2 d\bar{B}^2}{\sigma_{T-t}^8}}_{\text{(III)}},\\ \mathbb{E}\Big[\|\nabla\widehat{h}_t^\lambda - \nabla h_t^*\|_{L^2(p_{T-t})}^2\Big] \lesssim\ & \underbrace{\inf_{h_t\in\mathscr{H}_t}\Big\{\frac{1}{\lambda}\|h_t - h_t^*\|_{L^2(p_{T-t})}^2 + \|\nabla h_t - \nabla h_t^*\|_{L^2(p_{T-t})}^2\Big\}}_{\text{(I))}} \\ & + \underbrace{\frac{\bar{B}^2}{\lambda}\Big(\frac{\mathrm{VCdim}(\mathscr{H}_t)}{n\log^{-1} n}\Big)^{\frac12} + \frac{d\bar{B}^2}{\sigma_{T-t}^4}\Big(\frac{\mathrm{VCdim}(\nabla\mathscr{H}_t)}{n\log^{-1} n}\Big)^{\frac12}}_{\text{(II)}} + \underbrace{\frac{\lambda d\bar{B}^2}{\sigma_{T-t}^8}}_{\text{(III)}},\end{aligned}$$

where the notation $\lesssim$ hides absolute constants.

The proof of Lemma 5.2 is deferred to Appendix D. Both oracle inequalities for Doob's h-function and its gradient decompose the error into three components: approximation error, generalization error, and regularization gap.

(I) The **approximation error** is defined as the minimal H^1-distance between functions in the hypothesis class $\mathscr{H}_t$ and the ground-truth Doob's h-function h_t^*, measuring the approximation capability of $\mathscr{H}_t$.

(II) The **generalization error** captures the error arising from finite-sample approximation, which vanishes as the number of samples approaches infinity.

(III) The **regularization gap** is introduced by the gradient regularization in the objective functional, which causes the minimizer of the variationally stable objective (4.3) to deviate from the ground-truth Doob's h-function h_t^* (3.7). This gap has been analyzed in Proposition 4.2.

Comparison with oracle inequality of vanilla regression. Lemma 5.2 is analogous to the oracle inequality found in regression problems. Let $\widehat{h}_t$ be the vanilla estimator estimated by minimizing (4.5) over the hypothesis class $\mathscr{H}_t$. Informally, the following oracle inequality holds:

$$\mathbb{E}\Big[\|\widehat{h}_t - h_t^*\|^2_{L^2(p_{T-t})}\Big] \lesssim \underbrace{\inf_{h_t \in \mathscr{H}_t} \|h_t - h_t^*\|^2_{L^2(p_{T-t})}}_{\text{approximation}} + \underbrace{\bar{B}^2\Big(\frac{\mathrm{VCdim}(\mathscr{H}_t)}{n\log^{-1} n}\Big)^{\frac{1}{2}}}_{\text{generalization}}. \tag{5.1}$$

Comparing (5.1) with Lemma 5.2 reveals several crucial differences:

(i) The approximation error in Lemma 5.2 is measured in the H^1-norm, whereas in (5.1), it is measured in the L^2-norm. This distinction is natural because we require the estimator to converge in the H^1-norm; thus, simultaneous approximation of the function and its derivatives is essential. Simultaneous approximation using neural networks has been investigated in various contexts (Li et al., 2019; Gühring et al., 2020; Gühring and Raslan, 2021; Duan et al., 2022a,b; Lu et al., 2022b; Shen et al., 2022, 2024; Belomestny et al., 2023; Yakovlev and Puchkin, 2025b).

(ii) In vanilla regression, the generalization error in (5.1) depends only on the complexity of the hypothesis class. In contrast, the generalization error bounds in Lemma 5.2 also depend on the complexity of the derivative classes $\nabla\mathscr{H}_t$. This occurs because the objective functional of the gradient-regularized Doob's matching (4.3) includes the gradient norm term. Consequently, the error from finite-sample approximation is influenced not only by the complexity of the hypothesis class but also by that of the derivative class.

(iii) The most significant difference lies in the regularization error. If we focus solely on the oracle inequality for $\widehat{h}_t^\lambda$, letting λ go to zero reduces the expression to the vanilla regression oracle inequality (5.1). However, the bound for the gradient $\nabla\widehat{h}_t^\lambda$ diverges as the regularization parameter λ approaches zero. This highlights the key advantage of our gradient-regularized method: the gradient-regularized is essential for guaranteeing simultaneous convergence of both the estimator value and its gradient. Additionally, there exists a trade-off with respect to λ in the oracle inequality for $\widehat{h}_t^\lambda$: a larger λ leads to larger regularization error, while reduces the approximation and generalization errors.

Given the oracle inequality for a general hypothesis class $\mathscr{H}_t$, we consider the specifical case that $\mathscr{H}_t$ is chosen as a neural network class, with the aim of deriving non-asymptotic convergence rates. We begin by formally defining the neural network class.

Definition 2 (Neural network class)**.** A function implemented by a neural network $h : \mathbb{R}^{N_0} \to \mathbb{R}^{N_{L+1}}$ is defined by

$$h(\mathbf{x}) = T_L(\varrho(T_{L-1}(\cdots\varrho(T_0(\mathbf{x}))\cdots))),$$

where the activation function ϱ is applied component-wise and $T_\ell(\mathbf{x}) := \mathbf{A}_\ell\mathbf{x} + \mathbf{b}_\ell$ is an affine transformation with $\mathbf{A}_\ell \in \mathbb{R}^{N_{\ell+1}\times N_\ell}$ and $\mathbf{b}_\ell \in \mathbb{R}^{N_\ell}$ for $\ell = 0,\ldots,L$. In this paper, we consider the case where $N_0 = d+1$ and $N_{L+1} = 1$. The number L is called the depth of neural networks. Additionally, $S := \sum_{\ell=0}^L(\|\mathbf{A}_\ell\|_0 + \|\mathbf{b}_\ell\|_0)$ represents the total number of non-zero weights within the neural network. We denote by $N(L,S)$ the set of neural networks with depth at most L and the number of non-zero weights at most S.

The following theorem establishes the convergence rate of the estimated Doob's guidance given in (4.7).

Theorem 5.3 (Convergence rate of Doob's guidance)**.** *Suppose Assumptions 1 and 2 hold. Let $t \in (0, T)$. Set the hypothesis class $\mathscr{H}_t$ as*

$$\mathscr{H}_t := \left\{ h_t \in N(L, S) : \begin{array}{c} \sup_{\mathbf{x}\in\mathbb{R}^d} h_t(\mathbf{x}) \leq \bar{B}, \ \inf_{\mathbf{x}\in\mathbb{R}^d} h_t(\mathbf{x}) \geq \underline{B}, \\ \max_{1\leq k\leq d} \sup_{\mathbf{x}\in\mathbb{R}^d} |D_k h_t(\mathbf{x})| \leq 2\sigma_{T-t}^{-2}\bar{B} \end{array} \right\}, \tag{5.2}$$

where $L = \mathcal{O}(\log n)$ and $S = \mathcal{O}(n^{\frac{d}{d+8}})$. Let $\widehat{h}_t^\lambda$ be the gradient-regularized empirical risk minimizer defined as (4.6), and let h_t^* be the Doob's h-function defined as (3.7). Then the following inequality holds:

$$\mathbb{E}\Big[\|\nabla \log \widehat{h}_t^\lambda - \nabla \log h_t^*\|_{L^2(p_{T-t})}^2\Big] \leq C\sigma_{T-t}^{-8} n^{-\frac{2}{d+8}} \log^4 n,$$

provided that the regularization parameter λ is set as $\lambda = \mathcal{O}(n^{-\frac{2}{d+8}})$, where C is a constant depending only on d, $\bar{B}$, and $\underline{B}$.

The proof of Theorem 5.3 is deferred to Appendix D. This theorem demonstrates that the L^2-error of the Doob's guidance estimator (4.7) converges to the exact Doob's guidance in (3.11) as the sample size increases, provided that the size of the neural network is appropriately chosen. However, since the prefactor σ_{T-t}^{-8} diverges as t approaches the terminal time T, early stopping in controllable diffusion models (4.8) is required to ensure the validity of the Doob's guidance estimator.

Remark 5.4 (Comparisons with previous work). Simultaneous estimation of a function and its gradient using deep neural network has been investigated by Ding et al. (2025b). The most important distinction in our work lies in the elimination of the convexity assumption on the hypothesis class. Specifically, Ding et al. (2025b, Lemma 7) propose an oracle inequality under the assumption that the hypothesis class is convex. Furthermore, Ding et al. (2025b, Theorem 3) requires the estimator to be a minimizer of the gradient-regularized empirical risk over the convex hull of a neural network class, which is intractable in practice. In contrast, Lemma 5.2 eliminates the requirement of convexity for the hypothesis class, and Theorem 5.3 removes the need for the convex hull of the neural network class. This aligns the theoretical analysis more closely with practical computing.

5.3 Error bounds for the controllable diffusion models. In this subsection, we establish a non-asymptotic convergence rate for the controllable diffusion models (4.8). We begin by presenting an error decomposition for the KL-divergence between the early-stopping distribution q_{T_0} and the distribution of $\widehat{\mathbf{Z}}_{T-T_0}^{\leftarrow}$.

Lemma 5.5 (Error decomposition)**.** *Suppose Assumptions 1, 2, and 3 hold. Let $\widehat{q}_{T-T_0}$ be the marginal density of $\widehat{\mathbf{Z}}_{T-T_0}^{\leftarrow}$ defined in (4.8). Then it follows that*

$$\mathrm{KL}(q_{T_0}\|\widehat{q}_{T-T_0}) \lesssim \underbrace{\frac{\bar{B}}{\underline{B}} \sum_{k=0}^{K-1} h\mathbb{E}^{\mathbb{P}}\Big[\|\nabla \log \widehat{h}_{kh}(\mathbf{X}_{kh}^{\leftarrow}) - \nabla \log h_{kh}^*(\mathbf{X}_{kh}^{\leftarrow})\|_2^2\Big]}_{\text{(I)}} + \underbrace{\frac{\bar{B}}{\underline{B}} T\varepsilon_{\mathrm{ref}}^2}_{\text{(II)}} + \underbrace{d\exp(-T)}_{\text{(III)}} + \underbrace{\frac{d^2T^2}{\sigma_{T_0}^4 K}}_{\text{(IV)}},$$

where the notation $\lesssim$ hides absolute constants.

The proof of Lemma 5.5 is deferred to Appendix E. Lemma 5.5 decomposes the KL-divergence between the early-stopping distribution q_{T_0} and the distribution of $\widehat{\mathbf{Z}}^{\leftarrow}_{T-T_0}$ into four components: (I) Doob's guidance error, (II) the reference score error, (III) the initialization error, and (IV) the discretization error. Specifically, Doob's guidance error represents the average error of Doob's h-guidance estimator at each time point, which has been investigated in Theorem 5.3; the reference score error is the average error of the reference score estimator at each time point, which is discussed in Assumption 3; the initialization error arises from replacing the initial distribution $q_T = p_T$ with a Gaussian distribution in (4.8); and the discretization error is induced by the exponential integrator.

While Lemma 5.5 characterizes the error between the estimated distribution $\widehat{q}_{T-T_0}$ and the early-stopping distribution q_{T_0}, our primary interest lies in the discrepancy between $\widehat{q}_{T-T_0}$ and the target tilted distribution q_0 defined in (3.1). Since the KL-divergence does not satisfy the triangular inequality, we instead propose an error bound in 2-Wasserstein distance. The following theorem establishes the 2-Wasserstein distance between the scaled and truncated distribution $(\mathcal{M} \circ \mathcal{T}_R)_\sharp \widehat{q}_{T-T_0}$ defined in (4.9) and the target tilted distribution q_0.

Theorem 5.6 (Convergence rate of controllable diffusion models)**.** *Suppose Assumptions 1, 2, and 3 hold. Let $\varepsilon \in (0,1)$. Set the hypothesis classes $\{\mathscr{H}_{T-kh}\}_{k=0}^{K=1}$ as (5.2) with the same depth L and number of non-zero parameters S as Theorem 5.3. Let $\widehat{q}_{T-T_0}$ be the marginal density of $\widehat{\mathbf{Z}}^{\leftarrow}_{T-T_0}$ defined in (4.8), and let $(\mathcal{M} \circ \mathcal{T}_R)_\sharp \widehat{q}_{T-T_0}$ defined as (4.9). Then it follows that*

$$\mathbb{E}\Big[\mathcal{W}_2^2(q_0, (\mathcal{M} \circ \mathcal{T}_R)_\sharp \widehat{q}_{T-T_0})\Big] \leq C\varepsilon \log^3\Big(\frac{1}{\varepsilon}\Big),$$

provided that the truncation radius R, the terminal time T, the step size h, the number of steps K, the error of reference score $\varepsilon_{\mathrm{ref}}$, the number of samples n for Doob's matching, and the early-stopping time T_0 are set, respectively, as

$$R \asymp \log^{\frac{1}{2}}\Big(\frac{1}{\varepsilon}\Big), \quad T \asymp \log\Big(\frac{1}{\varepsilon^2}\Big), \quad K \gtrsim \frac{1}{\varepsilon^4}\log^2\Big(\frac{1}{\varepsilon^2}\Big), \quad h \lesssim \varepsilon^4 \log^{-1}\Big(\frac{1}{\varepsilon^2}\Big)$$
$$\varepsilon_{\mathrm{ref}}^2 \lesssim \varepsilon^2 \log^{-1}\Big(\frac{1}{\varepsilon^2}\Big), \quad n \gtrsim \frac{1}{\varepsilon^{3(d+8)}} \log^{\frac{d+8}{2}}\Big(\frac{1}{\varepsilon^2}\Big).$$

Here C is a constant depending only on d, $\bar{B}$, and $\underline{B}$.

The proof of Theorem 5.6 is deferred to Appendix E. This theorem establishes a non-asymptotic convergence rate for the controllable diffusion model (4.8) using the variationally stable Doob's matching (4.6). Crucially, it provides theoretical guidance for selecting hyper-parameters, including the truncation radius R in (4.9), the step size h, the number of steps K in (4.8), the early stopping time T_0, the terminal time T, the reference score error $\varepsilon_{\mathrm{ref}}$ (Assumption 3), and the sample size n for Doob's matching in (4.4).

However, this rate suffers from the curse of dimensionality (CoD), implying that the required number of samples n grows exponentially as the error tolerance ε decays. We address this challenge in the remainder of this section under a low-dimensional subspace assumption.

5.4 Adaptivity to low-dimensionality. In this subsection, we demonstrate that the convergence rate mitigates the curse of dimensionality under a low-dimensional subspace assumption, a setting previously explored in Chen et al. (2023b, Section 3) and Oko et al. (2023, Section 6).

Assumption 4 (Low-dimensional subspace)**.** Let $d^* \ll d$ be an integer, and $\mathbf{P} \in \mathbb{R}^{d\times d^*}$ be a column orthogonal matrix, i.e., $\mathbf{P}^\top\mathbf{P} = \mathbf{I}_{d^*}$. Let $\bar{p}_0$ be a probability density with a compact support contained within a hypercube $\{\bar{\mathbf{x}}_0 \in \mathbb{R}^{d^*} : \|\bar{\mathbf{x}}_0\|_\infty \le 1\}$. The reference density p_0 is a push-forward of $\bar{p}_0$ by the linear map $\mathbf{P}$, i.e., $p_0 := \mathbf{P}_\sharp \bar{p}_0$.

Consequently, the reference density p_0 is supported on a linear subspace $\{\mathbf{P}\bar{\mathbf{x}}_0 \in \mathbb{R}^d : \bar{\mathbf{x}}_0 \in \mathbb{R}^{d^*}\}$ with an ambient dimension d, and a much smaller intrinsic dimension $d^* \ll d$.

Before proceeding, we define the forward and time-reversal process in the low-dimensional latent space. Analogously to (2.1), the forward process reads

$$\mathrm{d}\bar{\mathbf{X}}_t = -\bar{\mathbf{X}}_t\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}\bar{\mathbf{B}}_t, \quad t \in (0,T),\ \bar{\mathbf{X}}_0 \sim \bar{p}_0,$$

where $\bar{\mathbf{B}}_t$ is a d^*-dimensional standard Brownian motion, and $T > 0$ is the terminal time. The transition distribution of this forward process is given by:

$$(\bar{\mathbf{X}}_t | \bar{\mathbf{X}}_0 = \bar{\mathbf{x}}_0) \sim \mathcal{N}(\mu_t\bar{\mathbf{x}}_0, \sigma_t^2\mathbf{I}_{d^*}). \tag{5.3}$$

Let $\bar{p}_t$ denote the marginal density of $\bar{\mathbf{X}}_t$ for $t \in (0,T)$. The corresponding time-reversal process (Anderson, 1982) is defined as

$$\begin{aligned}\mathrm{d}\bar{\mathbf{X}}_t^{\leftarrow} &= (\bar{\mathbf{X}}_t^{\leftarrow} + 2\nabla\log\bar{p}_{T-t}(\bar{\mathbf{X}}_t^{\leftarrow}))\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}\bar{\mathbf{B}}_t, \quad t \in (0,T),\\ \bar{\mathbf{X}}_0^{\leftarrow} &\sim \bar{p}_T.\end{aligned}$$

As shown by Anderson (1982), the path measure of the time-reversal process $(\bar{\mathbf{X}}_t^{\leftarrow})_{0\le t\le T}$ corresponds exactly to the reverse of the forward process $(\bar{\mathbf{X}}_t)_{0\le t\le T}$.

The following result establishes a relationship between the ground-truth Doob's h-function (3.7) and its analogue $\bar{h}_t^* : \mathbb{R}^{d^*} \to \mathbb{R}$ in the low-dimensional latent space. In other words, it provides a low-dimensional representation of the ground-truth Doob's h-function.

Proposition 5.7 (Low-dimensional representation)**.** *Suppose Assumptions 4 and 2 hold. Then for any $t \in (0,T)$ and $\mathbf{x} \in \mathbb{R}^d$, we have*

$$h_t^*(\mathbf{x}) = \bar{h}_t^*(\mathbf{P}^\top\mathbf{x}) := \mathbb{E}[w(\mathbf{P}\bar{\mathbf{X}}_T^{\leftarrow}) \mid \bar{\mathbf{X}}_t^{\leftarrow} = \mathbf{P}^\top\mathbf{x}]. \tag{5.4}$$

The proof of Proposition 5.7 is provided in Appendix F. Proposition 5.7 implies that estimating the ground-truth Doob's h-function reduces to estimating its low-dimensional counterpart $\bar{h}_t^* : \mathbb{R}^{d^*} \to \mathbb{R}$, thereby enabling the Doob's guidance estimator to adapt to low-dimensional structures.

Analogously to Proposition 5.1, we can establish the regularity properties of the low-dimensional representation of Doob's h-function $\bar{h}_t^* : \mathbb{R}^{d^*} \to \mathbb{R}$ in (5.4).

Proposition 5.8. *Suppose Assumptions 4 and 2 hold. Then for all $t \in (0,T)$ and $\bar{\mathbf{x}} \in \mathbb{R}^{d^*}$, the following bounds hold:*

(i) $\underline{B} \le \bar{h}_t^*(\bar{\mathbf{x}}) \le \bar{B}$;

(ii) $\max_{1\le k\le d} |D_k\bar{h}_t^*(\bar{\mathbf{x}})| \le 2\sigma_{T-t}^{-2}\bar{B}$; *and*

(iii) $\max_{1\le k,\ell\le d} |D_{k\ell}^2\bar{h}_t^*(\bar{\mathbf{x}})| \le 6\sigma_{T-t}^{-4}\bar{B}$,

where D_k and $D_{k\ell}^2$ denote the first-order and second-order partial derivatives with respect to the input coordinates, respectively.

The proof of Proposition 5.7 is provided in Appendix F. Based on these results, we derive the convergence rates for the variationally stable Doob's matching under the assumption of low-dimensional subspace.

Theorem 5.9 (Adaptivity to intrinsic dimension)**.** *Suppose Assumptions 4 and 2 hold. Let $t \in (0, T)$. Set the hypothesis class $\mathscr{H}_t$ as*

$$\mathscr{H}_t := \left\{ h_t \in N(L, S) : \begin{array}{l} \sup_{\mathbf{x} \in \mathbb{R}^d} h_t(\mathbf{x}) \le \bar{B}, \quad \inf_{\mathbf{x} \in \mathbb{R}^d} h_t(\mathbf{x}) \ge \underline{B}, \\ \max_{1 \le k \le d} \sup_{\mathbf{x} \in \mathbb{R}^d} |D_k h_t(\mathbf{x})| \le 2\sigma_{T-t}^{-2} \bar{B} \end{array} \right\}, \tag{5.5}$$

where $L = \mathcal{O}(\log n)$ and $S = \mathcal{O}(n^{\frac{d^*}{d^*+8}})$. Let $\widehat{h}_t^\lambda$ be the gradient-regularized empirical risk minimizer defined as (4.6), and let h_t^* be the Doob's h-function defined as (3.7). Then the following inequality holds:

$$\mathbb{E}\Big[\|\nabla \log \widehat{h}_t^\lambda - \nabla \log h_t^*\|_{L^2(p_{T-t})}^2 \Big] \le C \sigma_{T-t}^{-8} n^{-\frac{2}{d^*+8}} \log^4 n,$$

provided that the regularization parameter λ is set as $\lambda = \mathcal{O}(n^{-\frac{2}{d^*+8}})$, where C is a constant depending only on d^*, $\bar{B}$, and $\underline{B}$.

The proof of Theorem 5.9 is provided in Appendix F. This result confirms that the convergence rate eliminates the exponential dependence on the ambient dimension d, scaling exponentially solely with the intrinsic dimension $d^* \ll d$. This effectively mitigates the curse of dimensionality in Theorem 5.3.

The following corollary is a direct consequence of Theorem 5.9, derived using arguments similar to those in Theorem 5.6.

Corollary 5.10. *Suppose Assumptions 4, 2, and 3 hold. Let $\varepsilon \in (0, 1)$. Set the hypothesis classes $\{\mathscr{H}_{T-kh}\}_{k=0}^{K=1}$ as (5.2) with the same depth L and number of non-zero parameters S as Theorem 5.3. Let $\widehat{q}_{T-T_0}$ be the marginal density of $\widehat{\mathbf{Z}}_{T-T_0}^{\leftarrow}$ defined in (4.8), and let $(\mathcal{M} \circ \mathcal{T}_R)_\sharp \widehat{q}_{T-T_0}$ defined as (4.9). Then it follows that*

$$\mathbb{E}\Big[\mathcal{W}_2^2(q_0, (\mathcal{M} \circ \mathcal{T}_R)_\sharp \widehat{q}_{T-T_0}) \Big] \le C \varepsilon \log^3 \Big(\frac{1}{\varepsilon}\Big).$$

provided that the truncation radius R, the terminal time T, the step size h, the number of steps K, the error of reference score $\varepsilon_{\mathrm{ref}}$, the number of samples n for Doob's matching, and the early-stopping time T_0 are set, respectively, as

$$\begin{aligned} &R \asymp \log^{\frac{1}{2}} \Big(\frac{1}{\varepsilon}\Big), \quad T \asymp \log \Big(\frac{1}{\varepsilon^2}\Big), \quad K \gtrsim \frac{1}{\varepsilon^4} \log^2 \Big(\frac{1}{\varepsilon^2}\Big), \quad h \lesssim \varepsilon^4 \log^{-1} \Big(\frac{1}{\varepsilon^2}\Big) \\ &\varepsilon_{\mathrm{ref}}^2 \lesssim \varepsilon^2 \log^{-1} \Big(\frac{1}{\varepsilon^2}\Big), \quad n \gtrsim \frac{1}{\varepsilon^{3(d^*+8)}} \log^{\frac{d+8}{2}} \Big(\frac{1}{\varepsilon^2}\Big). \end{aligned}$$

Here C is a constant depending only on d^*, d, $\bar{B}$, and $\underline{B}$.

Crucially, the convergence rates in Corollary 5.10 depend only polynomially on the ambient dimension d, while the sample complexity depends exponentially solely on the intrinsic dimension $d^* \ll d$. This result significantly mitigates the curse of dimensionality.

