A STOCHASTIC ANALYSIS APPROACH TO CONDITIONAL DIFFUSION GUIDANCE

WENPIN TANG AND RENYUAN XU

Abstract.

Contents

1. Introduction

Recently, there has been growing interest in guiding, or fine tuning pretrained diffusion models for specific purposes, e.g., aesthetic quality of images, functional property of proteins, and downstream tasks in operations and management. Existing approaches include:

- supervised fine-tuning with regularization [35, 50, 55] (or *soft guidance*);
- supervised fine-tuning by conditioning [14, 27] (or *conditional guidance*);
- reinforcement learning [18, 19, 23, 38, 65, 66];
- Diffusion-DPO (Direct Preference Optimization) [58, 63].

In this paper, we focus on the supervised fine-tuning by (endogenous) conditioning. The idea is closely related to Doob's h-transform, and the key is to learn the function h. Other h transform papers [13, 16] rely on the stochastic control approach, while our viewpoint is probability-theoretic (martingale, quadratic variation, etc.)

This work is in the context of *classifier guidance* [14], see also [64] for statistical theory, and [3, 43] for further applications. It is crucial in the recent development of reinforcement learning from human feedback (RLHF) [40]. There is also work on *classifier-free guidance* [27] (see [61] for a study). Refer to [10, 42, 54, 60] for literature reviews of fine-tuning, or guiding diffusion models.

Date: December 29, 2024.

This work has also connections with *pluralistic alignment* [48] [1], in which a multidimensional reward function is allowed for diffusion guidance (see Section 3.3). This is also related to *multi-task learning.*(?)

Application to rare-event simulation [2, 8], which avoids importance sampling [6, 7]. There is also recent work of diffusion models in operations research/simulation [41].

Organization of the paper: The remainder of