Remark 5.11. Adaptivity to low dimensionality plays a pivotal role in the analysis of diffusion and flow-based models. One line of work studies the adaptivity of score or velocity estimator to low dimensionality under Assumption 4 or its variants; see e.g., Chen et al. (2023b); Oko et al. (2023); Yakovlev and Puchkin (2025a); Ding et al. (2025a). A second line of work focuses on the adaptivity of the sampling procedure; see e.g., Li and Yan (2024); Huang et al. (2024); Potaptchik et al. (2025). These works provide valuable insights for extending Corollary 5.10 to achieve provable adaptivity across reference score estimation, sampling, and guidance estimation, thereby completely eliminating dependence on the ambient dimension. We leave this unified analysis, which is outside the scope of the current work, for future research. Importantly, even within the current framework, the dependence on the ambient dimension d remains only polynomial.

6 Conclusions

In this work, we proposed variationally stable Doob's matching, a principled inference-time alignment framework for diffusion models grounded in the theory of Doob's h-transform. Our approach reformulates guidance as the gradient of the logarithm of an underlying Doob's h-function, providing a mathematically consistent mechanism for tilting a pre-trained diffusion model toward a target distribution without retraining the reference score network. By leveraging gradient-regularized regression, Doob's matching simultaneously estimates both the h-function and its gradient, thereby providing a consistent estimator for Doob's guidance.

From a theoretical perspective, we established non-asymptotic convergence rates for the proposed guidance estimator, showing that the estimated Doob's guidance converges to the true guidance under suitable choices of the hypothesis class and regularization parameter. Building on this result, we further derived non-asymptotic convergence guarantees for the induced controllable diffusion process, demonstrating that the generated distribution converges to the target distribution in the 2-Wasserstein distance. These results provide an end-to-end theoretical guarantees for inference-time aligned diffusion models that explicitly account for guidance estimation error, reference score estimation error, initialization bias, and discretization error. Furthermore, we show that our convergence rates depend solely on the intrinsic dimension of the linear subspace rather than the ambient dimension. This highlights the estimator's adaptivity to low-dimensional structures, effectively mitigating the curse of dimensionality.

References

Brian D.O. Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982. (Cited on pages 4 and 18.)

Martin Anthony, Peter L Bartlett, Peter L Bartlett, et al. *Neural network learning: Theoretical foundations*, volume 9. Cambridge University Press, 1999. (Cited on pages 13 and 42.)

Yara Bahram, Mohammadhadi Shateri, and Eric Granger. Dogfit: Domain-guided fine-tuning for efficient transfer learning of diffusion models, 2026. arXiv:2508.05685. (Cited on page 7.)

Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and Geometry of Markov Diffusion Operators*, volume 348 of *Grundlehren der mathematischen Wissenschaften (GL)*. Springer Cham, first edition, 2014. (Cited on page 5.)

Feng Bao, Zezhong Zhang, and Guannan Zhang. A score-based filter for nonlinear data assimilation. *Journal of Computational Physics*, 514:113207, 2024. (Cited on page 1.)

Peter L. Bartlett and Shahar Mendelson. Rademacher and gaussian complexities: Risk bounds and structural results. *Journal of Machine Learning Research*, 3:463–482, 2002. (Cited on pages 41 and 43.)

Peter L. Bartlett, Nick Harvey, Christopher Liaw, and Abbas Mehrabian. Nearly-tight VC-dimension and pseudodimension bounds for piecewise linear neural networks. *Journal of Machine Learning Research*, 20(63):1–17, 2019. (Cited on pages 13 and 47.)

Benedikt Bauer and Michael Kohler. On deep learning as a remedy for the curse of dimensionality in nonparametric regression. *The Annals of Statistics*, 47(4):2261–2285, 2019. (Cited on pages 4 and 13.)

Denis Belomestny, Alexey Naumov, Nikita Puchkin, and Sergey Samsonov. Simultaneous approximation of a smooth function and its derivatives by deep neural networks with piecewise-polynomial activations. *Neural Networks*, 161:242–253, 2023. (Cited on page 15.)

Eliot Beyler and Francis Bach. Convergence of deterministic and stochastic diffusion-model samplers: A simple analysis in wasserstein distance, 2025. arXiv:2508.03210. (Cited on pages 12 and 13.)

Kevin Black, Michael Janner, Yilun Du, Ilya Kostrikov, and Sergey Levine. Training diffusion models with reinforcement learning. In *The Twelfth International Conference on Learning Representations*, 2024. (Cited on page 2.)

Haim Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Universitext (UTX). Springer New York, NY, first edition, 2011. (Cited on page 33.)

Jinyuan Chang, Zhao Ding, Yuling Jiao, Ruoxuan Li, and Jerry Zhijian Yang. Deep conditional distribution learning via conditional föllmer flow, 2025a. arXiv:2402.01460. (Cited on page 12.)

Jinyuan Chang, Chenguang Duan, Yuling Jiao, Ruoxuan Li, Jerry Zhijian Yang, and Cheng Yuan. Provable diffusion posterior sampling for Bayesian inversion, 2025b. arXiv:2512.08022. (Cited on pages 1, 2, 5, and 13.)

Haoxuan Chen, Yinuo Ren, Martin Renqiang Min, Lexing Ying, and Zachary Izzo. Solving inverse problems via diffusion-based priors: An approximation-free ensemble sampling approach, 2025. arXiv:2506.03979. (Cited on pages 1 and 2.)

Hongrui Chen, Holden Lee, and Jianfeng Lu. Improved analysis of score-based generative modeling: User-friendly bounds under minimal smoothness assumptions. In *Proceedings of the 40th International Conference on Machine Learning*, pages 4735–4763, 2023a. (Cited on pages 11, 13, 48, and 49.)

Minshuo Chen, Kaixuan Huang, Tuo Zhao, and Mengdi Wang. Score approximation, estimation and distribution recovery of diffusion models on low-dimensional data. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 4672–4712. PMLR, 23–29 Jul 2023b. (Cited on pages 17 and 20.)

Sinho Chewi. Log-concave sampling, 2025. URL `https://chewisinho.github.io/main.pdf`. unfinished draft. (Cited on pages 2 and 7.)

Hyungjin Chung, Jeongsol Kim, Michael Thompson Mccann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *The Eleventh International Conference on Learning Representations*, 2023. (Cited on pages 1, 2, and 5.)

Kevin Clark, Paul Vicol, Kevin Swersky, and David J. Fleet. Directly fine-tuning diffusion models on differentiable rewards. In *The Twelfth International Conference on Learning Representations*, 2024. (Cited on page 2.)

Alexander Denker, Francisco Vargas, Shreyas Padhy, Kieran Didi, Simon Mathis, Vincent Dutordoir, Riccardo Barbano, Emile Mathieu, Urszula Julia Komorowska, and Pietro Lio. Deft: Efficient fine-tuning of diffusion models by learning the generalised h-transform. In *Advances in Neural Information Processing Systems*, volume 37, pages 19636–19682. Curran Associates, Inc., 2024. (Cited on pages 2 and 3.)

Alexander Denker, Shreyas Padhy, Francisco Vargas, and Johannes Hertrich. Iterative importance fine-tuning of diffusion models, 2025. arXiv:2502.04468. (Cited on pages 2, 3, and 6.)

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat GANs on image synthesis. In *Advances in Neural Information Processing Systems*, volume 34, pages 8780–8794. Curran Associates, Inc., 2021. (Cited on pages 2 and 10.)

Zhao Ding, Chenguang Duan, Yuling Jiao, Jerry Zhijian Yang, Cheng Yuan, and Pingwen Zhang. Nonlinear assimilation via score-based sequential Langevin sampling, 2024. arXiv:2411.13443. (Cited on pages 1, 5, and 13.)

Zhao Ding, Chenguang Duan, Yuling Jiao, Ruoxuan Li, Jerry Zhijian Yang, and Pingwen Zhang. Characteristic learning for provable one step generation, 2025a. arXiv:2405.05512. (Cited on pages 4, 13, and 20.)

Zhao Ding, Chenguang Duan, Yuling Jiao, and Jerry Zhijian Yang. Semi-supervised deep Sobolev regression: Estimation and variable selection by ReQU neural network. *IEEE Transactions on Information Theory*, 2025b. (Cited on pages 10, 16, 45, and 47.)

Carles Domingo-Enrich, Jiequn Han, Brandon Amos, Joan Bruna, and Ricky T. Q. Chen. Stochastic optimal control matching. In *Advances in Neural Information Processing Systems*, volume 37, pages 112459–112504. Curran Associates, Inc., 2024. (Cited on page 2.)

Carles Domingo-Enrich, Michal Drozdzal, Brian Karrer, and Ricky T. Q. Chen. Adjoint matching: Fine-tuning flow and diffusion generative models with memoryless stochastic optimal control. In *The Thirteenth International Conference on Learning Representations*, 2025. (Cited on pages 2 and 6.)

H. Drucker and Y. Le Cun. Double backpropagation increasing generalization performance. In *IJCNN-91-Seattle International Joint Conference on Neural Networks*, volume 2, pages 145–150, 1991. (Cited on page 10.)

H. Drucker and Y. Le Cun. Improving generalization performance using double backpropagation. *IEEE Transactions on Neural Networks*, 3(6):991–997, 1992. (Cited on page 10.)

Chenguang Duan, Yuling Jiao, Yanming Lai, Dingwei Li, Xiliang Lu, and Jerry Zhijian Yang. Convergence rate analysis for deep Ritz method. *Communications in Computational Physics*, 31(4):1020–1048, 2022a. (Cited on page 15.)

Chenguang Duan, Yuling Jiao, Yanming Lai, Xiliang Lu, Qimeng Quan, and Jerry Zhijian Yang. Deep Ritz methods for Laplace equations with Dirichlet boundary condition. *CSIAM Transactions on Applied Mathematics*, 3(4):761–791, 2022b. (Cited on page 15.)

Ying Fan, Olivia Watkins, Yuqing Du, Hao Liu, Moonkyung Ryu, Craig Boutilier, Pieter Abbeel, Mohammad Ghavamzadeh, Kangwook Lee, and Kimin Lee. DPOK: Reinforcement learning for fine-tuning text-to-image diffusion models. In *Advances in Neural Information Processing Systems*, volume 36, pages 79858–79885. Curran Associates, Inc., 2023. (Cited on page 2.)

Hengyu Fu, Zhuoran Yang, Mengdi Wang, and Minshuo Chen. Unveil conditional diffusion models with classifier-free guidance: A sharp statistical theory, 2024. arXiv:2403.11968. (Cited on pages 4 and 13.)

Leo Gao, John Schulman, and Jacob Hilton. Scaling laws for reward model overoptimization. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 10835–10866. PMLR, 23–29 Jul 2023. (Cited on page 2.)

Ingo Gühring and Mones Raslan. Approximation rates for neural networks with encodable weights in smoothness spaces. *Neural Networks*, 134:107–130, 2021. (Cited on page 15.)

Ingo Gühring, Gitta Kutyniok, and Philipp Petersen. Error bounds for approximations with deep relu neural networks in ws,p norms. *Analysis and Applications*, 18(05):803–859, 2020. (Cited on page 15.)

Yinbin Han, Meisam Razaviyayn, and Renyuan Xu. Stochastic control for fine-tuning diffusion models: Optimality, regularity, and convergence, 2024. arXiv:2412.18164. (Cited on page 2.)

Jeremy Heng, Valentin De Bortoli, and Arnaud Doucet. Diffusion schrödinger bridges for bayesian computation. *Statistical Science*, 39(1):90 – 99, 2024. (Cited on pages 2 and 7.)

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. In *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*, 2021. (Cited on page 2.)

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851. Curran Associates, Inc., 2020. (Cited on page 1.)

Marlis Hochbruck and Alexander Ostermann. Explicit exponential runge-kutta methods for semilinear parabolic problems. *SIAM Journal on Numerical Analysis*, 43(3):1069–1090, 2005. (Cited on page 5.)

Marlis Hochbruck and Alexander Ostermann. Exponential integrators. *Acta Numerica*, 19: 209–286, 2010. (Cited on page 5.)

Zhihan Huang, Yuting Wei, and Yuxin Chen. Denoising diffusion probabilistic models are optimally adaptive to unknown low dimensionality, 2024. arXiv:2410.18784. (Cited on page 20.)

Aapo Hyvärinen. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(24):695–709, 2005. (Cited on pages 4 and 13.)

Yuchen Jiao, Yuxin Chen, and Gen Li. Towards a unified framework for guided diffusion models, 2025. arXiv:2512.04985. (Cited on page 2.)

Yuling Jiao, Guohao Shen, Yuanyuan Lin, and Jian Huang. Deep nonparametric regression on approximate manifolds: Nonasymptotic error bounds with polynomial prefactors. *The Annals of Statistics*, 51(2):691 – 716, 2023. (Cited on pages 4 and 13.)

Nikolas Kantas, Alexandros Beskos, and Ajay Jasra. Sequential Monte Carlo methods for high-dimensional inverse problems: A case study for the Navier–Stokes equations. *SIAM/ASA Journal on Uncertainty Quantification*, 2(1):464–489, 2014. (Cited on page 5.)

Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*, volume 113 of *Graduate Texts in Mathematics (GTM)*. Springer New York, NY, second edition, 1998. (Cited on pages 8 and 32.)

Sunwoo Kim, Minkyu Kim, and Dongmin Park. Test-time alignment of diffusion models without reward over-optimization. In *The Thirteenth International Conference on Learning Representations*, 2025. (Cited on pages 2 and 6.)

Michael Kohler and Sophie Langer. On the rate of convergence of fully connected deep neural network regression estimates. *The Annals of Statistics*, 49(4):2231–2249, 2021. (Cited on pages 4 and 13.)

Gitte Kremling, Francesco Iafrate, Mahsa Taheri, and Johannes Lederer. Non-asymptotic error bounds for probability flow odes under weak log-concavity, 2025. arXiv:2510.17608. (Cited on page 13.)

Holden Lee, Jianfeng Lu, and Yixin Tan. Convergence of score-based generative modeling for general data distributions. In *Proceedings of The 34th International Conference on Algorithmic Learning Theory*, volume 201 of *Proceedings of Machine Learning Research*, pages 946–985. PMLR, 20 Feb–23 Feb 2023a. (Cited on pages 11, 12, and 13.)

Kimin Lee, Hao Liu, Moonkyung Ryu, Olivia Watkins, Yuqing Du, Craig Boutilier, Pieter Abbeel, Mohammad Ghavamzadeh, and Shixiang Shane Gu. Aligning text-to-image models using human feedback, 2023b. arXiv:2302.12192. (Cited on pages 2 and 6.)

Bo Li, Shanshan Tang, and Haijun Yu. Better approximations of high dimensional smooth functions by deep neural networks with rectified power units. *Communications in Computational Physics*, 27(2):379–411, 2019. (Cited on page 15.)

Gen Li and Yuling Yan. Adapting to unknown low-dimensional structures in score-based diffusion models. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. (Cited on page 20.)

Zhuoyuan Li, Bin Dong, and Pingwen Zhang. State-observation augmented diffusion model for nonlinear assimilation with unknown dynamics. *Journal of Computational Physics*, 539: 114240, 2025. (Cited on page 1.)

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan LI, and Jun Zhu. Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps. In *Advances in Neural Information Processing Systems*, volume 35, pages 5775–5787. Curran Associates, Inc., 2022a. (Cited on page 5.)

Yiping Lu, Haoxuan Chen, Jianfeng Lu, Lexing Ying, and Jose Blanchet. Machine learning for elliptic PDEs: Fast rate generalization bound, neural scaling law and minimax optimality. In *International Conference on Learning Representations*, 2022b. (Cited on page 15.)

Morteza Mardani, Jiaming Song, Jan Kautz, and Arash Vahdat. A variational perspective on solving inverse problems with diffusion models. In *The Twelfth International Conference on Learning Representations*, 2024. (Cited on page 5.)

Ségolène Tiffany Martin, Anne Gagneux, Paul Hagemann, and Gabriele Steidl. Pnp-flow: Plug-and-play image restoration with flow matching. In *The Thirteenth International Conference on Learning Representations*, 2025. (Cited on page 2.)

Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of Machine Learning*. MIT press, Second edition, 2018. (Cited on pages 41, 42, and 43.)

Wenlong Mou. Is rl fine-tuning harder than regression? a pde learning approach for diffusion models, 2025. arXiv:2509.02528. (Cited on pages 3 and 10.)

Kazusato Oko, Shunta Akiyama, and Taiji Suzuki. Diffusion models are minimax optimal distribution estimators. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 26517–26582. PMLR, 23–29 Jul 2023. (Cited on pages 4, 12, 13, 17, and 20.)

Bernt Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Universitext (UTX). Springer Berlin, Heidelberg, sixth edition, 2003. (Cited on pages 7, 8, 9, 30, and 32.)

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul F Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems*, volume 35, pages 27730–27744. Curran Associates, Inc., 2022. (Cited on page 6.)

Yidong Ouyang, Liyan Xie, Hongyuan Zha, and Guang Cheng. Transfer learning for diffusion models. In *Advances in Neural Information Processing Systems*, volume 37, pages 136962–136989. Curran Associates, Inc., 2024. (Cited on pages 7 and 10.)

Jean Pachebat, Giovanni Conforti, Alain Durmus, and Yazid Janati. Iterative tilting for diffusion fine-tuning, 2025. arXiv:2512.03234. (Cited on page 2.)

Peter Potaptchik, Iskander Azangulov, and George Deligiannidis. Linear convergence of diffusion models under the manifold hypothesis, 2025. arXiv:2410.09046. (Cited on page 20.)

Vishal Purohit, Matthew Repasky, Jianfeng Lu, Qiang Qiu, Yao Xie, and Xiuyuan Cheng. Consistency posterior sampling for diverse image synthesis. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025. (Cited on pages 2, 5, and 13.)

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. In *Advances in Neural Information Processing Systems*, volume 36, pages 53728–53741. Curran Associates, Inc., 2023. (Cited on page 6.)

Rafael Rafailov, Yaswanth Chittepu, Ryan Park, Harshit Sushil Sikchi, Joey Hejna, Brad Knox, Chelsea Finn, and Scott Niekum. Scaling laws for reward model overoptimization in direct alignment algorithms. In *Advances in Neural Information Processing Systems*, volume 37, pages 126207–126242. Curran Associates, Inc., 2024. (Cited on page 2.)

Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. Zero-shot text-to-image generation. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 8821–8831. PMLR, 18–24 Jul 2021. (Cited on page 1.)

Yinuo Ren, Wenhao Gao, Lexing Ying, Grant M. Rotskoff, and Jiequn Han. Driftlite: Lightweight drift control for inference-time scaling of diffusion models, 2025. arXiv:2509.21655. (Cited on pages 2 and 6.)

L. C. G. Rogers and David Williams. *Diffusions, Markov Processes and Martingales*, volume 2: Itô Calculus of *Cambridge Mathematical Library*. Cambridge University Press, 2 edition, 2000. (Cited on pages 2 and 7.)

Amirmojtaba Sabour, Michael S. Albergo, Carles Domingo-Enrich, Nicholas M. Boffi, Sanja Fidler, Karsten Kreis, and Eric Vanden-Eijnden. Test-time scaling of diffusions with flow maps, 2025. arXiv:2511.22688. (Cited on pages 2 and 6.)

Simo Särkkä and Arno Solin. *Applied Stochastic Differential Equations*. Institute of Mathematical Statistics Textbooks. Cambridge University Press, 2019. (Cited on pages 2 and 7.)

Johannes Schmidt-Hieber. Nonparametric regression using deep neural networks with ReLU activation function. *The Annals of Statistics*, 48(4):1875–1897, 2020. (Cited on pages 4 and 13.)

Guohao Shen, Yuling Jiao, Yuanyuan Lin, and Jian Huang. Approximation with cnns in sobolev space: with applications to classification. In *Advances in Neural Information Processing Systems*, volume 35, pages 2876–2888. Curran Associates, Inc., 2022. (Cited on page 15.)

Guohao Shen, Yuling Jiao, Yuanyuan Lin, and Jian Huang. Differentiable neural networks with repu activation: with applications to score estimation and isotonic regression, 2024. arXiv:2305.00608. (Cited on page 15.)

Phillip Si and Peng Chen. Latent-enSF: A latent ensemble score filter for high-dimensional data assimilation with sparse observation data. In *The Thirteenth International Conference on Learning Representations*, 2025. (Cited on page 1.)

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 2256–2265, Lille, France, 07–09 Jul 2015. PMLR. (Cited on page 1.)

Jiaming Song, Qinsheng Zhang, Hongxu Yin, Morteza Mardani, Ming-Yu Liu, Jan Kautz, Yongxin Chen, and Arash Vahdat. Loss-guided diffusion models for plug-and-play controllable generation. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 32483–32498. PMLR, 23–29 Jul 2023. (Cited on pages 1 and 2.)

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. (Cited on page 1.)

Yang Song, Sahaj Garg, Jiaxin Shi, and Stefano Ermon. Sliced score matching: A scalable approach to density and score estimation. In *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*, volume 115 of *Proceedings of Machine Learning Research*, pages 574–584. PMLR, 22–25 Jul 2020. (Cited on pages 4 and 13.)

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021. (Cited on pages 1 and 4.)

Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models. In *International Conference on Learning Representations*, 2022. (Cited on page 5.)

Nisan Stiennon, Long Ouyang, Jeffrey Wu, Daniel Ziegler, Ryan Lowe, Chelsea Voss, Alec Radford, Dario Amodei, and Paul F Christiano. Learning to summarize with human feedback. In *Advances in Neural Information Processing Systems*, volume 33, pages 3008–3021. Curran Associates, Inc., 2020. (Cited on page 6.)

Andrew M Stuart. Inverse problems: A Bayesian perspective. *Acta Numerica*, 19:451–559, 2010. (Cited on page 5.)

Rong Tang and Yun Yang. Adaptivity of diffusion models to manifold structures. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pages 1648–1656. PMLR, 02–04 May 2024. (Cited on pages 4 and 13.)

Wenpin Tang and Fuzhong Zhou. Fine-tuning of diffusion models via stochastic control: entropy regularization and beyond, 2025. arXiv:2403.06279. (Cited on pages 2 and 6.)

Wenping Tang and Renyuan Xu. A stochastic analysis approach to conditional diffusion guidance, 2024. Columbia University Preprint. (Cited on pages 2, 3, and 10.)

Masatoshi Uehara, Yulai Zhao, Kevin Black, Ehsan Hajiramezanali, Gabriele Scalia, Nathaniel Lee Diamant, Alex M Tseng, Tommaso Biancalani, and Sergey Levine. Fine-tuning of continuous-time diffusion models as entropy-regularized control, 2024a. arXiv:2402.15194. (Cited on page 6.)

Masatoshi Uehara, Yulai Zhao, Kevin Black, Ehsan Hajiramezanali, Gabriele Scalia, Nathaniel Lee Diamant, Alex M Tseng, Tommaso Biancalani, and Sergey Levine. Fine-tuning of continuous-time diffusion models as entropy-regularized control, 2024b. arXiv:2402.15194. (Cited on page 2.)

Masatoshi Uehara, Yulai Zhao, Kevin Black, Ehsan Hajiramezanali, Gabriele Scalia, Nathaniel Lee Diamant, Alex M Tseng, Sergey Levine, and Tommaso Biancalani. Feedback efficient online fine-tuning of diffusion models. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 48892–48918. PMLR, 21–27 Jul 2024c. (Cited on page 2.)

Masatoshi Uehara, Xingyu Su, Yulai Zhao, Xiner Li, Aviv Regev, Shuiwang Ji, Sergey Levine, and Tommaso Biancalani. Reward-guided iterative refinement in diffusion models at test-time with applications to protein and DNA design, 2025a. arXiv:2502.14944. (Cited on page 1.)

Masatoshi Uehara, Yulai Zhao, Chenyu Wang, Xiner Li, Aviv Regev, Sergey Levine, and Tommaso Biancalani. Inference-time alignment in diffusion models with reward-guided generation: Tutorial and review, 2025b. arXiv:2501.09685. (Cited on pages 2, 6, and 10.)

V. N. Vapnik and A. Ya. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability & Its Applications*, 16(2):264–280, 1971. (Cited on page 13.)

Santosh Vempala and Andre Wibisono. Rapid convergence of the unadjusted Langevin algorithm: isoperimetry suffices. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. (Cited on page 5.)

Cédric Villani. *Optimal Transport: Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften (GL)*. Springer Berlin, Heidelberg, first edition, 2009. (Cited on page 52.)

Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural Computation*, 23(7):1661–1674, 2011. (Cited on pages 4 and 13.)

Xiyu Wang, Baijiong Lin, Daochang Liu, Ying-Cong Chen, and Chang Xu. Bridging data gaps in diffusion models with adversarial noise-based transfer learning. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 50944–50959. PMLR, 21–27 Jul 2024. (Cited on page 7.)

Jiazheng Xu, Xiao Liu, Yuchen Wu, Yuxuan Tong, Qinkai Li, Ming Ding, Jie Tang, and Yuxiao Dong. ImageReward: Learning and evaluating human preferences for text-to-image generation. In *Advances in Neural Information Processing Systems*, volume 36, pages 15903–15935. Curran Associates, Inc., 2023. (Cited on page 2.)

Konstantin Yakovlev and Nikita Puchkin. Generalization error bound for denoising score matching under relaxed manifold assumption. In *Proceedings of Thirty Eighth Conference on Learning Theory*, volume 291 of *Proceedings of Machine Learning Research*, pages 5824–5891. PMLR, 30 Jun–04 Jul 2025a. (Cited on pages 13 and 20.)

Konstantin Yakovlev and Nikita Puchkin. Simultaneous approximation of the score function and its derivatives by deep neural networks, 2025b. arXiv:2512.23643. (Cited on page 15.)

Konstantin Yakovlev, Anna Markovich, and Nikita Puchkin. Implicit score matching meets denoising score matching: improved rates of convergence and log-density hessian estimation, 2025. arXiv:2512.24378. (Cited on page 13.)

Hui Yuan, Kaixuan Huang, Chengzhuo Ni, Minshuo Chen, and Mengdi Wang. Reward-directed conditional diffusion: Provable distribution estimation and reward improvement. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. (Cited on page 2.)

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3836–3847, October 2023. (Cited on page 2.)

Qinsheng Zhang and Yongxin Chen. Fast sampling of diffusion models with exponential integrator. In *The Eleventh International Conference on Learning Representations*, 2023. (Cited on page 5.)

Jincheng Zhong, Xiangcheng Zhang, Jianmin Wang, and Mingsheng Long. Domain guidance: A simple transfer approach for a pre-trained diffusion model, 2025. arXiv:2504.01521. (Cited on page 7.)

Outline of Appendices

A Derivations in Section 3

Lemma A.1. *The Doob's h-transform $h^*(t, \mathbf{X}_t^{\leftarrow})$ defined as (3.7) is a martingale, and satisfies the following SDE:*

$$\mathrm{d}h^*(t, \mathbf{X}_t^{\leftarrow}) = \nabla h^*(t, \mathbf{X}_t^{\leftarrow})^\top \sqrt{2}\, \mathrm{d}\mathbf{B}_t.$$

Proof of Lemma A.1. This proof is divided into two parts.
Part 1. Martingale. Due to the Markov property of the diffusion process $\mathbf{X}_t^{\leftarrow}$ (Øksendal, 2003, Theorem 7.1.2), using (3.7) implies

$$M_t := h^*(t, \mathbf{X}_t^{\leftarrow}) = \mathbb{E}^{\mathbb{P}}[w(\mathbf{X}_T^{\leftarrow})|\mathcal{F}_t].$$

It is apparent that M_t is $\mathcal{F}_t$-measurable for each $t \in (0, T)$, thus M_t is adapted to $\mathbb{F}$. Then we show that M_t is integrable under $\mathbb{P}$. Indeed,

$$\mathbb{E}^{\mathbb{P}}[|M_t|] \leq \mathbb{E}^{\mathbb{P}}[\mathbb{E}^{\mathbb{P}}[|w(\mathbf{X}_T^{\leftarrow})||\mathcal{F}_t]] = \mathbb{E}^{\mathbb{P}}[w(\mathbf{X}_T^{\leftarrow})] = Z < \infty,$$

where the first inequality holds from Jensen's inequality, and the first equality used the law of total expectation and the fact that $w(\mathbf{X}_T^{\leftarrow})$. We next show the martingale property. For each $0 \leq s \leq t \leq T$,

$$\mathbb{E}^{\mathbb{P}}[M_t|\mathcal{F}_s] = \mathbb{E}^{\mathbb{P}}[\mathbb{E}^{\mathbb{P}}[w(\mathbf{X}_T^{\leftarrow})|\mathcal{F}_t]|\mathcal{F}_s] = \mathbb{E}^{\mathbb{P}}[w(\mathbf{X}_T^{\leftarrow})|\mathcal{F}_s] = M_s,$$

where the second equality involves the tower property of conditional expectation, and the fact that $\mathcal{F}_s \subseteq \mathcal{F}_t$. Therefore, M_t is a martingale.
Part 2. Stochastic dynamics. Applying Itô's formula to $h^*(t, \mathbf{X}_t^{\leftarrow})$ yields

$$\begin{aligned}\mathrm{d}h^*(t, \mathbf{X}_t^{\leftarrow}) &= \partial_t h^*(t, \mathbf{X}_t^{\leftarrow})\, \mathrm{d}t + \nabla h^*(t, \mathbf{X}_t^{\leftarrow})^\top \mathrm{d}\mathbf{X}_t^{\leftarrow} + \frac{1}{2}(\mathrm{d}\mathbf{X}_t^{\leftarrow})^\top \nabla^2 h^*(t, \mathbf{X}_t^{\leftarrow})\, \mathrm{d}\mathbf{X}_t^{\leftarrow} \\ &= \nabla h^*(t, \mathbf{X}_t^{\leftarrow})^\top \sqrt{2}\, \mathrm{d}\mathbf{B}_t,\end{aligned}$$

where the last equality holds from the fact that martingale has zero drift. This completes the proof. □

Proposition 3.1. *The Radon-Nikodym derivative process L_t, as defined in (3.5), admits the following representation:*

$$L_t = \mathbb{E}^{\mathbb{P}}\Big[\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\Big|\mathcal{F}_t\Big] = \frac{h^*(t, \mathbf{X}_t^{\leftarrow})}{\mathbb{E}^{\mathbb{P}}[w(\mathbf{X}_T^{\leftarrow})]},$$

where $h^* : [0, T] \times \mathbb{R}^d \to \mathbb{R}$, referred to as Doob's h-function, is defined as the conditional expectation of the terminal weight:

$$h^*(t, \mathbf{x}) := \mathbb{E}^{\mathbb{P}}[w(\mathbf{X}_T^{\leftarrow}) \mid \mathbf{X}_t^{\leftarrow} = \mathbf{x}].$$

Furthermore, the log-likelihood ratio satisfies the following SDE:

$$\mathrm{d}(\log L_t) = \nabla \log h^*(t, \mathbf{X}_t^{\leftarrow})^\top \sqrt{2}\,\mathrm{d}\mathbf{B}_t - \|\nabla \log h^*(t, \mathbf{X}_t^{\leftarrow})\|_2^2\,\mathrm{d}t.$$

Proof of Proposition 3.1. The proof is divided into three parts.
Part 1. The equivalent definition of L_t. In this part, we aim to show the conditional expectation in Proposition 3.1 is identical to the likelihood L_t defined as (3.5). Define

$$\bar{L}_t = \mathbb{E}^{\mathbb{P}}\Big[\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\Big|\mathcal{F}_t\Big].$$

For each event $A \in \mathcal{F}_t \subseteq \mathcal{F}_T$, we have

$$\mathbb{Q}(A) = \int_A \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\,\mathrm{d}\mathbb{P} = \int_A \mathbb{E}\Big[\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\Big|\mathcal{F}_t\Big]\,\mathrm{d}\mathbb{P} = \int_A \bar{L}_t\,\mathrm{d}\mathbb{P},$$

where the second equality is due to $A \in \mathcal{F}_t$. This means $\bar{L}_t$ acts as the density for the measure $\mathbb{Q}$ with respect to $\mathbb{P}$ when restricting to the σ-algebra $\mathcal{F}_t$. Thus $L_t \equiv \bar{L}_t$ for each $t \in (0, T)$.
Part 2. The expression of Doob's h-transform. It is straightforward that

$$L_t = \mathbb{E}^{\mathbb{P}}\Big[\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\Big|\mathcal{F}_t\Big] = \frac{\mathbb{E}^{\mathbb{P}}[w(\mathbf{X}_T^{\leftarrow})|\mathcal{F}_t]}{\mathbb{E}^{\mathbb{P}}[w(\mathbf{X}_T^{\leftarrow})]} = \frac{\mathbb{E}^{\mathbb{P}}[w(\mathbf{X}_T^{\leftarrow})|\mathbf{X}_t^{\leftarrow}]}{\mathbb{E}^{\mathbb{P}}[w(\mathbf{X}_T^{\leftarrow})]} = \frac{h^*(t, \mathbf{X}_t^{\leftarrow})}{\mathbb{E}^{\mathbb{P}}[w(\mathbf{X}_T^{\leftarrow})]},$$

where the second equality follows from (3.6), the third equality involves the Markovity of the diffusion process $\mathbf{X}_t^{\leftarrow}$, and the last equality is due to the definition of the h-function (3.7).
Part 3. The stochastic dynamics of log-likelihood. We can now derive the dynamics of the likelihood L_t. Letting $Z = \mathbb{E}^{\mathbb{P}}[w(\mathbf{X}_T^{\leftarrow})]$, we have:

$$\begin{aligned}
\mathrm{d}L_t &= \frac{1}{Z}\,\mathrm{d}h^*(t, \mathbf{X}_t^{\leftarrow}) \\
&= \frac{1}{Z}(\nabla h^*(t, \mathbf{X}_t^{\leftarrow}))^\top \sqrt{2}\,\mathrm{d}\mathbf{B}_t \\
&= \frac{h^*(t, \mathbf{X}_t^{\leftarrow})}{Z}\frac{\nabla h^*(t, \mathbf{X}_t^{\leftarrow})^\top}{h^*(t, \mathbf{X}_t^{\leftarrow})}\sqrt{2}\,\mathrm{d}\mathbf{B}_t \\
&= L_t(\nabla \log h^*(t, \mathbf{X}_t^{\leftarrow}))^\top \sqrt{2}\,\mathrm{d}\mathbf{B}_t, \qquad \text{(A.1)}
\end{aligned}$$

where the second equality is due to Lemma A.1. Using Itô's formula for log-likelihood yields

$$\begin{aligned}
\mathrm{d}(\log L_t) &= \frac{\mathrm{d}L_t}{L_t} - \frac{1}{2}\Big\langle \frac{\mathrm{d}L_t}{L_t}, \frac{\mathrm{d}L_t}{L_t}\Big\rangle \\
&= \sqrt{2}\nabla \log h^*(t, \mathbf{X}_t^{\leftarrow})^\top\,\mathrm{d}\mathbf{B}_t - \frac{1}{2}\|\sqrt{2}\nabla \log h^*(t, \mathbf{X}_t^{\leftarrow})\|_2^2\,\mathrm{d}t \\
&= \sqrt{2}\nabla \log h^*(t, \mathbf{X}_t^{\leftarrow})^\top\,\mathrm{d}\mathbf{B}_t - \|\nabla \log h^*(t, \mathbf{X}_t^{\leftarrow})\|_2^2\,\mathrm{d}t,
\end{aligned}$$

where the second equality is due to (A.1). This completes the proof. □

Proposition 3.2. *Let the reference process* $\mathbf{X}_t^{\leftarrow}$ *satisfy the SDE* (2.3) *under the path measure* $\mathbb{P}$*, and let* $\mathbb{Q}$ *be the target path measure defined by* (3.6)*. Assume that the Novikov condition holds:*

$$\mathbb{E}^{\mathbb{P}}\Big[\exp\Big(\int_0^T \|\nabla \log h^*(s, \mathbf{X}_s^{\leftarrow})\|_2^2 \, \mathrm{d}s\Big)\Big] < \infty.$$

Define a process $(\widetilde{\mathbf{B}}_t)_{1\le t\le T}$ *by*

$$\mathrm{d}\widetilde{\mathbf{B}}_t = \mathrm{d}\mathbf{B}_t - \sqrt{2}\nabla \log h^*(t, \mathbf{X}_t^{\leftarrow}) \, \mathrm{d}t,$$

where $\mathbf{B}_t$ *is a Brownian motion under* $\mathbb{P}$*. Then* $\widetilde{\mathbf{B}}_t$ *is a standard Brownian motion under* $\mathbb{Q}$*. Further, under the path measure* $\mathbb{Q}$*, the reference process* $\mathbf{X}_t^{\leftarrow}$ *in* (2.3) *evolves according to:*

$$\mathrm{d}\mathbf{X}_t^{\leftarrow} = (\mathbf{X}_t^{\leftarrow} + 2\nabla \log p_{T-t}(\mathbf{X}_t^{\leftarrow}) + 2\nabla \log h^*(t, \mathbf{X}_t^{\leftarrow})) \, \mathrm{d}t + \sqrt{2} \, \mathrm{d}\widetilde{\mathbf{B}}_t.$$

Proof of Proposition 3.2. The derivation proceeds in two steps.
Step 1. The Brownian motion under the target path measure $\mathbb{Q}$. Using Proposition 3.1 and noting that $L_0 \equiv 1$ as (3.6), we have

$$L_t = \exp\Big(\int_0^t \nabla \log h^*(s, \mathbf{X}_s^{\leftarrow})^\top \sqrt{2} \, \mathrm{d}\mathbf{B}_s - \|\nabla \log h^*(s, \mathbf{X}_s^{\leftarrow})\|_2^2 \, \mathrm{d}s\Big).$$

According to Karatzas and Shreve (1998, Corollary 5.13), the Novikov condition (3.8) implies that L_t is a martingale. Then applying Girsanov's theorem (Øksendal, 2003, Theorem 8.6.6) yields that $\widetilde{\mathbf{B}}_t$ is a Brownian motion under $\mathbb{Q}$.
Step 2. Dynamics under the target path measure. Recall that under the reference path measure $\mathbb{P}$, the process evolves as

$$\mathrm{d}\mathbf{X}_t^{\leftarrow} = (\mathbf{X}_t^{\leftarrow} + 2\nabla \log p_{T-t}(\mathbf{X}_t^{\leftarrow})) \, \mathrm{d}t + \sqrt{2} \, \mathrm{d}\mathbf{B}_t,$$

where $\mathbf{B}_t$ is a $\mathbb{P}$-Brownian motion. Substituting the relationship (3.9) into this SDE implies

$$\begin{aligned}\mathrm{d}\mathbf{X}_t^{\leftarrow} &= (\mathbf{X}_t^{\leftarrow} + 2\nabla \log p_{T-t}(\mathbf{X}_t^{\leftarrow})) \, \mathrm{d}t + \sqrt{2}(\, \mathrm{d}\widetilde{\mathbf{B}}_t + \sqrt{2}\nabla \log h^*(t, \mathbf{X}_t^{\leftarrow}) \, \mathrm{d}t) \\ &= (\mathbf{X}_t^{\leftarrow} + 2\nabla \log p_{T-t}(\mathbf{X}_t^{\leftarrow}) + 2\nabla \log h^*(t, \mathbf{X}_t^{\leftarrow})) \, \mathrm{d}t + \sqrt{2} \, \mathrm{d}\widetilde{\mathbf{B}}_t,\end{aligned}$$

which is the dynamics of $\mathbf{X}_t^{\leftarrow}$ under the target path measure $\mathbb{Q}$. This completes the proof. ☐

B Derivations in Section 4

Proposition 4.1. *For every* $t \in (0, T)$*, the Doob's h-function* h_t^* *in* (3.7) *minimizes the implicit Doob's matching loss* (4.1)*. Further,*

$$\mathcal{J}_t(h_t) = \mathbb{E}^{\mathbb{P}}[\|h_t(\mathbf{X}_t^{\leftarrow}) - h_t^*(\mathbf{X}_t^{\leftarrow})\|_2^2] + V_t^2,$$

where $V_t^2 := \mathbb{E}^{\mathbb{P}}[\mathrm{Var}(w(\mathbf{X}_T^{\leftarrow})|\mathbf{X}_t^{\leftarrow})]$ *is a constant independent of* h_t*.*

Proof of Proposition 4.1. By a direct calculation, we have

$$\begin{aligned}\mathcal{J}_t(h_t) &= \mathbb{E}^{\mathbb{P}}[\|h_t(\mathbf{X}_t^{\leftarrow}) - w(\mathbf{X}_T^{\leftarrow})\|_2^2] \\ &= \mathbb{E}^{\mathbb{P}}[\|h_t(\mathbf{X}_t^{\leftarrow}) - h_t^*(\mathbf{X}_t^{\leftarrow}) + h_t^*(\mathbf{X}_t^{\leftarrow}) - w(\mathbf{X}_T^{\leftarrow})\|_2^2] \\ &= \mathbb{E}^{\mathbb{P}}[\|h_t(\mathbf{X}_t^{\leftarrow}) - h_t^*(\mathbf{X}_t^{\leftarrow})\|_2^2] + \mathbb{E}^{\mathbb{P}}[\|h_t^*(\mathbf{X}_t^{\leftarrow}) - w(\mathbf{X}_T^{\leftarrow})\|_2^2] \\ &\quad + 2\mathbb{E}^{\mathbb{P}}[\langle h_t(\mathbf{X}_t^{\leftarrow}) - h_t^*(\mathbf{X}_t^{\leftarrow}), h_t^*(\mathbf{X}_t^{\leftarrow}) - w(\mathbf{X}_T^{\leftarrow})\rangle],\end{aligned} \tag{B.1}$$

where h_t^* is defined as (3.7). For the second term in (B.1), we have

$$\begin{aligned}
&\mathbb{E}^{\mathbb{P}}[\|h_t^*(\mathbf{X}_t^{\leftarrow}) - w(\mathbf{X}_T^{\leftarrow})\|_2^2] \\
&= \mathbb{E}^{\mathbb{P}}[\|\mathbb{E}^{\mathbb{P}}[w(\mathbf{X}_T^{\leftarrow})|\mathbf{X}_t^{\leftarrow}] - w(\mathbf{X}_T^{\leftarrow})\|_2^2] \\
&= \mathbb{E}^{\mathbb{P}}\mathbb{E}^{\mathbb{P}}[\|\mathbb{E}^{\mathbb{P}}[w(\mathbf{X}_T^{\leftarrow})|\mathbf{X}_t^{\leftarrow}] - w(\mathbf{X}_T^{\leftarrow})\|_2^2|\mathbf{X}_t^{\leftarrow}] \\
&= \mathbb{E}^{\mathbb{P}}[\mathrm{Var}(w(\mathbf{X}_T^{\leftarrow})|\mathbf{X}_t^{\leftarrow})],
\end{aligned} \tag{B.2}$$

where the first equality holds from the definition of h_t^* (3.7), and the second equality is due to the law of the total expectation. For the third term in (B.1), we find

$$\begin{aligned}
&\mathbb{E}^{\mathbb{P}}[\langle h_t(\mathbf{X}_t^{\leftarrow}) - h_t^*(\mathbf{X}_t^{\leftarrow}), h_t^*(\mathbf{X}_t^{\leftarrow}) - w(\mathbf{X}_T^{\leftarrow})\rangle] \\
&= \mathbb{E}^{\mathbb{P}}[\langle h_t(\mathbf{X}_t^{\leftarrow}) - \mathbb{E}^{\mathbb{P}}[w(\mathbf{X}_T^{\leftarrow})|\mathbf{X}_t^{\leftarrow}], \mathbb{E}^{\mathbb{P}}[w(\mathbf{X}_T^{\leftarrow})|\mathbf{X}_t^{\leftarrow}] - w(\mathbf{X}_T^{\leftarrow})\rangle] \\
&= \mathbb{E}^{\mathbb{P}}[\langle h_t(\mathbf{X}_t^{\leftarrow}) - \mathbb{E}^{\mathbb{P}}[w(\mathbf{X}_T^{\leftarrow})|\mathbf{X}_t^{\leftarrow}], \mathbb{E}^{\mathbb{P}}[w(\mathbf{X}_T^{\leftarrow})|\mathbf{X}_t^{\leftarrow}] - \mathbb{E}^{\mathbb{P}}[w(\mathbf{X}_T^{\leftarrow})|\mathbf{X}_t^{\leftarrow}]\rangle] \\
&= 0,
\end{aligned} \tag{B.3}$$

where the first equality holds from the definition of h_t^* (3.7), and the second equality is due to the law of the total expectation. Substituting (B.2) and (B.3) into (B.1) completes the proof. □

Lemma B.1. *Suppose Assumptions 1 and 2 hold. Assume that $v_t \in H^1(p_{T-t})$. Then*

$$-(\nabla h_t^*, \nabla v_t)_{L^2(p_{T-t})} = (\Delta h_t^* + \nabla h_t^* \cdot \nabla \log p_{T-t}, v_t)_{L^2(p_{T-t})}.$$

Proof of Lemma B.1. We first construct a sequence of cut-off functions $\{\psi_k\}_{k=1}^\infty \subseteq C_0^\infty(\mathbb{R}^d)$, satisfying

(i) $\psi_k(\mathbf{x}) = 1$ for $\|\mathbf{x}\|_2 \le k$,
(ii) $\psi_k(\mathbf{x}) = 0$ for $\|\mathbf{x}\|_2 \ge 2k$,
(iii) $\psi_k(\mathbf{x}) \in (0,1)$ for $\mathbf{x} \in \mathbb{R}^d$, and
(iv) $\|\nabla \psi_k(\mathbf{x})\|_2 \le Ck^{-1}$ for some constant C independent of $\mathbf{x}$ and k.

See Brezis (2011, Theorem 8.7) for a detailed construction of such cut-off functions. Then we focus on the compactly supported approximations $\{\psi_k v_t\}_{k=1}^\infty$:

$$\begin{aligned}
&-\langle \nabla h_t^*, \nabla(\psi_k v_t)\rangle_{L^2(p_{T-t})} \\
&= -\int_{\mathbb{R}^d} \langle \nabla h_t^*(\mathbf{x}), \nabla(\psi_k v_t)(\mathbf{x})\rangle p_{T-t}(\mathbf{x})\,\mathrm{d}\mathbf{x} \\
&= -\int_{\mathbb{R}^d} \nabla\cdot(\nabla h_t^*(\mathbf{x}) p_{T-t}(\mathbf{x})(\psi_k v_t)(\mathbf{x}))\,\mathrm{d}\mathbf{x} + \int_{\mathbb{R}^d} \nabla\cdot(\nabla h_t^*(\mathbf{x}) p_{T-t}(\mathbf{x}))(\psi_k v_t)(\mathbf{x})\,\mathrm{d}\mathbf{x} \\
&= \int_{\mathbb{R}^d} \nabla\cdot(\nabla h_t^*(\mathbf{x}) p_{T-t}(\mathbf{x}))(\psi_k v_t)(\mathbf{x})\,\mathrm{d}\mathbf{x},
\end{aligned} \tag{B.4}$$

where we used the Gauss's divergence theorem and the fact that $\psi_k v_t \in H_0^1(B(\mathbf{0},k))$ with $B(\mathbf{0},k) := \{\mathbf{x} : \|\mathbf{x}\|_2 \le k\}$. For the left-hand side of (B.4), we have

$$\begin{aligned}
&\int_{\mathbb{R}^d} \langle \nabla h_t^*(\mathbf{x}), \nabla(\psi_k v_t)(\mathbf{x})\rangle p_{T-t}(\mathbf{x})\,\mathrm{d}\mathbf{x} \\
&= \int_{\mathbb{R}^d} \psi_k(\mathbf{x})\langle \nabla h_t^*(\mathbf{x}), \nabla v_t(\mathbf{x})\rangle p_{T-t}(\mathbf{x})\,\mathrm{d}\mathbf{x} + \int_{\mathbb{R}^d} \langle \nabla h_t^*(\mathbf{x}), \nabla \psi_k(\mathbf{x})\rangle v_t(\mathbf{x}) p_{T-t}(\mathbf{x})\,\mathrm{d}\mathbf{x}.
\end{aligned} \tag{B.5}$$

For the second term in (B.5), we have

$$
\begin{aligned}
&\Big| \int_{\mathbb{R}^d} \langle \nabla h_t^*(\mathbf{x}), \nabla \psi_k(\mathbf{x}) \rangle v_t(\mathbf{x}) p_{T-t}(\mathbf{x}) \,\mathrm{d}\mathbf{x} \Big| \\
&\leq \int_{\mathbb{R}^d} \|\nabla h_t^*(\mathbf{x})\|_2 \|\nabla \psi_k(\mathbf{x})\|_2 |v_t(\mathbf{x})| p_{T-t}(\mathbf{x}) \,\mathrm{d}\mathbf{x} \\
&\leq \frac{C}{k} \int_{\mathbb{R}^d} \|\nabla h_t^*(\mathbf{x})\|_2 |v_t(\mathbf{x})| p_{T-t}(\mathbf{x}) \,\mathrm{d}\mathbf{x} \\
&\leq \frac{C}{k} \|\nabla h_t^*\|_{L^2(p_{T-t})} \|v_t\|_{L^2(p_{T-t})},
\end{aligned}
$$

where the second inequality holds from the definition of the cut-off function, and the last inequality is due to Cauchy-Schwarz inequality and the fact that $h_t^* \in H^1(p_{T-t})$, which is a direct conclusion of Lemmas C.1 and C.4. Taking limitation with respect to $k \to \infty$ and using Lebesgue's dominated convergence theorem yields

$$
\lim_{k \to \infty} \int_{\mathbb{R}^d} \langle \nabla h_t^*(\mathbf{x}), \nabla \psi_k(\mathbf{x}) \rangle v_t(\mathbf{x}) p_{T-t}(\mathbf{x}) \,\mathrm{d}\mathbf{x} = 0. \tag{B.6}
$$

Combining (B.4), (B.5), and (B.6) and taking limitation with respect to $k \to \infty$ completes the proof. □

Proposition 4.2. *Let $\lambda > 0$, h_t^* be the Doob's h-function defined as (3.7), and h_t^λ be the minimizer of $\mathcal{J}_t^\lambda$ defined as (4.3). Then $h_t^* \in H^2(p_{T-t})$, and*

$$
\begin{aligned}
\|h_t^\lambda - h_t^*\|_{L^2(p_{T-t})}^2 &\leq \lambda^2 \|\Delta h_t^* + \nabla h_t^* \cdot \nabla \log p_{T-t}\|_{L^2(p_{T-t})}^2, \\
\|\nabla h_t^\lambda - \nabla h_t^*\|_{L^2(p_{T-t})}^2 &\leq \lambda \|\Delta h_t^* + \nabla h_t^* \cdot \nabla \log p_{T-t}\|_{L^2(p_{T-t})}^2.
\end{aligned}
$$

Proof of Proposition 4.2. First, $h_t^* \in H^2(p_{T-t})$ is a direct conclusion of Lemmas C.1, C.4, and C.5. It remains to prove two inequalities. Using Proposition 4.1 and (4.3), we have

$$
\mathcal{J}_t^\lambda(h_t) = \|h_t - h_t^*\|_{L^2(p_{T-t})}^2 + V_t^2 + \lambda \|\nabla h_t\|_{L^2(p_{T-t})}^2
$$

Since h_t^λ is the minimizer of $\mathcal{J}_t^\lambda$, the methods of variation imply that for any $v_t \in H^1(p_{T-t})$,

$$
\delta \mathcal{J}_t^\lambda(h_t^\lambda, v_t) = \langle h_t^\lambda - h_t^*, v_t \rangle_{L^2(p_{T-t})} + \lambda \langle \nabla h_t^\lambda, \nabla v_t \rangle_{L^2(p_{T-t})} = 0,
$$

which implies

$$
\begin{aligned}
&\langle h_t^\lambda - h_t^*, v_t \rangle_{L^2(p_{T-t})} + \lambda \langle \nabla h_t^\lambda - \nabla h_t^*, \nabla v_t \rangle_{L^2(p_{T-t})} \\
&= -\lambda \langle \nabla h_t^*, \nabla v_t \rangle_{L^2(p_{T-t})} \\
&= \lambda \langle \Delta h_t^* + \nabla h_t^* \cdot \nabla \log p_{T-t}, v_t \rangle_{L^2(p_{T-t})},
\end{aligned}
$$

where the last equality invokes Lemma B.1. Substituting $v_t := h_t^\lambda - h_t^*$ yields

$$
\begin{aligned}
&\|h_t^\lambda - h_t^*\|_{L^2(p_{T-t})}^2 + \lambda \|\nabla h_t^\lambda - \nabla h_t^*\|_{L^2(p_{T-t})}^2 \\
&= \lambda \langle \Delta h_t^* + \nabla h_t^* \cdot \nabla \log p_{T-t}, h_t^\lambda - h_t^* \rangle_{L^2(p_{T-t})} \\
&\leq \lambda \|\Delta h_t^* + \nabla h_t^* \cdot \nabla \log p_{T-t}\|_{L^2(p_{T-t})} \|h_t^\lambda - h_t^*\|_{L^2(p_{T-t})}, \qquad \text{(B.7)}
\end{aligned}
$$

where the last inequality is due to Cauchy-Schwarz inequality. A direct conclusion is

$$\|h_t^\lambda - h_t^*\|^2_{L^2(p_{T-t})} \le \lambda^2 \|\Delta h_t^* + \nabla h_t^* \cdot \nabla \log p_{T-t}\|^2_{L^2(p_{T-t})}.$$

Then plugging this equality into (B.7) yields

$$\|\nabla h_t^\lambda - \nabla h_t^*\|^2_{L^2(p_{T-t})} \le \lambda \|\Delta h_t^* + \nabla h_t^* \cdot \nabla \log p_{T-t}\|^2_{L^2(p_{T-t})},$$

which completes the proof. □

Proposition 4.3. *Let $\lambda > 0$, and h_t^λ be the minimizer of $\mathcal{J}_t^\lambda$ defined as (4.3). Then for any $h_t \in H^1(p_{T-t})$, we have*

$$\frac{1}{\max\{\lambda, 1\}}\{\mathcal{J}_t^\lambda(h_t) - \mathcal{J}_t^\lambda(h_t^\lambda)\} \le \|h_t - h_t^\lambda\|^2_{H^1(p_{T-t})} \le \frac{1}{\min\{\lambda, 1\}}\{\mathcal{J}_t^\lambda(h_t) - \mathcal{J}_t^\lambda(h_t^\lambda)\}.$$

Proof of Proposition 4.3. Using Proposition 4.1 and (4.3), we have

$$\mathcal{J}_t^\lambda(h_t) = \|h_t - h_t^*\|^2_{L^2(p_{T-t})} + V_t^2 + \lambda \|\nabla h_t\|^2_{L^2(p_{T-t})}.$$

Since h_t^λ is the minimizer of $\mathcal{J}_t^\lambda$, the methods of variation imply

$$\delta \mathcal{J}_t^\lambda(h_t^\lambda, v_t) = \langle h_t^\lambda - h_t^*, v_t\rangle_{L^2(p_{T-t})} + \lambda \langle \nabla h_t^\lambda, \nabla v_t\rangle_{L^2(p_{T-t})} = 0, \tag{B.8}$$

for any $v_t \in H^1(p_{T-t})$. A direct calculation yields

$$\begin{aligned}
\mathcal{J}_t^\lambda(h_t) &= \mathcal{J}_t^\lambda(h_t - h_t^\lambda + h_t^\lambda) \\
&= \|h_t - h_t^\lambda + h_t^\lambda - h_t^*\|^2_{L^2(p_{T-t})} + V_t^2 + \lambda \|\nabla h_t - \nabla h_t^\lambda + \nabla h_t^\lambda\|^2_{L^2(p_{T-t})} \\
&= \mathcal{J}_t^\lambda(h_t^\lambda) + \|h_t - h_t^\lambda\|^2_{L^2(p_{T-t})} + \lambda \|\nabla h_t - \nabla h_t^\lambda\|^2_{L^2(p_{T-t})} \\
&\quad + 2\langle h_t - h_t^\lambda, h_t^\lambda - h_t^*\rangle_{L^2(p_{T-t})} + 2\lambda \langle \nabla h_t - \nabla h_t^\lambda, \nabla h_t^\lambda\rangle_{L^2(p_{T-t})} \\
&= \mathcal{J}_t^\lambda(h_t^\lambda) + \|h_t - h_t^\lambda\|^2_{L^2(p_{T-t})} + \lambda \|\nabla h_t - \nabla h_t^\lambda\|^2_{L^2(p_{T-t})},
\end{aligned}$$

where the last equality holds from (B.8). This completes the proof. □

C Derivations in Section 5.1

Lemma C.1. *Suppose Assumption 2 holds. Then for all $t \in (0, T)$,*

$$\underline{B} \le h_t^*(\mathbf{x}) \le \bar{B}, \quad \mathbf{x} \in \mathbb{R}^d.$$

Proof of Lemma C.1. A direct conclusion of Assumption 2 and the definition of Doob's h-function h_t^* (3.7). □

Lemma C.2 (Tweedie's formula)**.** *Let $t \in (0, T)$, and let $\mathbf{X}_t$ be defined as (2.1). Then*

$$\nabla \log p_t(\mathbf{x}) + \frac{\mathbf{x}}{\sigma_t^2} = \frac{\mu_t}{\sigma_t^2}\mathbb{E}[\mathbf{X}_0 | \mathbf{X}_t = \mathbf{x}], \quad \mathbf{x} \in \mathbb{R}^d.$$

Proof of Lemma C.2. It is straightforward that

$$\begin{aligned}\nabla \log p_t(\mathbf{x}) &= \frac{\nabla p_t(\mathbf{x})}{p_t(\mathbf{x})} \\ &= \frac{1}{p_t(\mathbf{x})} \int \nabla_{\mathbf{x}} \varphi_d(\mathbf{x}; \mu_t \mathbf{x}_0, \sigma_t^2 \mathbf{I}_d) p_0(\mathbf{x}_0) \, \mathrm{d}\mathbf{x}_0 \\ &= -\frac{1}{p_t(\mathbf{x})} \int \Big(\frac{\mathbf{x} - \mu_t \mathbf{x}_0}{\sigma_t^2} \Big) \varphi_d(\mathbf{x}; \mu_t \mathbf{x}_0, \sigma_t^2 \mathbf{I}_d) p_0(\mathbf{x}_0) \, \mathrm{d}\mathbf{x}_0 \\ &= -\frac{1}{p_t(\mathbf{x})} \frac{\mathbf{x}}{\sigma_t^2} \int \varphi_d(\mathbf{x}; \mu_t \mathbf{x}_0, \sigma_t^2 \mathbf{I}_d) p_0(\mathbf{x}_0) \, \mathrm{d}\mathbf{x}_0 \\ &\quad + \frac{\mu_t}{\sigma_t^2} \int \mathbf{x}_0 \varphi_d(\mathbf{x}; \mu_t \mathbf{x}_0, \sigma_t^2 \mathbf{I}_d) \frac{p_0(\mathbf{x}_0)}{p_t(\mathbf{x})} \, \mathrm{d}\mathbf{x}_0 \\ &= -\frac{\mathbf{x}}{\sigma_t^2} + \frac{\mu_t}{\sigma_t^2} \mathbb{E}[\mathbf{X}_0 | \mathbf{X}_t = \mathbf{x}],\end{aligned}$$

where the second equality is due to (2.2), and last equality invokes the Bayes' rule. This completes the proof. ☐

Lemma C.3. *Let $g : \mathbb{R}^d \to \mathbb{R}^d$ be an integrable function. Let $t \in (0, T)$, and let $\mathbf{X}_t$ be defined as (2.1). Then for each $\mathbf{x} \in \mathbb{R}^d$,*

$$\nabla_{\mathbf{x}} \mathbb{E}[g(\mathbf{X}_0) | \mathbf{X}_t = \mathbf{x}] = \frac{\mu_t}{\sigma_t^2} \mathrm{Cov}(\mathbf{X}_0, g(\mathbf{X}_0) | \mathbf{X}_t = \mathbf{x}),$$

where the k-th entry of $\mathrm{Cov}(\mathbf{X}_0, g(\mathbf{X}_0) | \mathbf{X}_t = \mathbf{x})$ *is defined as* $\mathrm{Cov}(X_{0,k}, g(\mathbf{X}_0) | \mathbf{X}_t = \mathbf{x})$ *with* $\mathbf{X}_0 = (X_{0,1}, \ldots, X_{0,d})$.

Proof of Lemma C.3. According to Bayes' rule, we have

$$\mathbb{E}[g(\mathbf{X}_0) | \mathbf{X}_t = \mathbf{x}] = \frac{1}{p_t(\mathbf{x})} \int g(\mathbf{x}_0) \varphi_d(\mathbf{x}; \mu_t \mathbf{x}_0, \sigma_t^2 \mathbf{I}_d) p_0(\mathbf{x}_0) \, \mathrm{d}\mathbf{x}_0.$$

Taking gradient with respect to $\mathbf{x}$ on both sides of the equality yields

$$\begin{aligned}\nabla_{\mathbf{x}} \mathbb{E}[g(\mathbf{X}_0) | \mathbf{X}_t = \mathbf{x}] &= \frac{1}{p_t(\mathbf{x})} \int g(\mathbf{x}_0) \nabla_{\mathbf{x}} \varphi_d(\mathbf{x}; \mu_t \mathbf{x}_0, \sigma_t^2 \mathbf{I}_d) p_0(\mathbf{x}_0) \, \mathrm{d}\mathbf{x}_0 \\ &\quad - \frac{\nabla p_t(\mathbf{x})}{p_t^2(\mathbf{x})} \int g(\mathbf{x}_0) \varphi_d(\mathbf{x}; \mu_t \mathbf{x}_0, \sigma_t^2 \mathbf{I}_d) p_0(\mathbf{x}_0) \, \mathrm{d}\mathbf{x}_0 \\ &= -\frac{1}{p_t(\mathbf{x})} \int g(\mathbf{x}_0) \Big(\frac{\mathbf{x} - \mu_t \mathbf{x}_0}{\sigma_t^2} \Big) \varphi_d(\mathbf{x}; \mu_t \mathbf{x}_0, \sigma_t^2 \mathbf{I}_d) p_0(\mathbf{x}_0) \, \mathrm{d}\mathbf{x}_0 \\ &\quad - \frac{\nabla \log p_t(\mathbf{x})}{p_t(\mathbf{x})} \int g(\mathbf{x}_0) \varphi_d(\mathbf{x}; \mu_t \mathbf{x}_0, \sigma_t^2 \mathbf{I}_d) p_0(\mathbf{x}_0) \, \mathrm{d}\mathbf{x}_0 \\ &= -\Big(\nabla \log p_t(\mathbf{x}) + \frac{\mathbf{x}}{\sigma_t^2} \Big) \int g(\mathbf{x}_0) \varphi_d(\mathbf{x}; \mu_t \mathbf{x}_0, \sigma_t^2 \mathbf{I}_d) \frac{p_0(\mathbf{x}_0)}{p_t(\mathbf{x})} \, \mathrm{d}\mathbf{x}_0 \\ &\quad + \frac{\mu_t}{\sigma_t^2} \int \mathbf{x}_0 g(\mathbf{x}_0) \varphi_d(\mathbf{x}; \mu_t \mathbf{x}_0, \sigma_t^2 \mathbf{I}_d) \frac{p_0(\mathbf{x}_0)}{p_t(\mathbf{x})} \, \mathrm{d}\mathbf{x}_0 \\ &= \frac{\mu_t}{\sigma_t^2} (-\mathbb{E}[\mathbf{X}_0 | \mathbf{X}_t = \mathbf{x}] \mathbb{E}[g(\mathbf{X}_0) | \mathbf{X}_t = \mathbf{x}] + \mathbb{E}[\mathbf{X}_0 g(\mathbf{X}_0) | \mathbf{X}_t = \mathbf{x}]) \\ &= \frac{\mu_t}{\sigma_t^2} \mathrm{Cov}(\mathbf{X}_0, g(\mathbf{X}_0) | \mathbf{X}_t = \mathbf{x}),\end{aligned}$$

where the fourth equality follows from Lemma C.2 and the Bayes' rule. Here the k-th entry of the conditional covariance $\mathrm{Cov}(\mathbf{X}_0, g(\mathbf{X}_0)|\mathbf{X}_t = \mathbf{x})$ is defined as $\mathrm{Cov}(X_{0,k}, g(\mathbf{X}_0)|\mathbf{X}_t = \mathbf{x})$, where $\mathbf{X}_0 = (X_{0,1}, \dots, X_{0,d})$. This completes the proof. □

Lemma C.4. *Suppose Assumptions 1 and 2 hold. Then for each* $\mathbf{x} \in \mathbb{R}^d$ *and* $t > 0$,

$$\max_{1 \le k \le d} |D_k h_t^*(\mathbf{x})| \le \frac{2\bar{B}}{\sigma_{T-t}^2},$$

where D_k *denote the differential operator with respect to the* k*-th entry of* $\mathbf{x}$.

Proof of Lemma C.4. According to the definition of Doob's h-function h_t^* (3.7) and the property of the time-reversal process (2.3), we have

$$h^*(t, \mathbf{x}) := \mathbb{E}^{\mathbb{P}}[w(\mathbf{X}_T^{\leftarrow})|\mathbf{X}_t^{\leftarrow} = \mathbf{x}] = \mathbb{E}[w(\mathbf{X}_0)|\mathbf{X}_{T-t} = \mathbf{x}],$$

where the second expectation is with respect to the path measure of the forward process (2.1). Then it follows from Lemma C.3 that

$$\nabla h_t^*(\mathbf{x}) = \frac{\mu_{T-t}}{\sigma_{T-t}^2} \mathrm{Cov}(\mathbf{X}_0, w(\mathbf{X}_0)|\mathbf{X}_{T-t} = \mathbf{x}). \tag{C.1}$$

Then it follows from Assumptions 1 and 2 that for each $\mathbf{x} \in \mathbb{R}^d$,

$$\|\mathrm{Cov}(\mathbf{X}_0, w(\mathbf{X}_0)|\mathbf{X}_{T-t} = \mathbf{x})\|_\infty = \max_{1 \le k \le d} \mathrm{Cov}(X_{0,k}, w(\mathbf{X}_0)|\mathbf{X}_{T-t} = \mathbf{x}) \le 2\bar{B}. \tag{C.2}$$

Substituting (C.2) into (C.1) yields

$$\|\nabla h_t^*(\mathbf{x})\|_\infty = \frac{\mu_t}{\sigma_t^2} \|\mathrm{Cov}(\mathbf{X}_0, w(\mathbf{X}_0)|\mathbf{X}_{T-t} = \mathbf{x})\|_\infty \le \frac{2\bar{B}}{\sigma_{T-t}^2},$$

where we used the fact that $\mu_t = \exp(-t) < 1$. This completes the proof. □

Lemma C.5. *Suppose Assumptions 1 and 2 hold. Then for each* $\mathbf{x} \in \mathbb{R}^d$ *and* $t > 0$,

$$|D_{k\ell}^2 h_t^*(\mathbf{x})| \le \frac{6\bar{B}}{\sigma_{T-t}^4},$$

where $D_{k\ell}^2$ *denote the second-order differential operator with respect to* k*-th and* ℓ*-th entry.*

Proof of Lemma C.5. Taking derivative with respect to the ℓ-th entry of $\mathbf{x}$ on both sides of (C.1) implies

$$D_{k\ell}^2 h_t^*(\mathbf{x}) = \frac{\mu_{T-t}}{\sigma_{T-t}^2} D_\ell \mathrm{Cov}(X_{0,k}, w(\mathbf{X}_0)|\mathbf{X}_{T-t} = \mathbf{x}). \tag{C.3}$$

It remains to estimate the derivative of the conditional covariance. Indeed,

$$\begin{aligned}
&D_\ell \mathrm{Cov}(X_{0,k}, w(\mathbf{X}_0)|\mathbf{X}_{T-t} = \mathbf{x}) \\
&= D_\ell \mathbb{E}[X_{0,k}|\mathbf{X}_{T-t} = \mathbf{x}]\mathbb{E}[w(\mathbf{X}_0)|\mathbf{X}_{T-t} = \mathbf{x}] \\
&\quad + \mathbb{E}[X_{0,k}|\mathbf{X}_{T-t} = \mathbf{x}] D_\ell \mathbb{E}[w(\mathbf{X}_0)|\mathbf{X}_{T-t} = \mathbf{x}] \\
&\quad - D_\ell \mathbb{E}[X_{0,k} w(\mathbf{X}_0)|\mathbf{X}_{T-t} = \mathbf{x}] \\
&= \frac{\mu_{T-t}}{\sigma_{T-t}^2} \mathrm{Cov}(X_{0,\ell}, X_{0,k}|\mathbf{X}_t = \mathbf{x}) \mathbb{E}[w(\mathbf{X}_0)|\mathbf{X}_{T-t} = \mathbf{x}] \\
&\quad + \mathbb{E}[X_{0,k}|\mathbf{X}_{T-t} = \mathbf{x}] \frac{\mu_{T-t}}{\sigma_{T-t}^2} \mathrm{Cov}(w(\mathbf{X}_0), X_{0,\ell}|\mathbf{X}_{T-t} = \mathbf{x}) \\
&\quad - \frac{\mu_{T-t}}{\sigma_{T-t}^2} \mathrm{Cov}(X_{0,k} w(\mathbf{X}_0), X_{0,\ell}|\mathbf{X}_{T-t} = \mathbf{x}),
\end{aligned}$$

where the last equality holds from Lemma C.3. Concequently, for each $\mathbf{x} \in \mathbb{R}^d$ and $t > 0$,

$$|D_\ell \mathrm{Cov}(X_{0,k}, w(\mathbf{X}_0)|\mathbf{X}_{T-t} = \mathbf{x})| \leq \frac{6\mu_{T-t}\bar{B}}{\sigma_{T-t}^2}, \tag{C.4}$$

where we used Assumptions 1 and 2. Substituting (C.4) into (C.3) completes the proof. □

Proposition 5.1. *Suppose Assumptions 1 and 2 hold. Then for all $t \in (0, T)$ and $\mathbf{x} \in \mathbb{R}^d$, the following bounds hold:*

(i) $\underline{B} \leq h_t^*(\mathbf{x}) \leq \bar{B}$;

(ii) $\max_{1\leq k\leq d} |D_k h_t^*(\mathbf{x})| \leq 2\sigma_{T-t}^{-2}\bar{B}$; *and*

(iii) $\max_{1\leq k,\ell\leq d} |D_{k\ell}^2 h_t^*(\mathbf{x})| \leq 6\sigma_{T-t}^{-4}\bar{B}$,

where D_k and $D_{k\ell}^2$ denote the first-order and second-order partial derivatives with respect to the input coordinates, respectively.

Proof of Proposition 5.1. A direct conclusion of Lemmas C.1, C.4, and C.5. □

D Derivations in Section 5.2

D.1 Oracle inequality of variationally stable Doob's matching.

Lemma 5.2. *Suppose Assumptions 1 and 2 hold. Let $t \in (0, T)$ and let $\mathscr{H}_t$ be a hypothesis class. Let $\widehat{h}_t^\lambda$ be the gradient-regularized empirical risk minimizer defined as (4.6), and let h_t^* be the Doob's h-function defined as (3.7). Then the following inequalities hold:*

$$\begin{aligned}
\mathbb{E}\Big[\|\widehat{h}_t^\lambda - h_t^*\|_{L^2(p_{T-t})}^2\Big] &\lesssim \inf_{h_t\in\mathscr{H}_t}\Big\{\|h_t - h_t^*\|_{L^2(p_{T-t})}^2 + \lambda\|\nabla h_t - \nabla h_t^*\|_{L^2(p_{T-t})}^2\Big\} \\
&\quad + \bar{B}^2\Big(\frac{\mathrm{VCdim}(\mathscr{H}_t)}{n\log^{-1} n}\Big)^{\frac{1}{2}} + \frac{\lambda d\bar{B}^2}{\sigma_{T-t}^4}\Big(\frac{\mathrm{VCdim}(\nabla\mathscr{H}_t)}{n\log^{-1} n}\Big)^{\frac{1}{2}} + \frac{\lambda^2 d\bar{B}^2}{\sigma_{T-t}^8}, \\
\mathbb{E}\Big[\|\nabla\widehat{h}_t^\lambda - \nabla h_t^*\|_{L^2(p_{T-t})}^2\Big] &\lesssim \inf_{h_t\in\mathscr{H}_t}\Big\{\frac{1}{\lambda}\|h_t - h_t^*\|_{L^2(p_{T-t})}^2 + \|\nabla h_t - \nabla h_t^*\|_{L^2(p_{T-t})}^2\Big\} \\
&\quad + \frac{\bar{B}^2}{\lambda}\Big(\frac{\mathrm{VCdim}(\mathscr{H}_t)}{n\log^{-1} n}\Big)^{\frac{1}{2}} + \frac{d\bar{B}^2}{\sigma_{T-t}^4}\Big(\frac{\mathrm{VCdim}(\nabla\mathscr{H}_t)}{n\log^{-1} n}\Big)^{\frac{1}{2}} + \frac{\lambda d\bar{B}^2}{\sigma_{T-t}^8},
\end{aligned}$$

where the notation $\lesssim$ hides absolute constants.

Proof of Lemma 5.2. It follows from Proposition 4.3 and Lemma D.1 that

$$\begin{aligned}
&\mathbb{E}\Big[\|\widehat{h}_t^\lambda - h_t^\lambda\|_{L^2(p_{T-t})}^2\Big] + \lambda\mathbb{E}\Big[\|\nabla\widehat{h}_t^\lambda - \nabla h_t^\lambda\|_{L^2(p_{T-t})}^2\Big] \\
&= \mathbb{E}\Big[\mathcal{J}_t^\lambda(\widehat{h}_t^\lambda) - \mathcal{J}_t^\lambda(h_t^\lambda)\Big] \\
&\leq \inf_{h_t\in\mathscr{H}_t}\Big\{\|h_t - h_t^\lambda\|_{L^2(p_{T-t})}^2 + \lambda\|\nabla h_t - \nabla h_t^\lambda\|_{L^2(p_{T-t})}^2\Big\} \\
&\quad + 80\bar{B}^2\Big(\frac{\mathrm{VCdim}(\mathscr{H}_t)}{n\log^{-1} n}\Big)^{\frac{1}{2}} + 8\frac{\lambda d\bar{B}^2}{\sigma_{T-t}^4}\Big(\frac{\mathrm{VCdim}(\nabla\mathscr{H}_t)}{n\log^{-1} n}\Big)^{\frac{1}{2}} \\
&\leq \inf_{h_t\in\mathscr{H}_t}\Big\{2\|h_t - h_t^*\|_{L^2(p_{T-t})}^2 + 2\lambda\|\nabla h_t - \nabla h_t^*\|_{L^2(p_{T-t})}^2\Big\} \\
&\quad + 80\bar{B}^2\Big(\frac{\mathrm{VCdim}(\mathscr{H}_t)}{n\log^{-1} n}\Big)^{\frac{1}{2}} + 8\frac{\lambda d\bar{B}^2}{\sigma_{T-t}^4}\Big(\frac{\mathrm{VCdim}(\nabla\mathscr{H}_t)}{n\log^{-1} n}\Big)^{\frac{1}{2}} \\
&\quad + 2\|h_t^* - h_t^\lambda\|_{L^2(p_{T-t})}^2 + 2\lambda\|\nabla h_t^* - \nabla h_t^\lambda\|_{L^2(p_{T-t})}^2,
\end{aligned}$$

where the last inequality holds from the triangular inequality. Using the triangular inequality again, we have

$$
\begin{aligned}
&\mathbb{E}\Big[\|\widehat{h}_t^\lambda - h_t^*\|_{L^2(p_{T-t})}^2\Big] + \lambda\mathbb{E}\Big[\|\nabla\widehat{h}_t^\lambda - \nabla h_t^*\|_{L^2(p_{T-t})}^2\Big] \\
&\leq 2\mathbb{E}\Big[\|\widehat{h}_t^\lambda - h_t^\lambda\|_{L^2(p_{T-t})}^2\Big] + 2\lambda\mathbb{E}\Big[\|\nabla\widehat{h}_t^\lambda - \nabla h_t^\lambda\|_{L^2(p_{T-t})}^2\Big] \\
&\quad + 2\|h_t^* - h_t^\lambda\|_{L^2(p_{T-t})}^2 + 2\lambda\|\nabla h_t^* - \nabla h_t^\lambda\|_{L^2(p_{T-t})}^2 \\
&\leq \underbrace{\inf_{h_t\in\mathscr{H}_t}\Big\{4\|h_t - h_t^*\|_{L^2(p_{T-t})}^2 + 4\lambda\|\nabla h_t - \nabla h_t^*\|_{L^2(p_{T-t})}^2\Big\}}_{\text{approximation error}} \\
&\quad \underbrace{+ 160\bar{B}^2\Big(\frac{\mathrm{VCdim}(\mathscr{H}_t)}{n\log^{-1}n}\Big)^{\frac{1}{2}} + 16\frac{\lambda d\bar{B}^2}{\sigma_{T-t}^4}\Big(\frac{\mathrm{VCdim}(\nabla\mathscr{H}_t)}{n\log^{-1}n}\Big)^{\frac{1}{2}}}_{\text{generalization error}} \\
&\quad \underbrace{+ 6\|h_t^* - h_t^\lambda\|_{L^2(p_{T-t})}^2 + 6\lambda\|\nabla h_t^* - \nabla h_t^\lambda\|_{L^2(p_{T-t})}^2}_{\text{regularization gap}}.
\end{aligned}
\tag{D.1}
$$

Combining Proposition 4.2 and Lemma D.6 yields

$$
\begin{aligned}
\|h_t^* - h_t^\lambda\|_{L^2(p_{T-t})}^2 &\leq 144\lambda^2\frac{d\bar{B}^2}{\sigma_{T-t}^8}, \\
\|\nabla h_t^* - \nabla h_t^\lambda\|_{L^2(p_{T-t})}^2 &\leq 144\lambda\frac{d\bar{B}^2}{\sigma_{T-t}^8}.
\end{aligned}
\tag{D.2}
$$

Substituting (D.2) into (D.1) completes the proof. □

Lemma D.1. *Suppose Assumptions 1 and 2 hold. Let $t \in (0, T)$ and let $\mathscr{H}_t$ be a hypothesis class. Let $\widehat{h}_t^\lambda$ be the gradient-regularized empirical risk minimizer defined as (4.6). Then we have*

$$
\begin{aligned}
\mathbb{E}[\mathcal{J}_t^\lambda(\widehat{h}_t^\lambda) - \mathcal{J}_t^\lambda(h_t^\lambda)] \leq \inf_{h_t\in\mathscr{H}_t}&\Big\{\|h_t - h_t^\lambda\|_{L^2(p_{T-t})}^2 + \lambda\|\nabla h_t - \nabla h_t^\lambda\|_{L^2(p_{T-t})}^2\Big\} \\
&+ 80\bar{B}^2\Big(\frac{\mathrm{VCdim}(\mathscr{H}_t)}{n\log^{-1}n}\Big)^{\frac{1}{2}} + 8\frac{\lambda d\bar{B}^2}{\sigma_{T-t}^4}\Big(\frac{\mathrm{VCdim}(\nabla\mathscr{H}_t)}{n\log^{-1}n}\Big)^{\frac{1}{2}},
\end{aligned}
$$

where the notation $\lesssim$ hides absolute constants.

Proof of Lemma D.1. For any $h_t \in \mathscr{H}_t$, we have

$$
\begin{aligned}
&\mathcal{J}_t^\lambda(\widehat{h}_t^\lambda) - \mathcal{J}_t^\lambda(h_t^\lambda) \\
&= \mathcal{J}_t^\lambda(\widehat{h}_t^\lambda) - \widehat{\mathcal{J}}_t^\lambda(\widehat{h}_t^\lambda) + \widehat{\mathcal{J}}_t^\lambda(\widehat{h}_t^\lambda) - \widehat{\mathcal{J}}_t^\lambda(h_t) + \widehat{\mathcal{J}}_t^\lambda(h_t) - \mathcal{J}_t^\lambda(h_t) + \mathcal{J}_t^\lambda(h_t) - \mathcal{J}_t^\lambda(h_t^\lambda) \\
&\leq \mathcal{J}_t^\lambda(\widehat{h}_t^\lambda) - \widehat{\mathcal{J}}_t^\lambda(\widehat{h}_t^\lambda) + \widehat{\mathcal{J}}_t^\lambda(h_t) - \mathcal{J}_t^\lambda(h_t) + \mathcal{J}_t^\lambda(h_t) - \mathcal{J}_t^\lambda(h_t^\lambda),
\end{aligned}
$$

where the inequality holds from that fact that $\widehat{h}_t^\lambda$ is the minimizer of $\widehat{\mathcal{J}}_t^\lambda$ over the hypothesis class $\mathscr{H}_t$. Taking expectation on both sides of the inequality yields

$$
\begin{aligned}
\mathbb{E}[\mathcal{J}_t^\lambda(\widehat{h}_t^\lambda) - \mathcal{J}_t^\lambda(h_t^\lambda)] &= \mathbb{E}[\mathcal{J}_t^\lambda(\widehat{h}_t^\lambda) - \widehat{\mathcal{J}}_t^\lambda(\widehat{h}_t^\lambda)] + \mathcal{J}_t^\lambda(h_t) - \mathcal{J}_t^\lambda(h_t^\lambda) \\
&\leq \mathbb{E}\Big[\sup_{h_t\in\mathscr{H}_t} \mathcal{J}_t^\lambda(h_t) - \widehat{\mathcal{J}}_t^\lambda(h_t)\Big] + \mathcal{J}_t^\lambda(h_t) - \mathcal{J}_t^\lambda(h_t^\lambda),
\end{aligned}
$$

where the equality holds from $\mathbb{E}[\widehat{\mathcal{J}}_t(h_t)] = \mathcal{J}_t(h_t)$ for each fixed h_t, and the inequality is due to $h_t \in \mathscr{H}_t$. By taking infimum on both sides of the inequality with respect to $h_t \in \mathscr{H}_t$, we have

$$\mathbb{E}[\mathcal{J}_t^\lambda(\widehat{h}_t^\lambda) - \mathcal{J}_t^\lambda(h_t^\lambda)] \leq \underbrace{\mathbb{E}\left[\sup_{h_t \in \mathscr{H}_t} \mathcal{J}_t^\lambda(h_t) - \widehat{\mathcal{J}}_t^\lambda(h_t)\right]}_{\text{generalization error}} + \underbrace{\inf_{h_t \in \mathscr{H}_t} \left\{\mathcal{J}_t^\lambda(h_t) - \mathcal{J}_t^\lambda(h_t^\lambda)\right\}}_{\text{approximation error}}. \tag{D.3}$$

The rest of the proof is divided into three steps.

Step 1. Generalization error in (D.3).

For the generalization error in (D.3), we have the following decomposition:

$$\begin{aligned}
&\mathbb{E}\left[\sup_{h_t \in \mathscr{H}_t} \mathcal{J}_t^\lambda(h_t) - \widehat{\mathcal{J}}_t^\lambda(h_t)\right] \\
&= \underbrace{\mathbb{E}\left[\sup_{h_t \in \mathscr{H}_t} \mathbb{E}[(h_t(\mathbf{X}_{T-t}) - w(\mathbf{X}_0))^2] - \frac{1}{n}\sum_{i=1}^n (h_t(\mathbf{X}_{T-t}^i) - w(\mathbf{X}_0^i))^2\right]}_{\text{(G1)}} \\
&\quad + \underbrace{\lambda\,\mathbb{E}\left[\sup_{h_t \in \mathscr{H}_t} \mathbb{E}[\|\nabla h_t(\mathbf{X}_{T-t})\|_2^2] - \frac{1}{n}\sum_{i=1}^n \|\nabla h_t(\mathbf{X}_{T-t}^i)\|_2^2\right]}_{\text{(G2)}}.
\end{aligned} \tag{D.4}$$

We start from the term (G1) in (D.4). First, recall Proposition 4.1:

$$\mathbb{E}[(h_t(\mathbf{X}_{T-t}) - w(\mathbf{X}_0))^2] = \mathbb{E}[(h_t(\mathbf{X}_{T-t}) - h_t^*(\mathbf{X}_{T-t}))^2] + V_t. \tag{D.5}$$

For the empirical counterpart, it is straightforward that

$$\begin{aligned}
&-\frac{1}{n}\sum_{i=1}^n (h_t(\mathbf{X}_{T-t}^i) - w(\mathbf{X}_0^i))^2 \\
&= -\frac{1}{n}\sum_{i=1}^n (h_t(\mathbf{X}_{T-t}^i) - h_t^*(\mathbf{X}_{T-t}^i) + \mathbb{E}[w(\mathbf{X}_0^i)|\mathbf{X}_{T-t}^i] - w(\mathbf{X}_0^i))^2 \\
&= -\frac{1}{n}\sum_{i=1}^n (h_t(\mathbf{X}_{T-t}^i) - h_t^*(\mathbf{X}_{T-t}^i))^2 - \frac{1}{n}\sum_{i=1}^n (\mathbb{E}[w(\mathbf{X}_0^i)|\mathbf{X}_{T-t}^i] - w(\mathbf{X}_0^i))^2 \\
&\quad - \frac{2}{n}\sum_{i=1}^n (h_t(\mathbf{X}_{T-t}^i) - h_t^*(\mathbf{X}_{T-t}^i))(\mathbb{E}[w(\mathbf{X}_0^i)|\mathbf{X}_{T-t}^i] - w(\mathbf{X}_0^i)),
\end{aligned}$$

where the first equality invokes the definition of the Doob's h-function h_t^* in (3.7). Taking expectation with respect to $\{(\mathbf{X}_0^i, \mathbf{X}_{T-t}^i)\}_{i=1}^n$ yields

$$\mathbb{E}\Big[\sup_{h_t\in\mathscr{H}_t} \mathbb{E}[(h_t(\mathbf{X}_{T-t}) - w(\mathbf{X}_0))^2] - \frac{1}{n}\sum_{i=1}^n (h_t(\mathbf{X}_{T-t}^i) - w(\mathbf{X}_0^i))^2\Big]$$

$$= \mathbb{E}\Big[\sup_{h_t\in\mathscr{H}_t} \mathbb{E}[(h_t(\mathbf{X}_{T-t}) - h_t^*(\mathbf{X}_{T-t}))^2] - \frac{1}{n}\sum_{i=1}^n (h_t(\mathbf{X}_{T-t}^i) - h_t^*(\mathbf{X}_{T-t}^i))^2\Big]$$

$$+ 2\mathbb{E}\Big[\frac{1}{n}\sum_{i=1}^n h_t(\mathbf{X}_{T-t}^i)(\mathbb{E}[w(\mathbf{X}_0^i)|\mathbf{X}_{T-t}^i] - w(\mathbf{X}_0^i))\Big]$$

$$\le 64\bar{B}^2\Big(\frac{\mathrm{VCdim}(\mathscr{H}_t)}{n\log^{-1} n}\Big)^{\frac{1}{2}} + 16\bar{B}^2\Big(\frac{\mathrm{VCdim}(\mathscr{H}_t)}{n\log^{-1} n}\Big)^{\frac{1}{2}} = 80\bar{B}^2\Big(\frac{\mathrm{VCdim}(\mathscr{H}_t)}{n\log^{-1} n}\Big)^{\frac{1}{2}}, \tag{D.6}$$

where the equality invokes (D.5), and the inequality holds from Lemmas D.2 and D.5. For the term (G2) in (D.4), using Lemma D.3 implies

$$\mathbb{E}\Big[\sup_{h_t\in\mathscr{H}_t} \mathbb{E}[\|\nabla h_t(\mathbf{X}_{T-t})\|_2^2] - \frac{1}{n}\sum_{i=1}^n \|\nabla h_t(\mathbf{X}_{T-t}^i)\|_2^2\Big] \le \frac{8d\bar{B}^2}{\sigma_{T-t}^4}\Big(\frac{\mathrm{VCdim}(\nabla\mathscr{H}_t)}{n\log^{-1} n}\Big)^{\frac{1}{2}}. \tag{D.7}$$

Substituting (D.6) and (D.7) into (D.4) yields a generalization error bound:

$$\mathbb{E}\Big[\sup_{h_t\in\mathscr{H}_t} \mathcal{J}_t^\lambda(h_t) - \widehat{\mathcal{J}}_t^\lambda(h_t)\Big] \le 80\bar{B}^2\Big(\frac{\mathrm{VCdim}(\mathscr{H}_t)}{n\log^{-1} n}\Big)^{\frac{1}{2}} + \frac{8d\bar{B}^2}{\sigma_{T-t}^4}\Big(\frac{\mathrm{VCdim}(\nabla\mathscr{H}_t)}{n\log^{-1} n}\Big)^{\frac{1}{2}}. \tag{D.8}$$

Step 2. Approximation error in (D.3). According to the proof of Proposition 4.3, we have

$$\mathcal{J}_t^\lambda(h_t) - \mathcal{J}_t^\lambda(h_t^\lambda) = \|h_t - h_t^\lambda\|_{L^2(p_{T-t})}^2 + \lambda\|\nabla h_t - \nabla h_t^\lambda\|_{L^2(p_{T-t})}^2. \tag{D.9}$$

Step 3. Conclusion. Substituting (D.8) and (D.9) into (D.3) completes the proof. □

D.2 Auxilary lemmas for the oracle inequality. According to the standard techniques of symmetrization (Mohri et al., 2018, Theorem 3.3), we have the following generalization bounds. We introduce the concept of Rademacher complexity (Bartlett and Mendelson, 2002; Mohri et al., 2018), which is crucial for analyzing the generalization error.

Definition 3 (Rademacher complexity)**.** Let $\mathscr{H}$ be a function class, and let $\mathbf{X}^{1:n} := (\mathbf{X}^1, \ldots, \mathbf{X}^n)$ be a set of samples. The empirical Rademacher complexity of $\mathscr{H}$ with respect to $\mathbf{X}_{1:n}$ is defined as

$$\mathfrak{R}(\mathscr{H}|\mathbf{X}^{1:n}) := \mathbb{E}\Big[\sup_{h\in\mathscr{H}} \frac{1}{n}\sum_{i=1}^n \varepsilon_i h(\mathbf{X}^i)\Big|\mathbf{X}^{1:n}\Big],$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. Rademacher random variables. The Rademacher complexity of $\mathscr{H}$ is the expectation of empirical Rademacher complexity with respect to the distribution of $\mathbf{X}^{1:n}$ defined as

$$\mathfrak{R}_n(\mathscr{H}) := \mathbb{E}[\widehat{\mathfrak{R}}(\mathscr{H}|\mathbf{X}^{1:n})] = \mathbb{E}\Big[\sup_{h\in\mathscr{H}} \frac{1}{n}\sum_{i=1}^n \varepsilon_i h(\mathbf{X}^i)\Big].$$

Lemma D.2. *Suppose Assumptions 1 and 2 hold. Then*

$$\mathbb{E}\Big[\sup_{h_t\in\mathscr{H}_t} \|h_t - h_t^*\|_{L^2(p_{T-t})}^2 - \frac{1}{n}\sum_{i=1}^n (h_t(\mathbf{X}_{T-t}^i) - h_t^*(\mathbf{X}_{T-t}^i))^2\Big] \le 64\bar{B}^2\Big(\frac{\mathrm{VCdim}(\mathscr{H}_t)}{n\log^{-1} n}\Big)^{\frac{1}{2}},$$

where the expectation is taken with respect to $\mathbf{X}_{T-t}^1, \ldots, \mathbf{X}_{T-t}^n \sim^{\text{i.i.d.}} p_{T-t}$.

Proof of Lemma D.2. Let $\mathbf{X}_{T-t}^{1,\prime}, \dots, \mathbf{X}_{T-t}^{n,\prime}$ be independent copies of $\mathbf{X}_{T-t}^{1}, \dots, \mathbf{X}_{T-t}^{n}$. Let $\varepsilon_1, \dots, \varepsilon_n$ be a set of i.i.d. Rademacher variables, which are independent of $\mathbf{X}_{T-t}^{1,\prime}, \dots, \mathbf{X}_{T-t}^{n,\prime}$ and $\mathbf{X}_{T-t}^{1}, \dots, \mathbf{X}_{T-t}^{n}$. It follows that

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{h_t \in \mathscr{H}_t} \|h_t - h_t^*\|_{L^2(p_{T-t})}^2 - \frac{1}{n}\sum_{i=1}^n (h_t(\mathbf{X}_{T-t}^i) - h_t^*(\mathbf{X}_{T-t}^i))^2\Big] \\
&\leq \mathbb{E}\Big[\sup_{h_t \in \mathscr{H}_t} \mathbb{E}\big[(h_t(\mathbf{X}_{T-t}) - h_t^*(\mathbf{X}_{T-t}))^2\big] - \frac{1}{n}\sum_{i=1}^n (h_t(\mathbf{X}_{T-t}^i) - h_t^*(\mathbf{X}_{T-t}^i))^2\Big] \\
&= \mathbb{E}\Big[\sup_{h_t \in \mathscr{H}_t} \mathbb{E}\Big[\frac{1}{n}\sum_{i=1}^n (h_t(\mathbf{X}_{T-t}^{i,\prime}) - h_t^*(\mathbf{X}_{T-t}^{i,\prime}))^2\Big] - \frac{1}{n}\sum_{i=1}^n (h_t(\mathbf{X}_{T-t}^i) - h_t^*(\mathbf{X}_{T-t}^i))^2\Big] \\
&\leq \mathbb{E}\Big[\sup_{h_t \in \mathscr{H}_t} \frac{1}{n}\sum_{i=1}^n \Big\{(h_t(\mathbf{X}_{T-t}^{i,\prime}) - h_t^*(\mathbf{X}_{T-t}^{i,\prime}))^2 - (h_t(\mathbf{X}_t^i) - h_t^*(\mathbf{X}_t^i))^2\Big\}\Big] \\
&= \mathbb{E}\Big[\sup_{h_t \in \mathscr{H}_t} \frac{1}{n}\sum_{i=1}^n \varepsilon_i\Big\{(h_t(\mathbf{X}_{T-t}^{i,\prime}) - h_t^*(\mathbf{X}_{T-t}^{i,\prime}))^2 - (h_t(\mathbf{X}_{T-t}^i) - h_t^*(\mathbf{X}_{T-t}^i))^2\Big\}\Big] \\
&= 2\mathbb{E}\Big[\sup_{h_t \in \mathscr{H}_t} \frac{1}{n}\sum_{i=1}^n \varepsilon_i (h_t(\mathbf{X}_{T-t}^i) - h_t^*(\mathbf{X}_{T-t}^i))^2\Big] \leq 8\bar{B}\mathfrak{R}_n(\mathscr{H}_t),
\end{aligned} \tag{D.10}$$

where the second inequality holds from Jensen's inequality, and last inequality is due to Ledoux-Talagrand contraction inequality (Mohri et al., 2018, Lemma 5.7) and Lemma C.1.

It remains to bound the Rademacher complexity $\mathfrak{R}_n(\mathscr{H}_t)$ in (D.10). Let $\delta > 0$ and $\mathscr{H}_t^\delta$ be an $L^\infty(\mathbf{X}_{T-t}^{1:n})$ δ-cover of $\mathscr{H}_t$ satisfying $|\mathscr{H}_t^\delta| = N(\delta, \mathscr{H}_t, L^\infty(\mathbf{X}_{T-t}^{1:n}))$. Then for any $h_t \in \mathscr{H}_t$, there exists $h_t^\delta \in \mathscr{H}_t^\delta$ such that

$$\frac{1}{n}\sum_{i=1}^n \varepsilon_i h_t(\mathbf{X}_t^i) - \frac{1}{n}\sum_{i=1}^n \varepsilon_i h_t^\delta(\mathbf{X}_t^i) \leq \delta.$$

As a consequence,

$$\begin{aligned}
\mathfrak{R}(\mathscr{H}_t \mid \mathbf{X}_{T-t}^{1:n}) &= \mathbb{E}\Big[\sup_{h_t \in \mathscr{H}_t} \frac{1}{n}\sum_{i=1}^n \varepsilon_i h_t(\mathbf{X}_t^i) \mid \mathbf{X}_{T-t}^{1:n}\Big] \\
&\leq \mathbb{E}\Big[\sup_{h_t^\delta \in \mathscr{H}_t^\delta} \frac{1}{n}\sum_{i=1}^n \varepsilon_i h_t^\delta(\mathbf{X}_t^i) \mid \mathbf{X}_{T-t}^{1:n}\Big] + \delta \\
&\leq \bar{B}\Big(\frac{2\log|\mathscr{H}_t^\delta|}{n}\Big)^{\frac{1}{2}} + \delta \\
&= \bar{B}\Big(\frac{2\log N(\delta, \mathscr{H}_t, L^\infty(\mathbf{X}_{T-t}^{1:n}))}{n}\Big)^{\frac{1}{2}} + \delta,
\end{aligned}$$

where $\varepsilon_1, \dots, \varepsilon_n$ are a sequence of i.i.d. Rademacher variables, the second inequality follows from Massart's lemma (Mohri et al., 2018, Theorem 3.7), and the equality is due to the definition of $\mathscr{H}_t^\delta$. Then setting $\delta = \bar{B}/\sqrt{n}$ yields

$$\mathfrak{R}(\mathscr{H}_t \mid \mathbf{X}_{T-t}^{1:n}) \leq \bar{B}\Big(\frac{2\log N(B/\sqrt{n}, \mathscr{H}_t, L^\infty(\mathbf{X}_{T-t}^{1:n}))}{n}\Big)^{\frac{1}{2}} \leq 8\bar{B}\Big(\frac{\mathrm{VCdim}(\mathscr{H}_t)}{n\log^{-1} n}\Big)^{\frac{1}{2}}, \tag{D.11}$$

where the last inequality holds from Anthony et al. (1999, Theorem 12.2). Substituting (D.11) into (D.10) completes the proof. ∎

By a similar argument as Lemma D.2, we have the following generalization bounds for the gradient term.

Lemma D.3. *Suppose Assumptions 1 and 2 hold. Then*

$$\mathbb{E}\left[\sup_{h_t \in \mathscr{H}_t} \|\nabla h_t\|^2_{L^2(p_{T-t})} - \frac{1}{n}\sum_{i=1}^n \|\nabla h_t(\mathbf{X}^i_{T-t})\|_2^2\right] \le \frac{8d\bar{B}^2}{\sigma^4_{T-t}}\Big(\frac{\mathrm{VCdim}(\nabla \mathscr{H}_t)}{n \log^{-1} n}\Big)^{\frac{1}{2}},$$

where the expectation is taken with respect to $\mathbf{X}^1_{T-t}, \dots, \mathbf{X}^i_{T-t} \sim^{\text{i.i.d.}} p_{T-t}$.

Proof of Lemma D.3. It is straightforward that

$$\begin{aligned}
&\mathbb{E}\left[\sup_{h_t \in \mathscr{H}_t} \|\nabla h_t\|^2_{L^2(p_{T-t})} - \frac{1}{n}\sum_{i=1}^n \|\nabla h_t(\mathbf{X}^i_{T-t})\|_2^2\right] \\
&\le \sum_{k=1}^d \mathbb{E}\left[\sup_{h_t \in \mathscr{H}_t} \|D_k h_t\|^2_{L^2(p_{T-t})} - \frac{1}{n}\sum_{i=1}^n (D_k h_t(\mathbf{X}^i_{T-t}))^2\right] \\
&\le \sum_{k=1}^d \frac{8\bar{B}^2}{\sigma^4_{T-t}}\Big(\frac{\mathrm{VCdim}(D_k \mathscr{H}_t)}{n \log^{-1} n}\Big)^{\frac{1}{2}} \le \frac{8d\bar{B}^2}{\sigma^4_{T-t}}\Big(\frac{\mathrm{VCdim}(\nabla \mathscr{H}_t)}{n \log^{-1} n}\Big)^{\frac{1}{2}},
\end{aligned}$$

where the first inequality holds from the convexity of supremum and Jensen's inequality, the second inequality invokes a similar argument as Lemma D.2, and the last inequality holds from the definition of $\mathrm{VCdim}(\nabla \mathscr{H}_t)$. This completes the proof. □

The following lemma is an extension of Bartlett and Mendelson (2002, Lemma 4).

Lemma D.4. *Let* $\mathbf{z} = (z_1, \dots, z_n) \in \mathcal{Z} \subseteq \mathbb{R}^n$. *Let* $\xi_1, \dots, \xi_n$ *be a sequence of i.i.d. random variables with* $|\xi_i| < K$ *and* $\mathbb{E}[\xi_i] = 0$ *for each* $1 \le i \le n$. *Then it follows that*

$$\mathbb{E}\left[\sup_{\mathbf{z} \in \mathcal{Z}} \frac{1}{n}\sum_{i=1}^n \xi_i z_i\right] \le 2K\mathbb{E}\left[\sup_{\mathbf{z} \in \mathcal{Z}} \frac{1}{n}\sum_{i=1}^n \varepsilon_i z_i\right],$$

where $\varepsilon_1, \dots, \varepsilon_n$ *is a sequence of i.i.d. Rademacher variables.*

Proof of Lemma D.4. The proof relies on the symmetrization technique. Let $\xi'_1, \dots, \xi'_n$ be independent copies of $\xi_1, \dots, \xi_n$. It follows that

$$\begin{aligned}
\mathbb{E}\left[\sup_{\mathbf{z} \in \mathcal{Z}} \frac{1}{n}\sum_{i=1}^n \xi_i z_i\right] &= \mathbb{E}\left[\sup_{\mathbf{z} \in \mathcal{Z}} \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n (\xi_i - \xi'_i) z_i \Big| \xi_1, \dots, \xi_n\right]\right] \\
&\le \mathbb{E}\left[\sup_{\mathbf{z} \in \mathcal{Z}} \frac{1}{n}\sum_{i=1}^n (\xi_i - \xi'_i) z_i\right] = \mathbb{E}\left[\sup_{\mathbf{z} \in \mathcal{Z}} \frac{1}{n}\sum_{i=1}^n \varepsilon_i(\xi_i - \xi'_i) z_i\right] \\
&\le 2\mathbb{E}\left[\sup_{\mathbf{z} \in \mathcal{Z}} \frac{1}{n}\sum_{i=1}^n \varepsilon_i \xi_i z_i\right] \le 2K\mathbb{E}\left[\sup_{\mathbf{z} \in \mathcal{Z}} \frac{1}{n}\sum_{i=1}^n \varepsilon_i z_i\right],
\end{aligned}$$

where the first equality due to $\mathbb{E}[\xi'_i] = 0$, the first inequality holds from Jensen's inequality, and the second equality follows from the fact that distribution of $(\xi_i - \xi'_i)$ is symmetric around zero, so it has the same distribution as $\varepsilon_i(\xi_i - \xi'_i)$. The second inequality comes from the triangular inequality for the supremum, and we used the fact that ξ_i and ξ'_i are identically distributed. The last inequality invokes Ledoux-Talagrand contraction inequality (Mohri et al., 2018, Lemma 5.7) and $\max_{1 \le i \le n} |\xi_i| \le K$. This completes the proof. □

Lemma D.5. *Suppose Assumptions 1 and 2 hold. Then*

$$\mathbb{E}\Big[\sup_{h_t \in \mathscr{H}_t} \frac{1}{n}\sum_{i=1}^{n} h_t(\mathbf{X}_{T-t}^i)(w(\mathbf{X}_0^i) - \mathbb{E}[w(\mathbf{X}_0^i) \mid \mathbf{X}_{T-t}^i]) \mid \mathbf{X}_{T-t}^{1:n}\Big] \leq 16\bar{B}^2\Big(\frac{\mathrm{VCdim}(\mathscr{H}_t)}{n\log^{-1} n}\Big)^{\frac{1}{2}},$$

where the expectation is taken with respect to $\mathbf{X}_{T-t}^1, \dots, \mathbf{X}_{T-t}^i \sim^{\text{i.i.d.}} p_{T-t}$.

Proof of Lemma D.5. Define a sequence of auxilary random variables

$$\xi_i := w(\mathbf{X}_0^i) - \mathbb{E}[w(\mathbf{X}_0^i) \mid \mathbf{X}_{T-t}^i].$$

It is apparent that $\mathbb{E}[\xi_i \mid \mathbf{X}_{T-t}^i] = 0$, and $|\xi_i| \leq \bar{B}$. Using Lemma D.4 yields

$$\begin{aligned}\mathbb{E}\Big[\sup_{h_t \in \mathscr{H}_t} \frac{1}{n}\sum_{i=1}^{n} \xi_i h_t(\mathbf{X}_{T-t}^i) \mid \mathbf{X}_{T-t}^{1:n}\Big] &\leq 2\bar{B}\mathbb{E}\Big[\sup_{h_t \in \mathscr{H}_t} \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i h_t(\mathbf{X}_{T-t}^i) \mid \mathbf{X}_{T-t}^{1:n}\Big] \\ &= 2\bar{B}\mathfrak{R}(\mathscr{H}_t \mid \mathbf{X}_{T-t}^{1:n}),\end{aligned}$$

where $\varepsilon_1, \dots, \varepsilon_n$ is a sequence of i.i.d. Rademacher variables. Here the first inequality follows from Lemma D.4, and the second inequality is due to the fact that $\widehat{h}_t^\lambda \in \mathscr{H}_t$, the second inequality holds from Lemma D.4. Finally, using (D.11) completes the proof. □

Lemma D.6. *Suppose Assumptions 1 and 2 hold. Then*

$$\|\Delta h_t^* + \nabla h_t^* \cdot \nabla \log p_{T-t}\|_{L^2(p_{T-t})} \leq \frac{12\sqrt{d}\bar{B}}{\sigma_{T-t}^4}.$$

Proof of Lemma D.6. By applying the triangular inequality, we have

$$\begin{aligned}&\|\Delta h_t^* + \nabla h_t^* \cdot \nabla \log p_{T-t}\|_{L^2(p_{T-t})} \\ &\leq \|\Delta h_t^*\|_{L^2(p_{T-t})} + \|\nabla h_t^* \cdot \nabla \log p_{T-t}\|_{L^2(p_{T-t})} \\ &\leq \frac{6\sqrt{d}\bar{B}}{\sigma_{T-t}^4} + \frac{2\bar{B}}{\sigma_{T-t}^2}\|\nabla \log p_{T-t}\|_{L^2(p_{T-t})},\end{aligned} \tag{D.12}$$

where the last inequality holds from Lemmas C.4 and C.5. It remains to estimate the $L^2(p_{T-t})$-norm of the score $\nabla \log p_{T-t}$ in (D.12). Indeed,

$$\begin{aligned}&\|\nabla \log p_{T-t}\|_{L^2(p_{T-t})}^2 \\ &= \int \Big\|\frac{\mathbf{x}}{\sigma_{T-t}^2} - \frac{\mu_{T-t}}{\sigma_{T-t}^2}\mathbb{E}[\mathbf{X}_0|\mathbf{X}_{T-t} = \mathbf{x}]\Big\|_2^2 p_{T-t}(\mathbf{x})\,\mathrm{d}\mathbf{x} \\ &\leq \frac{2}{\sigma_{T-t}^4}\mathbb{E}\big[\|\mathbf{X}_{T-t}\|_2^2\big] + \frac{2d\mu_{T-t}^2}{\sigma_{T-t}^4} \\ &= \frac{2}{\sigma_{T-t}^4}\Big\{\mu_{T-t}^2\mathbb{E}\big[\|\mathbf{X}_0\|_2^2\big] + 2\mu_{T-t}\sigma_{T-t}\mathbb{E}\big[\langle \mathbf{X}_0, \boldsymbol{\varepsilon}\rangle\big] + \sigma_{T-t}^2\mathbb{E}\big[\|\boldsymbol{\varepsilon}\|_2^2\big]\Big\} + \frac{2d\mu_{T-t}^2}{\sigma_{T-t}^4} \\ &= \frac{2}{\sigma_{T-t}^4}\Big\{\mu_{T-t}^2\mathbb{E}\big[\|\mathbf{X}_0\|_2^2\big] + \sigma_{T-t}^2\mathbb{E}\big[\|\boldsymbol{\varepsilon}\|_2^2\big]\Big\} + \frac{2d\mu_{T-t}^2}{\sigma_{T-t}^4} \leq \frac{6d}{\sigma_{T-t}^4},\end{aligned} \tag{D.13}$$

where the first equality is owing to Lemma C.2, the first inequality used Assumption 1, and the second and third equalities hold from $\mathbf{X}_{T-t} \overset{\mathrm{d}}{=} \mu_{T-t}\mathbf{X}_0 + \sigma_{T-t}\boldsymbol{\varepsilon}$ where $\mathbf{X}_0$ is independent $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$. The last inequality also uses Assumption 1. Substituting (D.13) into (D.12) completes the proof. □

D.3 Convergence rate of the Doob's guidance estimation.

Lemma D.7. *Suppose Assumptions 1 and 2 hold. Let $R \geq 1$, and let the hypothesis class $\mathscr{H}_t$ be defined as (5.2) with $L \leq C \log N$ and $S \leq N^d$, then there exists $h_t \in \mathscr{H}_t$ such that*

$$\|h_t - h_t^*\|_{L^\infty(B(\mathbf{0},R))} \leq \frac{C\bar{B}R^2}{\sigma_{T-t}^4 N^2},$$
$$\|\nabla h_t - \nabla h_t^*\|_{L^\infty(B(\mathbf{0},R))} \leq \frac{C\bar{B}R}{\sigma_{T-t}^4 N},$$

where C is a constant only depending on d.

Proof of Lemma D.7. We first rescale the target function h_t^* to $B(\mathbf{0},1)$ by $g_t^*(\mathbf{z}) := h_t^*(R\mathbf{z})$. According to Ding et al. (2025b, Lemma 6), there exists $g_t \in N(L,S)$ such that

$$\|g_t - g_t^*\|_{L^\infty(B(\mathbf{0},1))} \leq \frac{C'}{N^2}\|g_t^*\|_{C^2(\mathbb{R}^d)},$$
$$\|\nabla g_t - \nabla g_t^*\|_{L^\infty(B(\mathbf{0},1))} \leq \frac{C'}{N}\|g_t^*\|_{C^2(\mathbb{R}^d)},$$

where C' is a constant only depending on d. Note that $D_k g_t^*(\mathbf{z}) = RD_k h_t^*(R\mathbf{z})$ for each $1 \leq k \leq d$, and $D_{k\ell}^2 g_t^*(\mathbf{z}) = R^2 D_{k\ell}^2 h_t^*(R\mathbf{z})$ for each $1 \leq k,\ell \leq d$. Thus

$$\|g_t(R^{-1}\cdot) - h_t^*\|_{L^\infty(B(\mathbf{0},R))} = \|g_t(R^{-1}\cdot) - g_t^*(R^{-1}\cdot)\|_{L^\infty(B(\mathbf{0},1))} \leq \frac{C'R^2}{N^2}\|h_t^*\|_{C^2(\mathbb{R}^d)},$$
$$\|\nabla g_t(R^{-1}\cdot) - \nabla h_t^*\|_{L^\infty(B(\mathbf{0},R))} = \frac{1}{R}\|\nabla g_t(R^{-1}\cdot) - \nabla g_t^*(R^{-1}\cdot)\|_{L^\infty(B(\mathbf{0},1))} \leq \frac{C'R}{N}\|h_t^*\|_{C^2(\mathbb{R}^d)}.$$

Setting $h_t := g_t(R^{-1}\cdot)$, and using Lemmas C.1, C.4, and C.5 complete the proof. ☐

Lemma D.8 (Approximation error)**.** *Suppose Assumptions 1 and 2 hold. Let $R \geq 1$, and let the hypothesis class $\mathscr{H}_t$ be defined as (5.2) with $L \leq C \log N$ and $S \leq N^d$, then*

$$\|h_t - h_t^*\|_{L^2(p_{T-t})}^2 \leq C\frac{\bar{B}^2 \log^4 N}{\sigma_{T-t}^8 N^4},$$
$$\|\nabla h_t - \nabla h_t^*\|_{L^2(p_{T-t})}^2 \leq C\frac{\bar{B}^2 \log^2 N}{\sigma_{T-t}^8 N^2}.$$

provided that $R^2 = (4d\mu_t^2 + 8\sigma_t^2)\log N^4$, where C is a constant only depending on d.

Proof of Lemma D.8. It is straightforward that for each $R \geq 1$,

$$\begin{aligned}\|h_t - h_t^*\|_{L^2(p_{T-t})}^2 = &\underbrace{\int (h_t(\mathbf{x}) - h_t^*(\mathbf{x}))^2 \mathbb{1}\{\|\mathbf{x}\|_2 \leq R\} p_{T-t}(\mathbf{x})\,\mathrm{d}\mathbf{x}}_{\text{(i)}} \\ &+ \underbrace{\int (h_t(\mathbf{x}) - h_t^*(\mathbf{x}))^2 \mathbb{1}\{\|\mathbf{x}\|_2 > R\} p_{T-t}(\mathbf{x})\,\mathrm{d}\mathbf{x}}_{\text{(ii)}}.\end{aligned} \tag{D.14}$$

For term (i) in (D.14), we have

$$\begin{aligned}&\int (h_t(\mathbf{x}) - h_t^*(\mathbf{x}))^2 \mathbb{1}\{\|\mathbf{x}\|_2 \leq R\} p_{T-t}(\mathbf{x}) \, \mathrm{d}\mathbf{x} \\ &\leq \sup_{\|\mathbf{x}\|_2 \leq R} (h_t(\mathbf{x}) - h_t^*(\mathbf{x}))^2 \leq \frac{C^2 \bar{B}^2 R^4}{\sigma_{T-t}^8 N^4},\end{aligned} \tag{D.15}$$

where the second inequality holds from Lemma D.7. For term (ii) in (D.14), we have

$$\begin{aligned}&\int \|h_t(\mathbf{x}) - h_t^*(\mathbf{x})\|_2^2 \mathbb{1}\{\|\mathbf{x}\|_2 > R\} p_{T-t}(\mathbf{x}) \, \mathrm{d}\mathbf{x} \\ &\leq 4\bar{B}^2 \Pr\{\|\mathbf{X}_{T-t}\|_2 > R\} \leq 2^{d+3} \bar{B}^2 \exp\Big(-\frac{R^2}{4d\mu_t^2 + 8\sigma_t^2}\Big),\end{aligned} \tag{D.16}$$

where the first inequality holds from Lemma C.1, and the second inequality is due to Lemma G.1. Substituting (D.15) and (D.16) into (D.14) yields

$$\|h_t - h_t^*\|_{L^2(p_{T-t})}^2 \leq \frac{C^2 \bar{B}^2 R^4}{\sigma_{T-t}^8 N^4} + 2^{d+3} \bar{B}^2 \exp\Big(-\frac{R^2}{4d\mu_t^2 + 8\sigma_t^2}\Big). \tag{D.17}$$

Similarly, for the gradient term, we have

$$\begin{aligned}\|\nabla h_t - \nabla h_t^*\|_{L^2(p_{T-t})}^2 &= \underbrace{\int \|\nabla h_t(\mathbf{x}) - \nabla h_t^*(\mathbf{x})\|_2^2 \mathbb{1}\{\|\mathbf{x}\|_2 \leq R\} p_{T-t}(\mathbf{x}) \, \mathrm{d}\mathbf{x}}_{\text{(i)}} \\ &\quad + \underbrace{\int \|\nabla h_t(\mathbf{x}) - \nabla h_t^*(\mathbf{x})\|_2^2 \mathbb{1}\{\|\mathbf{x}\|_2 > R\} p_{T-t}(\mathbf{x}) \, \mathrm{d}\mathbf{x}}_{\text{(ii)}}.\end{aligned} \tag{D.18}$$

For term (i) in (D.18), we have

$$\begin{aligned}&\int \|\nabla h_t(\mathbf{x}) - \nabla h_t^*(\mathbf{x})\|_2^2 \mathbb{1}\{\|\mathbf{x}\|_2 \leq R\} p_{T-t}(\mathbf{x}) \, \mathrm{d}\mathbf{x} \\ &\leq \sup_{\|\mathbf{x}\|_2 \leq R} \|\nabla h_t(\mathbf{x}) - \nabla h_t^*(\mathbf{x})\|_2^2 \leq \frac{C^2 \bar{B}^2 R^2}{\sigma_{T-t}^8 N^2},\end{aligned} \tag{D.19}$$

where the second inequality holds from Lemma D.7. For term (ii) in (D.18), we have

$$\begin{aligned}&\int \|\nabla h_t(\mathbf{x}) - \nabla h_t^*(\mathbf{x})\|_2^2 \mathbb{1}\{\|\mathbf{x}\|_2 > R\} p_{T-t}(\mathbf{x}) \, \mathrm{d}\mathbf{x} \\ &\leq \frac{16\bar{B}^2}{\sigma_{T-t}^4} \Pr\{\|\mathbf{X}_{T-t}\|_2 > R\} \leq 2^{d+5} \frac{\bar{B}^2}{\sigma_{T-t}^4} \exp\Big(-\frac{R^2}{4d\mu_t^2 + 8\sigma_t^2}\Big),\end{aligned} \tag{D.20}$$

where the first inequality holds from Lemma C.4, and the second inequality is due to Lemma G.1. Substituting (D.19) and (D.20) into (D.18) yields

$$\|\nabla h_t - \nabla h_t^*\|_{L^2(p_{T-t})}^2 \leq \frac{C^2 \bar{B}^2 R^2}{\sigma_{T-t}^8 N^2} + 2^{d+5} \frac{\bar{B}^2}{\sigma_{T-t}^4} \exp\Big(-\frac{R^2}{4d\mu_t^2 + 8\sigma_t^2}\Big). \tag{D.21}$$

Setting $R^2 = (4d\mu_t^2 + 8\sigma_t^2) \log N^4$ in (D.17) and (D.21) completes the proof. □

Lemma D.9 (Generalization error)**.** *Suppose Assumptions 1 and 2 hold. Let the hypothesis class $\mathscr{H}_t$ be defined as (5.2), then*

$$\mathrm{VCdim}(\mathscr{H}_t) \le cLS\log(S),$$
$$\mathrm{VCdim}(\nabla\mathscr{H}_t) \le cL^2S\log(LS),$$

where c is an absolute constant.

Proof of Lemma D.9. Since $\mathscr{H}_t \subseteq N(L,S)$, using Bartlett et al. (2019, Theorem 7) implies

$$\mathrm{VCdim}(\mathscr{H}_t) \le \mathrm{VCdim}(N(L,S)) \le c_1LS\log(S),$$

where c_1 is an absolute constant. According to Ding et al. (2025b, Lemma 13), we have $\nabla\mathscr{H}_t \subseteq N(c_2L, c_3LS)$, where c_2 and c_3 are absolute constants. Using Bartlett et al. (2019, Theorem 7) again implies

$$\mathrm{VCdim}(\nabla\mathscr{H}_t) \le c_4L^2S\log(LS),$$

where c_4 is an absolute constant. This completes the proof. □

Theorem 5.3. *Suppose Assumptions 1 and 2 hold. Let $t \in (0,T)$. Set the hypothesis class $\mathscr{H}_t$ as*

$$\mathscr{H}_t := \left\{ h_t \in N(L,S) : \begin{array}{c} \sup_{\mathbf{x}\in\mathbb{R}^d} h_t(\mathbf{x}) \le \bar{B}, \ \inf_{\mathbf{x}\in\mathbb{R}^d} h_t(\mathbf{x}) \ge \underline{B}, \\ \max_{1\le k\le d} \sup_{\mathbf{x}\in\mathbb{R}^d} |D_k h_t(\mathbf{x})| \le 2\sigma_{T-t}^{-2}\bar{B} \end{array} \right\},$$

where $L = \mathcal{O}(\log n)$ and $S = \mathcal{O}(n^{\frac{d}{d+8}})$. Let $\widehat{h}_t^\lambda$ be the gradient-regularized empirical risk minimizer defined as (4.6), and let h_t^* be the Doob's h-function defined as (3.7). Then the following inequality holds:

$$\mathbb{E}\Big[\|\nabla\log\widehat{h}_t^\lambda - \nabla\log h_t^*\|_{L^2(p_{T-t})}^2\Big] \le C\sigma_{T-t}^{-8}n^{-\frac{2}{d+8}},$$

provided that the regularization parameter λ is set as $\lambda = \mathcal{O}(n^{-\frac{2}{d+8}})$, where C is a constant depending only on d, $\bar{B}$, and $\underline{B}$.

Proof of Theorem 5.3. Substituting Lemmas D.8, and D.9 into Lemma 5.2 yields

$$\begin{aligned}&\mathbb{E}\Big[\|\widehat{h}_t^\lambda - h_t^*\|_{L^2(p_{T-t})}^2\Big] \\ &\le C\frac{\log^4 N}{\sigma_{T-t}^8N^4} + C\lambda\frac{\log^2 N}{\sigma_{T-t}^8N^2} + C\Big(\frac{N^d\log^2 N}{n\log^{-1}n}\Big)^{\frac{1}{2}} + C\frac{\lambda}{\sigma_{T-t}^4}\Big(\frac{N^d\log^4 N}{n\log^{-1}n}\Big)^{\frac{1}{2}} + C\frac{\lambda^2}{\sigma_{T-t}^8},\end{aligned}$$

where C is a constant only depending on d and $\bar{B}$, and we used the fact $L \le C'\log N$ and $S \le N^d$ in Lemma D.8. Similarly,

$$\begin{aligned}&\mathbb{E}[\|\nabla\widehat{h}_t^\lambda - \nabla h_t^*\|_{L^2(p_{T-t})}^2] \\ &\le \frac{C}{\lambda}\frac{\log^4 N}{\sigma_{T-t}^8N^4} + C\frac{\log^2 N}{\sigma_{T-t}^8N^2} + \frac{C}{\lambda}\Big(\frac{N^d\log^2 N}{n\log^{-1}n}\Big)^{\frac{1}{2}} + \frac{C}{\sigma_{T-t}^4}\Big(\frac{N^d\log^4 N}{n\log^{-1}n}\Big)^{\frac{1}{2}} + \frac{C\lambda}{\sigma_{T-t}^8}.\end{aligned}$$

By setting $N = \mathcal{O}(n^{\frac{1}{d+8}})$ and $\lambda = \mathcal{O}(n^{-\frac{2}{d+8}})$, we have

$$\begin{aligned} \mathbb{E}[\|\widehat{h}_t^\lambda - h_t^*\|_{L^2(p_{T-t})}^2] &\lesssim \frac{1}{\sigma_{T-t}^8} n^{-\frac{4}{d+8}} \log^4 n, \\ \mathbb{E}[\|\nabla\widehat{h}_t^\lambda - \nabla h_t^*\|_{L^2(p_{T-t})}^2] &\lesssim \frac{1}{\sigma_{T-t}^8} n^{-\frac{2}{d+8}} \log^4 n. \end{aligned} \tag{D.22}$$

Consequently,

$$\begin{aligned} &\|\nabla \log \widehat{h}_t^\lambda - \nabla \log h_t^*\|_{L^2(p_{T-t})}^2 \\ &= \Big\| \frac{\nabla \widehat{h}_t^\lambda}{\widehat{h}_t^\lambda} - \frac{\nabla h_t^*}{\widehat{h}_t^\lambda} + \frac{\nabla h_t^*}{\widehat{h}_t^\lambda} - \frac{\nabla h_t^*}{h_t^*} \Big\|_{L^2(p_{T-t})}^2 \\ &\leq 2\Big\| \frac{\nabla \widehat{h}_t^\lambda}{\widehat{h}_t^\lambda} - \frac{\nabla h_t^*}{\widehat{h}_t^\lambda} \Big\|_{L^2(p_{T-t})}^2 + 2\Big\| \frac{\nabla h_t^*}{\widehat{h}_t^\lambda} - \frac{\nabla h_t^*}{h_t^*} \Big\|_{L^2(p_{T-t})}^2 \\ &\leq \frac{2}{\underline{B}^2} \|\nabla \widehat{h}_t^\lambda - \nabla h_t^*\|_{L^2(p_{T-t})}^2 + 2\frac{\bar{B}^2}{\underline{B}^4} \|\widehat{h}_t^\lambda - h_t^*\|_{L^2(p_{T-t})}^2 \\ &\leq C' \frac{1}{\sigma_{T-t}^8} n^{-\frac{2}{d+8}} \log^4 n, \end{aligned}$$

where the second inequality is owing to Lemmas C.1 and C.4, and the last inequality holds from (D.22). This completes the proof. □

E Derivations in Section 5.3

E.1 Error decomposition of the controllable diffusion models.

Lemma 5.5. *Suppose Assumptions 1, 2, and 3 hold. Let $\widehat{q}_{T-T_0}$ be the marginal density of $\widehat{\mathbf{Z}}_{T-T_0}^{\leftarrow}$ defined in (4.8). Then it follows that*

$$\begin{aligned} \mathrm{KL}(q_{T_0}\|\widehat{q}_{T-T_0}) \lesssim \frac{\bar{B}}{\underline{B}} \sum_{k=0}^{K-1} h \mathbb{E}^{\mathbb{P}} \Big[\|\nabla \log \widehat{h}_{kh}(\mathbf{X}_{kh}^{\leftarrow}) - \nabla \log h_{kh}^*(\mathbf{X}_{kh}^{\leftarrow})\|_2^2 \Big] \\ + \frac{\bar{B}}{\underline{B}} T \varepsilon_{\mathrm{ref}}^2 + d \exp(-T) + \frac{d^2 T^2}{\sigma_{T_0}^4 K}, \end{aligned}$$

where the notation $\lesssim$ hides absolute constants.

Proof of Lemma 5.5. According to Chen et al. (2023a, Proposition C.3), we have

$$\begin{aligned} \mathrm{KL}(q_{T_0}\|\widehat{q}_{T-T_0}) \lesssim \underbrace{\mathrm{KL}(q_T\|\gamma_d)}_{\text{initial error}} + \underbrace{\sum_{k=0}^{K-1} h \mathbb{E}^{\mathbb{Q}} \Big[\|\widehat{\mathbf{s}}(kh, \mathbf{Z}_{kh}^{\leftarrow}) - \nabla \log p_{T-kh}(\mathbf{Z}_{kh}^{\leftarrow})\|_2^2 \Big]}_{\text{base score error}} \\ + \underbrace{\sum_{k=0}^{K-1} h \mathbb{E}^{\mathbb{Q}} \Big[\|\nabla \log \widehat{h}_{kh}(\mathbf{Z}_{kh}^{\leftarrow}) - \nabla \log h_{kh}^*(\mathbf{Z}_{kh}^{\leftarrow})\|_2^2 \Big]}_{\text{Doob's guidance error}} \\ + \underbrace{\sum_{k=0}^{K-1} \int_{kh}^{(k+1)h} \mathbb{E}^{\mathbb{Q}} \Big[\|\nabla \log q_{T-t}(\mathbf{Z}_t^{\leftarrow}) - \nabla \log q_{T-kh}(\mathbf{Z}_{kh}^{\leftarrow})\|_2^2 \Big] \, \mathrm{d}t}_{\text{discretization error}}, \end{aligned} \tag{E.1}$$

where γ_d is the density of a standard Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{I}_d)$.
Step 1. Initial error in (E.1). Using Chen et al. (2023a, Lemma C.4), we have

$$\mathrm{KL}(q_T \| \gamma_d) \lesssim d \exp(-T). \tag{E.2}$$

Step 2. Reference score error and Doob's guidance error in (E.1). Under Assumption 2, it is apparent that $\mathbb{E}^{\mathbb{P}}[w(\mathbf{X}_T^{\leftarrow})] \geq \underline{B}$ and $\sup_{\mathbf{x}\in\mathbb{R}^d} h^*(t,\mathbf{x}) \leq \bar{B}$ for each $t \in (T_0, T)$. Hence, the density ratio is uniformly bounded

$$\sup_{\mathbf{x}\in\mathbb{R}^d} \frac{q_{T-t}(\mathbf{x})}{p_{T-t}(\mathbf{x})} = \frac{h^*(t,\mathbf{x})}{\mathbb{E}^{\mathbb{P}}[w(\mathbf{X}_T^{\leftarrow})]} \leq \frac{\bar{B}}{\underline{B}}. \tag{E.3}$$

For the reference score error term in (E.1), it follows for each $0 \leq k \leq K-1$ that

$$\begin{aligned}
&\mathbb{E}^{\mathbb{Q}}\left[\|\widehat{\mathbf{s}}(kh, \mathbf{Z}_{kh}^{\leftarrow}) - \nabla \log p_{T-kh}(\mathbf{Z}_{kh}^{\leftarrow})\|_2^2\right] \\
&= \int \|\widehat{\mathbf{s}}(kh, \mathbf{z}) - \nabla \log p_{T-kh}(\mathbf{z})\|_2^2 q_{T-kh}(\mathbf{z})\,\mathrm{d}\mathbf{z} \\
&= \int \|\widehat{\mathbf{s}}(kh, \mathbf{x}) - \nabla \log p_{T-kh}(\mathbf{x})\|_2^2 \frac{q_{T-kh}(\mathbf{x})}{p_{T-kh}(\mathbf{x})} p_{T-kh}(\mathbf{x})\,\mathrm{d}\mathbf{x} \\
&\leq \frac{\bar{B}}{\underline{B}} \int \|\widehat{\mathbf{s}}(kh, \mathbf{x}) - \nabla \log p_{T-kh}(\mathbf{x})\|_2^2 p_{T-kh}(\mathbf{x})\,\mathrm{d}\mathbf{x} \\
&= \frac{\bar{B}}{\underline{B}} \mathbb{E}^{\mathbb{P}}\left[\|\widehat{\mathbf{s}}(kh, \mathbf{X}_{kh}^{\leftarrow}) - \nabla \log p_{T-kh}(\mathbf{X}_{kh}^{\leftarrow})\|_2^2\right],
\end{aligned}$$

where the inequality holds from (E.3) and Hölder's inequality. Consequently,

$$\begin{aligned}
&\sum_{k=0}^{K-1} h\mathbb{E}^{\mathbb{Q}}\left[\|\widehat{\mathbf{s}}(kh, \mathbf{Z}_{kh}^{\leftarrow}) - \nabla \log p_{T-kh}(\mathbf{Z}_{kh}^{\leftarrow})\|_2^2\right] \\
&\leq \frac{\bar{B}}{\underline{B}} \sum_{k=0}^{K-1} h\mathbb{E}^{\mathbb{P}}\left[\|\widehat{\mathbf{s}}(kh, \mathbf{X}_{kh}^{\leftarrow}) - \nabla \log p_{T-kh}(\mathbf{X}_{kh}^{\leftarrow})\|_2^2\right] \leq T\frac{\bar{B}}{\underline{B}}\varepsilon_{\mathrm{ref}}^2,
\end{aligned} \tag{E.4}$$

where the last inequality is owing to Assumption 3. By a similar argument, we have

$$\begin{aligned}
&\sum_{k=0}^{K-1} h\mathbb{E}^{\mathbb{Q}}\left[\|\nabla \log \widehat{h}_{kh}(\mathbf{Z}_{kh}^{\leftarrow}) - \nabla \log h_{kh}^*(\mathbf{Z}_{kh}^{\leftarrow})\|_2^2\right] \\
&\leq \frac{\bar{B}}{\underline{B}} \sum_{k=0}^{K-1} h\mathbb{E}^{\mathbb{P}}\left[\|\nabla \log \widehat{h}_{kh}(\mathbf{X}_{kh}^{\leftarrow}) - \nabla \log h_{kh}^*(\mathbf{X}_{kh}^{\leftarrow})\|_2^2\right].
\end{aligned} \tag{E.5}$$

Step 3. Discretization error in (E.1). According to Chen et al. (2023a, Lemma D.1), we have

$$\mathbb{E}^{\mathbb{Q}}[\|\nabla \log q_{T-t}(\mathbf{Z}_t^{\leftarrow}) - \nabla \log q_{T-kh}(\mathbf{Z}_{kh}^{\leftarrow})\|_2^2] \lesssim \frac{dG_k T}{K}, \tag{E.6}$$

for any $t \in (kh, (k+1)h)$, provided that $\nabla \log q_{T-t}$ is G-Lipschitz for any $t \in (kh, (k+1)h)$. Then it remains to estimate the Lipschitz constant G_k. Using Lemmas C.2 and C.3 yields

$$\begin{aligned}
\nabla^2 \log q_{T-t}(\mathbf{z}) &= -\frac{1}{\sigma_{T-t}^2}\mathbf{I}_d + \frac{\mu_{T-t}}{\sigma_{T-t}^2}\nabla\mathbb{E}[\mathbf{Z}_0|\mathbf{Z}_t = \mathbf{z}] \\
&= -\frac{1}{\sigma_{T-t}^2}\mathbf{I}_d + \frac{\mu_{T-t}^2}{\sigma_{T-t}^4}\mathrm{Cov}(\mathbf{Z}_0|\mathbf{Z}_t = \mathbf{z}).
\end{aligned}$$

As a consequence, for each $0 \le k \le K-1$,

$$
\begin{aligned}
G_k &\le \sup_{t\in(T_0,T)} \sup_{\mathbf{z}\in\mathbb{R}^d} \|\nabla^2 \log q_{T-t}(\mathbf{z})\|_{\mathrm{op}} \\
&\le \sup_{t\in(T_0,T)} \frac{1}{\sigma_{T-t}^2} + \frac{\mu_{T-t}^2}{\sigma_{T-t}^4}\|\mathrm{Cov}(\mathbf{Z}_0|\mathbf{Z}_t=\mathbf{z})\|_{\mathrm{op}} \lesssim \frac{d}{\sigma_{T_0}^4},
\end{aligned}
\tag{E.7}
$$

where the first inequality holds from the triangular inequality, and the second inequality is due to the boundedness of $\mathbf{Z}_0$ under Assumptions 1 and 2. Combining (E.6) and (E.7) implies

$$
\sum_{k=0}^{K-1}\int_{kh}^{(k+1)h} \mathbb{E}^{\mathbb{Q}}\Big[\|\nabla \log q_{T-t}(\mathbf{Z}_t^{\leftarrow}) - \nabla \log q_{T-kh}(\mathbf{Z}_{kh}^{\leftarrow})\|_2^2\Big]\,\mathrm{d}t \lesssim \frac{d^2T^2}{\sigma_{T_0}^4 K}. \tag{E.8}
$$

Step 4. Conclusions. Substituting (E.2), (E.4), (E.5), (E.8) into (E.1) completes the proof. □

Corollary E.1. *Suppose Assumptions 1, 2, and 3 hold. Let $\delta \in (0,1)$. Set the hypothesis classes $\{\mathscr{H}_{T-kh}\}_{k=0}^{K=1}$ as (5.2) with the same depth L and number of non-zero parameters S. Let $\widehat{q}_{T-T_0}$ be the marginal density of $\widehat{\mathbf{Z}}_{T-T_0}^{\leftarrow}$ defined in (4.8). Then it follows that*

$$
\|q_{T_0} - \widehat{q}_{T-T_0}\|_{\mathrm{TV}}^2 \le \frac{C\delta^2}{\sigma_{T_0}^8}\log\Big(\frac{1}{\delta}\log\Big(\frac{\sigma_{T_0}^8}{\delta^2}\Big)\Big),
$$

where C is a constant depending only on d, $\bar{B}$, and $\underline{B}$, and

$$
\begin{aligned}
&T \asymp \log\Big(\frac{\sigma_{T_0}^8}{\delta^2}\Big), \quad K \gtrsim \frac{\sigma_{T_0}^4}{\delta^2}\log^2\Big(\frac{\sigma_{T_0}^8}{\delta^2}\Big), \quad h \lesssim \frac{\delta^2}{\sigma_{T_0}^4}\log^{-1}\Big(\frac{\sigma_{T_0}^8}{\delta^2}\Big) \\
&\varepsilon_{\mathrm{ref}}^2 \lesssim \frac{\delta^2}{\sigma_{T_0}^8}\log^{-1}\Big(\frac{\sigma_{T_0}^8}{\delta^2}\Big), \quad n \gtrsim \frac{1}{\delta^{d+8}}\log^{\frac{d+8}{2}}\Big(\frac{\sigma_{T_0}^8}{\delta^2}\Big).
\end{aligned}
$$

Proof of Corollary E.1. Combining Theorem 5.3 and Lemma 5.5 yields

$$
\mathrm{KL}(q_{T_0}\|\widehat{q}_{T-T_0}) \le \frac{C}{\sigma_{T_0}^8}\Big\{\underbrace{Tn^{-\frac{2}{d+8}}\log^4 n}_{\text{(i)}} + \underbrace{T\sigma_{T_0}^8\varepsilon_{\mathrm{ref}}^2}_{\text{(ii)}} + \underbrace{\sigma_{T_0}^8\exp(-T)}_{\text{(iii)}} + \underbrace{T^2\sigma_{T_0}^4\frac{1}{K}}_{\text{(iv)}}\Big\},
$$

where C is a constant depending only on d, $\bar{B}$, and $\underline{B}$. By setting

$$
\begin{aligned}
&T \asymp \log\Big(\frac{\sigma_{T_0}^8}{\delta^2}\Big), \quad K \gtrsim \frac{\sigma_{T_0}^4}{\delta^2}\log^2\Big(\frac{\sigma_{T_0}^8}{\delta^2}\Big), \\
&\varepsilon_{\mathrm{ref}}^2 \lesssim \frac{\delta^2}{\sigma_{T_0}^8}\log^{-1}\Big(\frac{\sigma_{T_0}^8}{\delta^2}\Big), \quad n \gtrsim \frac{1}{\delta^{d+8}}\log^{\frac{d+8}{2}}\Big(\frac{\sigma_{T_0}^8}{\delta^2}\Big),
\end{aligned}
$$

we find

$$
\mathrm{KL}(q_{T_0}\|\widehat{q}_{T-T_0}) \le \frac{C\delta^2}{\sigma_{T_0}^8}\log\Big(\frac{1}{\delta}\log\Big(\frac{\sigma_{T_0}^8}{\delta^2}\Big)\Big).
$$

Finally, using Pinker's inequality completes the proof. □

E.2 Convergence rate of the controllable diffusion models.

Theorem 5.6. *Suppose Assumptions 1, 2, and 3 hold. Let $\varepsilon \in (0,1)$. Set the hypothesis classes $\{\mathscr{H}_{T-kh}\}_{k=0}^{K=1}$ as (5.2) with the same depth L and number of non-zero parameters S as Theorem 5.3. Let $\widehat{q}_{T-T_0}$ be the marginal density of $\widehat{\mathbf{Z}}^{\leftarrow}_{T-T_0}$ defined in (4.8), and let $(\mathcal{M}\circ\mathcal{T}_R)_\sharp\widehat{q}_{T-T_0}$ defined as (4.9). Then it follows that*

$$\mathbb{E}\Big[\mathcal{W}_2^2(q_0,(\mathcal{M}\circ\mathcal{T}_R)_\sharp\widehat{q}_{T-T_0})\Big] \leq C\varepsilon\log^3\Big(\frac{1}{\varepsilon}\Big).$$

provided that the truncation radius R, the terminal time T, the step size h, the number of steps K, the error of reference score $\varepsilon_{\mathrm{ref}}$, the number of samples n for Doob's matching, and the early-stopping time T_0 are set, respectively, as

$$R \asymp \log^{\frac{1}{2}}\Big(\frac{1}{\varepsilon}\Big), \quad T \asymp \log\Big(\frac{1}{\varepsilon^2}\Big), \quad K \gtrsim \frac{1}{\varepsilon^4}\log^2\Big(\frac{1}{\varepsilon^2}\Big), \quad h \lesssim \varepsilon^4\log^{-1}\Big(\frac{1}{\varepsilon^2}\Big)$$
$$\varepsilon_{\mathrm{ref}}^2 \lesssim \varepsilon^2\log^{-1}\Big(\frac{1}{\varepsilon^2}\Big), \quad n \gtrsim \frac{1}{\varepsilon^{3(d+8)}}\log^{\frac{d+8}{2}}\Big(\frac{1}{\varepsilon^2}\Big).$$

Here C is a constant depending only on d, $\bar{B}$, and $\underline{B}$.

Proof of Theorem 5.6. According to the triangular inequality, we have

$$\begin{aligned}\mathcal{W}_2^2(q_0,(\mathcal{M}\circ\mathcal{T}_R)_\sharp\widehat{q}_{T-T_0}) &= 3\underbrace{\mathcal{W}_2^2(q_0,\mathcal{M}_\sharp q_{T_0})}_{\text{(i)}} + 3\underbrace{\mathcal{W}_2^2(\mathcal{M}_\sharp q_{T_0},(\mathcal{M}\circ\mathcal{T}_R)_\sharp q_{T_0})}_{\text{(ii)}} \\ &\quad + 3\underbrace{\mathcal{W}_2^2((\mathcal{M}\circ\mathcal{T}_R)_\sharp q_{T_0},(\mathcal{M}\circ\mathcal{T}_R)_\sharp\widehat{q}_{T-T_0})}_{\text{(iii)}}.\end{aligned} \tag{E.9}$$

Here the term (i) represents the early-stopping error, the term (ii) represents the truncation error, while the term (iii) represents the error of controllable diffusion models (4.8). In the rest of the proof, we bound these three errors, respectively.

Step 1. Bound the term (i) in (E.9). To estimate the 2-Wasserstein distance between the target distribution q_0 and the scaled early-stopping distribution $\mathcal{M}_\sharp q_{T_0}$, we begin by producing a coupling of them. Let $\mathbf{Z}_0 \sim q_0$, and let $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0},\mathbf{I}_d)$ be independent of $\mathbf{Z}_0$. Define $\widetilde{\mathbf{Z}}_{T_0} := \mathbf{Z}_0 + \sigma_{T_0}\mu_{T_0}^{-1}\boldsymbol{\varepsilon}$. It is apparent that $\widetilde{\mathbf{Z}}_{T_0} \sim \mathcal{M}_\sharp q_{T_0}$. Then

$$\mathcal{W}_2^2(q_0,\mathcal{M}_\sharp q_{T_0}) \leq \mathbb{E}[\|\mathbf{Z}_0 - \widetilde{\mathbf{Z}}_{T_0}\|_2^2] = \frac{\sigma_{T_0}^2}{\mu_{T_0}^2}\mathbb{E}[\|\boldsymbol{\varepsilon}\|_2^2] = \frac{d\sigma_{T_0}^2}{\mu_{T_0}^2}. \tag{E.10}$$

Step 2. Bound the term (ii) in (E.9). Let $\mathbf{Z}_{T_0} \sim q_{T_0}$. According to the definition of the truncation operator $\mathcal{T}_R$, the joint law of $(\mu_{T_0}^{-1}\mathbf{Z}_{T_0}, \mu_{T_0}^{-1}\mathbf{Z}_{T_0}\mathbb{1}_{B(\mathbf{0},R)}(\mathbf{Z}_{T_0}))$ is a coupling of $\mathcal{M}_\sharp q_{T_0}$ and $(\mathcal{M}\circ\mathcal{T}_R)_\sharp q_{T_0}$. Therefore,

$$\begin{aligned}\mathcal{W}_2^2(\mathcal{M}_\sharp q_{T_0},(\mathcal{M}\circ\mathcal{T}_R)_\sharp q_{T_0}) &\leq \mathbb{E}[\|\mu_{T_0}^{-1}\mathbf{Z}_{T_0} - \mu_{T_0}^{-1}\mathbf{Z}_{T_0}\mathbb{1}_{B(\mathbf{0},R)}(\mathbf{Z}_{T_0})\|_2^2] \\ &= \frac{1}{\mu_{T_0}^2}\int \|\mathbf{z} - \mathbf{z}\mathbb{1}_{B(\mathbf{0},R)}(\mathbf{z})\|_2^2 q_{T_0}(\mathbf{z})\,\mathrm{d}\mathbf{z} \\ &= \frac{1}{\mu_{T_0}^2}\int \|\mathbf{z}\|_2^2\mathbb{1}_{\mathbb{R}^d - B(\mathbf{0},R)}(\mathbf{z}) q_{T_0}(\mathbf{z})\,\mathrm{d}\mathbf{z} \\ &\leq \frac{1}{\mu_{T_0}^2}\mathbb{E}^{\frac{1}{2}}[\|\mathbf{Z}_{T_0}\|_2^4]\,\mathrm{Pr}^{\frac{1}{2}}\{\|\mathbf{Z}_{T_0}\|_2 > R\} \\ &\lesssim \frac{1}{\mu_{T_0}^2}d2^{d+1}\exp\Big(-\frac{R^2}{4d\mu_{T_0}^2 + 8\sigma_{T_0}^2}\Big),\end{aligned}$$

where the second ineq holds from Cauchy-Schwarz inequality, and the last inequality is due to Lemma G.3 and Corollary G.2. By setting $R^2 = (4d\mu_{T_0}^2 + 8\sigma_{T_0}^2)\log(\varepsilon^{-1})$, we have

$$\mathcal{W}_2^2(\mathcal{M}_\sharp q_{T_0}, (\mathcal{M}\circ\mathcal{T}_R)_\sharp q_{T_0}) \lesssim \frac{d2^d}{\mu_{T_0}^2}\varepsilon. \tag{E.11}$$

Step 3. Bound the term (iii) in (E.9). Let $\mathbf{Z}_{T_0}^R \sim (\mathcal{T}_R)_\sharp q_{T_0}$ and $\widehat{\mathbf{Z}}_{T_0}^R \sim (\mathcal{T}_R)_\sharp \widehat{q}_{T-T_0}$ be optimal coupled. This means

$$\mathcal{W}_2^2((\mathcal{T}_R)_\sharp q_{T_0}, (\mathcal{T}_R)_\sharp \widehat{q}_{T-T_0}) = \mathbb{E}[\|\mathbf{Z}_{T_0}^R - \widehat{\mathbf{Z}}_{T_0}^R\|_2^2]. \tag{E.12}$$

On the other hand, $\mu_{T_0}^{-1}\mathbf{Z}_{T_0}^R \sim (\mathcal{M}\circ\mathcal{T}_R)_\sharp q_{T_0}$ and $\mu_{T_0}^{-1}\widehat{\mathbf{Z}}_{T_0}^R \sim (\mathcal{M}\circ\mathcal{T}_R)_\sharp \widehat{q}_{T-T_0}$. Hence,

$$\begin{aligned} &\mathcal{W}_2^2((\mathcal{M}\circ\mathcal{T}_R)_\sharp q_{T_0}, (\mathcal{M}\circ\mathcal{T}_R)_\sharp \widehat{q}_{T-T_0}) \\ &\leq \mathbb{E}[\|\mu_{T_0}^{-1}\mathbf{Z}_{T_0}^R - \mu_{T_0}^{-1}\widehat{\mathbf{Z}}_{T_0}^R\|_2^2] = \frac{1}{\mu_{T_0}^2}\mathcal{W}_2^2((\mathcal{T}_R)_\sharp q_{T_0}, (\mathcal{T}_R)_\sharp \widehat{q}_{T-T_0}), \end{aligned} \tag{E.13}$$

where the equality holds from (E.12). Then using Villani (2009, Theorem 6.15) and the data processing inequality, we have

$$\begin{aligned} \mathcal{W}_2^2((\mathcal{T}_R)_\sharp q_{T_0}, (\mathcal{T}_R)_\sharp \widehat{q}_{T-T_0}) &= 2R^2\|(\mathcal{T}_R)_\sharp q_{T_0} - (\mathcal{T}_R)_\sharp \widehat{q}_{T-T_0}\|_{\mathrm{TV}} \\ &\leq 2R^2\|q_{T_0} - \widehat{q}_{T-T_0}\|_{\mathrm{TV}}. \end{aligned} \tag{E.14}$$

Combining (E.13) and (E.14) yields

$$\begin{aligned} \mathcal{W}_2^2((\mathcal{M}\circ\mathcal{T}_R)_\sharp q_{T_0}, (\mathcal{M}\circ\mathcal{T}_R)_\sharp \widehat{q}_{T-T_0}) &\leq \frac{2R^2}{\mu_{T_0}^2}\|q_{T_0} - \widehat{q}_{T-T_0}\|_{\mathrm{TV}} \\ &\leq \frac{2R^2}{\mu_{T_0}^2}\frac{C'\varepsilon^3}{\sigma_{T_0}^4} \\ &\leq \frac{2(4d\mu_{T_0}^2 + 8\sigma_{T_0}^2)\log(\varepsilon^{-1})}{\mu_{T_0}^2}\frac{C'\varepsilon^3}{\sigma_{T_0}^4}\log^2\Big(\frac{1}{\varepsilon}\Big), \end{aligned} \tag{E.15}$$

where C is a constant depending only on d, $\bar{B}$, and $\underline{B}$, and the second inequality holds from Corollary E.1 with $\delta = \varepsilon^3$.

Step 4. Conclusion. Substituting (E.10), (E.11), and (E.15) into (E.9) yields

$$\begin{aligned} \mathcal{W}_2^2(q_0, (\mathcal{M}\circ\mathcal{T}_R)_\sharp \widehat{q}_{T-T_0}) &\lesssim \frac{d\sigma_{T_0}^2}{\mu_{T_0}^2} + \frac{d2^d}{\mu_{T_0}^2}\varepsilon + \frac{2(4d\mu_{T_0}^2 + 8\sigma_{T_0}^2)\log(\varepsilon^{-1})}{\mu_{T_0}^2}\frac{C'\varepsilon^3}{\sigma_{T_0}^4}\log^2\Big(\frac{1}{\varepsilon}\Big) \\ &\leq C\Big\{\sigma_{T_0}^2 + \varepsilon + \frac{\varepsilon^3}{\sigma_{T_0}^4}\log^3\Big(\frac{1}{\varepsilon}\Big)\Big\}, \end{aligned}$$

where C is a constant depending only on d, $\bar{B}$, and $\underline{B}$. Letting $\sigma_{T_0}^2 \asymp \varepsilon$, i.e., $T_0 \asymp \varepsilon$, completes the proof. □

F Derivations in Section 5.4

Proposition 5.7. *Suppose Assumptions 4 and 2 hold. Then for any $t \in (0, T)$ and $\mathbf{x} \in \mathbb{R}^d$, we have*

$$h_t^*(\mathbf{x}) = \bar{h}_t^*(\mathbf{P}^\top\mathbf{x}) := \mathbb{E}[w(\mathbf{P}\bar{\mathbf{X}}_T^{\leftarrow}) \mid \bar{\mathbf{X}}_t^{\leftarrow} = \mathbf{P}^\top\mathbf{x}].$$

Proof of Proposition 5.7. According to Assumption 4, a particle $\mathbf{X}_0$ following p_0 satisfies

$$\mathbf{X}_0 \overset{\mathrm{d}}{=} \mathbf{P}\bar{\mathbf{X}}_0, \quad \bar{\mathbf{X}}_0 \sim \bar{p}_0.$$

We first establish the relations between p_t and $\bar{p}_t$. It is straightforward that

$$\begin{aligned}
p_t(\mathbf{x}) &= \int \varphi_d(\mathbf{x}; \mu_t \mathbf{x}_0, \sigma_t^2 \mathbf{I}_d) p_0(\mathbf{x}_0)\, \mathrm{d}\mathbf{x}_0 \\
&= \int \varphi_d(\mathbf{x}; \mu_t \mathbf{x}_0, \sigma_t^2 \mathbf{I}_d) \Big(\int \delta_{\mathbf{P}\bar{\mathbf{x}}_0}(\mathbf{x}_0) \bar{p}_0(\bar{\mathbf{x}}_0)\, \mathrm{d}\bar{\mathbf{x}}_0 \Big)\, \mathrm{d}\mathbf{x}_0 \\
&= \int \Big(\int \varphi_d(\mathbf{x}; \mu_t \mathbf{x}_0, \sigma_t^2 \mathbf{I}_d) \delta_{\mathbf{P}\bar{\mathbf{x}}_0}(\mathbf{x}_0)\, \mathrm{d}\mathbf{x}_0 \Big) \bar{p}_0(\bar{\mathbf{x}}_0)\, \mathrm{d}\bar{\mathbf{x}}_0 \\
&= \int \varphi_d(\mathbf{x}; \mu_t \mathbf{P}\bar{\mathbf{x}}_0, \sigma_t^2 \mathbf{I}_d) \bar{p}_0(\bar{\mathbf{x}}_0)\, \mathrm{d}\bar{\mathbf{x}}_0 \\
&= (2\pi\sigma_t^2)^{-\frac{d}{2}} \int \exp\Big(-\frac{\|\mathbf{x} - \mu_t \mathbf{P}\bar{\mathbf{x}}_0\|_2^2}{2\sigma_t^2} \Big) \bar{p}_0(\bar{\mathbf{x}}_0)\, \mathrm{d}\bar{\mathbf{x}}_0 \\
&= (2\pi\sigma_t^2)^{-\frac{d}{2}} \int \exp\Big(-\frac{\|(\mathbf{I}_d - \mathbf{P}\mathbf{P}^\top)\mathbf{x} + \mathbf{P}\mathbf{P}^\top \mathbf{x} - \mu_t \mathbf{P}\bar{\mathbf{x}}_0\|_2^2}{2\sigma_t^2} \Big) \bar{p}_0(\bar{\mathbf{x}}_0)\, \mathrm{d}\bar{\mathbf{x}}_0 \\
&= (2\pi\sigma_t^2)^{-\frac{d}{2}} \int \exp\Big(-\frac{\|(\mathbf{I}_d - \mathbf{P}\mathbf{P}^\top)\mathbf{x}\|_2^2 + \|\mathbf{P}\mathbf{P}^\top \mathbf{x} - \mu_t \mathbf{P}\bar{\mathbf{x}}_0\|_2^2}{2\sigma_t^2} \Big) \bar{p}_0(\bar{\mathbf{x}}_0)\, \mathrm{d}\bar{\mathbf{x}}_0 \\
&= (2\pi\sigma_t^2)^{-\frac{d}{2}} \exp\Big(-\frac{\|(\mathbf{I}_d - \mathbf{P}\mathbf{P}^\top)\mathbf{x}\|_2^2}{2\sigma_t^2} \Big) \int \exp\Big(-\frac{\|\mathbf{P}^\top \mathbf{x} - \mu_t \bar{\mathbf{x}}_0\|_2^2}{2\sigma_t^2} \Big) \bar{p}_0(\bar{\mathbf{x}}_0)\, \mathrm{d}\bar{\mathbf{x}}_0 \\
&= \exp\Big(-\frac{\|(\mathbf{I}_d - \mathbf{P}\mathbf{P}^\top)\mathbf{x}\|_2^2}{2\sigma_t^2} \Big) \bar{p}_t(\mathbf{P}^\top \mathbf{x}),
\end{aligned} \tag{F.1}$$

where the seventh equality invokes the fact that $(\mathbf{I}_d - \mathbf{P}\mathbf{P}^\top)\mathbf{x}$ is orthogonal to $\mathbf{P}\mathbf{P}^\top \mathbf{x} - \mu_t \mathbf{P}\bar{\mathbf{x}}_0$, the eighth equality is due to $\|\mathbf{P}\mathbf{v}\|_2 = \|\mathbf{v}\|_2$ for each $\mathbf{v} \in \mathbb{R}^d$, and the last equality used (5.3). Then by a similar argument as the density, we find

$$\begin{aligned}
&h_{T-t}^*(\mathbf{x}) = \mathbb{E}[w(\mathbf{X}_0) \mid \mathbf{X}_t = \mathbf{x}] \\
&= \frac{1}{p_t(\mathbf{x})} \int w(\mathbf{x}_0) \varphi_d(\mathbf{x}; \mu_t \mathbf{x}_0, \sigma_t^2 \mathbf{I}_d) p_0(\mathbf{x}_0)\, \mathrm{d}\mathbf{x}_0 \\
&= \frac{1}{p_t(\mathbf{x})} \int w(\mathbf{P}\bar{\mathbf{x}}_0) \varphi_d(\mathbf{x}; \mu_t \mathbf{P}\bar{\mathbf{x}}_0, \sigma_t^2 \mathbf{I}_d) \bar{p}_0(\bar{\mathbf{x}}_0)\, \mathrm{d}\bar{\mathbf{x}}_0 \\
&= \frac{1}{p_t(\mathbf{x})} (2\pi\sigma_t^2)^{-\frac{d}{2}} \int w(\mathbf{P}\bar{\mathbf{x}}_0) \exp\Big(-\frac{\|\mathbf{x} - \mu_t \mathbf{P}\bar{\mathbf{x}}_0\|_2^2}{2\sigma_t^2} \Big) \bar{p}_0(\bar{\mathbf{x}}_0)\, \mathrm{d}\bar{\mathbf{x}}_0 \\
&= \frac{1}{p_t(\mathbf{x})} (2\pi\sigma_t^2)^{-\frac{d}{2}} \int w(\mathbf{P}\bar{\mathbf{x}}_0) \exp\Big(-\frac{\|(\mathbf{I}_d - \mathbf{P}\mathbf{P}^\top)\mathbf{x} + \mathbf{P}\mathbf{P}^\top \mathbf{x} - \mu_t \mathbf{P}\bar{\mathbf{x}}_0\|_2^2}{2\sigma_t^2} \Big) \bar{p}_0(\bar{\mathbf{x}}_0)\, \mathrm{d}\bar{\mathbf{x}}_0 \\
&= \frac{1}{p_t(\mathbf{x})} \exp\Big(-\frac{\|(\mathbf{I}_d - \mathbf{P}\mathbf{P}^\top)\mathbf{x}\|_2^2}{2\sigma_t^2} \Big) \int w(\mathbf{P}\bar{\mathbf{x}}_0) \varphi_d(\mathbf{P}^\top \mathbf{x}; \mu_t \bar{\mathbf{x}}_0, \sigma_t^2 \mathbf{I}_d) \bar{p}_0(\bar{\mathbf{x}}_0)\, \mathrm{d}\bar{\mathbf{x}}_0 \\
&= \frac{1}{\bar{p}_t(\mathbf{P}^\top \mathbf{x})} \int w(\mathbf{P}\bar{\mathbf{x}}_0) \varphi_d(\mathbf{P}^\top \mathbf{x}; \mu_t \bar{\mathbf{x}}_0, \sigma_t^2 \mathbf{I}_d) \bar{p}_0(\bar{\mathbf{x}}_0)\, \mathrm{d}\bar{\mathbf{x}}_0 \\
&= \mathbb{E}[w(\mathbf{P}\bar{\mathbf{X}}_0) \mid \bar{\mathbf{X}}_t = \mathbf{P}^\top \mathbf{x}] = \mathbb{E}[w(\mathbf{P}\bar{\mathbf{X}}_T^{\leftarrow}) \mid \bar{\mathbf{X}}_{T-t}^{\leftarrow} = \mathbf{P}^\top \mathbf{x}],
\end{aligned}$$

where the second and the eighth equalities are due to Bayes' rule, the seventh equality holds from (F.1). This completes the proof. ∎

Proposition 5.8. *Suppose Assumptions 4 and 2 hold. Then for all $t \in (0, T)$ and $\bar{\mathbf{x}} \in \mathbb{R}^{d^*}$, the following bounds hold:*

(i) $\underline{B} \le \bar{h}_t^*(\bar{\mathbf{x}}) \le \bar{B}$;

(ii) $\max_{1\le k\le d} |D_k \bar{h}_t^*(\bar{\mathbf{x}})| \le 2\sigma_{T-t}^{-2}\bar{B}$; *and*

(iii) $\max_{1\le k,\ell\le d} |D_{k\ell}^2 \bar{h}_t^*(\bar{\mathbf{x}})| \le 6\sigma_{T-t}^{-4}\bar{B}$,

where D_k and $D_{k\ell}^2$ denote the first-order and second-order partial derivatives with respect to the input coordinates, respectively.

Proof of Proposition 5.8. By the simialr argument as Lemmas C.1, C.4, and C.5, we conclude the desired results. □

By a similar argument as Lemma D.8, we have the following approximation error bounds for low-dimensional Doob's h-function $\bar{h}_t^*$ (5.4).

Lemma F.1 (Approximation error)**.** *Suppose Assumptions 4 and 2 hold. Let $R \ge 1$, and let the hypothesis class $\mathscr{H}_t$ be defined as (5.5) with $L \le C \log N$ and $S \le N^{d^*}$, then*

$$\|h_t - \bar{h}_t^*\|_{L^2(p_{T-t})}^2 \le C\frac{\bar{B}^2 \log^4 N}{\sigma_{T-t}^8 N^4},$$
$$\|\nabla h_t - \nabla \bar{h}_t^*\|_{L^2(p_{T-t})}^2 \le C\frac{\bar{B}^2 \log^2 N}{\sigma_{T-t}^8 N^2}.$$

provided that $R^2 = (4d\mu_t^2 + 8\sigma_t^2)\log N^4$, where C is a constant only depending on d^*.

Theorem 5.9. *Suppose Assumptions 4 and 2 hold. Let $t \in (0, T)$. Set the hypothesis class $\mathscr{H}_t$ as*

$$\mathscr{H}_t := \left\{ h_t \in N(L, S) : \begin{array}{c} \sup_{\mathbf{x}\in\mathbb{R}^d} h_t(\mathbf{x}) \le \bar{B}, \ \inf_{\mathbf{x}\in\mathbb{R}^d} h_t(\mathbf{x}) \ge \underline{B}, \\ \max_{1\le k\le d} \sup_{\mathbf{x}\in\mathbb{R}^d} |D_k h_t(\mathbf{x})| \le 2\sigma_{T-t}^{-2}\bar{B} \end{array} \right\},$$

where $L = \mathcal{O}(\log n)$ and $S = \mathcal{O}(n^{\frac{d^*}{d^*+8}})$. Let $\widehat{h}_t^\lambda$ be the gradient-regularized empirical risk minimizer defined as (4.6), and let h_t^* be the Doob's h-function defined as (3.7). Then the following inequality holds:

$$\mathbb{E}\Big[\|\nabla \log \widehat{h}_t^\lambda - \nabla \log h_t^*\|_{L^2(p_{T-t})}^2\Big] \le C\sigma_{T-t}^{-8} n^{-\frac{2}{d^*+8}} \log^4 n,$$

provided that the regularization parameter λ is set as $\lambda = \mathcal{O}(n^{-\frac{2}{d^*+8}})$, where C is a constant depending only on d^*, $\bar{B}$, and $\underline{B}$.

Proof of Theorem 5.9. Using the same arguments as the proof of Theorem 5.3 and applying Lemma F.1 completes the proof. □

G Auxilary Lemmas

Lemma G.1. *Suppose Assumption 1 holds. Let* $\mathbf{X}_t \sim p_t$*. Then for each* $\xi > 0$,

$$\Pr\left\{\|\mathbf{X}_t\|_2 \geq \xi\right\} \leq 2^{d+1} \exp\Big(-\frac{\xi^2}{4d\mu_t^2 + 8\sigma_t^2}\Big).$$

Proof of Lemma G.1. According to Assumption 1, we have

$$\mathbb{E}\Big[\exp\Big(\frac{\|\mu_t \mathbf{X}_0\|_2^2}{2d\mu_t^2}\Big)\Big] \leq 2. \tag{G.1}$$

Let $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{I}_d)$. Then it follows that

$$\begin{aligned}
\mathbb{E}\Big[\exp\Big(\frac{\|\sigma_t \boldsymbol{\varepsilon}\|_2^2}{4\sigma_t^2}\Big)\Big] &= \mathbb{E}\Big[\exp\Big(\frac{\|\boldsymbol{\varepsilon}\|_2^2}{4}\Big)\Big] \\
&= (2\pi)^{-\frac{d}{2}} \int \exp\Big(\frac{\|\boldsymbol{\varepsilon}\|_2^2}{4}\Big) \exp\Big(-\frac{\|\boldsymbol{\varepsilon}\|_2^2}{2}\Big)\, \mathrm{d}\boldsymbol{\varepsilon} \\
&= (2\pi)^{-\frac{d}{2}} \int \exp\Big(-\frac{\|\boldsymbol{\varepsilon}\|_2^2}{4}\Big)\, \mathrm{d}\boldsymbol{\varepsilon} \leq 2^d. \qquad \text{(G.2)}
\end{aligned}$$

Notice that $\mathbf{X}_t \stackrel{\mathrm{d}}{=} \mu_t \mathbf{X}_0 + \sigma_t \boldsymbol{\varepsilon}$, where $\mathbf{X}_0 \sim p_0$ and $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{I}_d)$ are independent. Therefore,

$$\begin{aligned}
\mathbb{E}\Big[\exp\Big(\frac{\|\mathbf{X}_t\|_2^2}{4d\mu_t^2 + 8\sigma_t^2}\Big)\Big] &= \mathbb{E}\Big[\exp\Big(\frac{\|\mu_t \mathbf{X}_0 + \sigma_t \boldsymbol{\varepsilon}\|_2^2}{4d\mu_t^2 + 8\sigma_t^2}\Big)\Big] \\
&\leq \mathbb{E}\Big[\exp\Big(\frac{\|\mu_t \mathbf{X}_0\|_2^2}{2d\mu_t^2 + 4\sigma_t^2} + \frac{\|\sigma_t \boldsymbol{\varepsilon}\|_2^2}{2d\mu_t^2 + 4\sigma_t^2}\Big)\Big] \\
&\leq \mathbb{E}\Big[\exp\Big(\frac{\|\mu_t \mathbf{X}_0\|_2^2}{2d\mu_t^2 + 4\sigma_t^2}\Big)\Big] \mathbb{E}\Big[\exp\Big(\frac{\|\sigma_t \boldsymbol{\varepsilon}\|_2^2}{2d\mu_t^2 + 4\sigma_t^2}\Big)\Big] \\
&\leq \mathbb{E}\Big[\exp\Big(\frac{\|\mu_t \mathbf{X}_0\|_2^2}{2d\mu_t^2}\Big)\Big] \mathbb{E}\Big[\exp\Big(\frac{\|\sigma_t \boldsymbol{\varepsilon}\|_2^2}{4\sigma_t^2}\Big)\Big] \leq 2^{d+1}, \qquad \text{(G.3)}
\end{aligned}$$

where the the first inequality follows from Cauchy-Schwarz inequality, the second inequality holds from the independence of $\mathbf{X}_0$ and $\boldsymbol{\varepsilon}$, and the last inequality is due to (G.1) and (G.2). Then we aim to bound the tail probability. For each $\xi > 0$, we have

$$\begin{aligned}
\Pr\left\{\|\mathbf{X}_t\|_2 \geq \xi\right\} &= \Pr\Big\{\frac{\|\mathbf{X}_t\|_2^2}{4d\mu_t^2 + 8\sigma_t^2} \geq \frac{\xi^2}{4d\mu_t^2 + 8\sigma_t^2}\Big\} \\
&= \Pr\Big\{\exp\Big(\frac{\|\mathbf{X}_t\|_2^2}{4d\mu_t^2 + 8\sigma_t^2}\Big) \geq \exp\Big(\frac{\xi^2}{4d\mu_t^2 + 8\sigma_t^2}\Big)\Big\} \\
&\leq \exp\Big(-\frac{\xi^2}{4d\mu_t^2 + 8\sigma_t^2}\Big) \mathbb{E}\Big[\exp\Big(\frac{\|\mathbf{X}_t\|_2^2}{4d\mu_t^2 + 8\sigma_t^2}\Big)\Big] \\
&\leq 2^{d+1} \exp\Big(-\frac{\xi^2}{4d\mu_t^2 + 8\sigma_t^2}\Big),
\end{aligned}$$

where the first inequality invokes Markov's inequality, and the last inequality is due to (G.3). This completes the proof. □

Corollary G.2. *Suppose Assumptions 1 and 2 hold. Let* $\mathbf{Z}_t \sim q_t$*. Then for each* $\xi > 0$,

$$\Pr\left\{\|\mathbf{Z}_t\|_2 \geq \xi\right\} \leq 2^{d+1} \exp\Big(-\frac{\xi^2}{4d\mu_t^2 + 8\sigma_t^2}\Big).$$

Proof of Corollary G.2. Under Assumptions 1 and 2, $\mathrm{supp}(q_0) = \mathrm{supp}(p_0)$. The same argument as Lemma G.1 completes the proof. □

Lemma G.3. *Suppose Assumptions 1 and 2 hold. Let* $\mathbf{Z}_t \sim q_t$*. Then for each* $\xi > 0$,

$$\mathbb{E}[\|\mathbf{Z}_t\|_2^4] \lesssim d^2.$$

Proof of Lemma G.3. Let $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{I}_d)$. It is straightforward that

$$\mathbb{E}[\|\boldsymbol{\varepsilon}\|_2^4] = 4\Gamma\Big(\frac{d+4}{2}\Big)\Gamma\Big(\frac{d}{2}\Big) \leq (d+4)^2.$$

Since $\mathbf{Z}_t \overset{\mathrm{d}}{=} \mu_t \mathbf{Z}_0 + \sigma_t \boldsymbol{\varepsilon}$ with $\mathbf{Z}_0 \sim q_0$ independent of $\boldsymbol{\varepsilon}$, it follows from the triangular inequality that

$$\begin{aligned}\mathbb{E}[\|\mathbf{Z}_t\|_2^4] &\leq 8\mu_t^4 \mathbb{E}[\|\mathbf{Z}_0\|_2^4] + 8\sigma_t^4 \mathbb{E}[\|\boldsymbol{\varepsilon}\|_2^4] \\ &\leq 8(d^2 + (d+4)^2),\end{aligned}$$

where we used the fact that $\mu_t, \sigma_t \leq 1$, and $\mathrm{supp}(q_0) = \mathrm{supp}(p_0)$ under Assumptions 1 and 2. This completes the proof. □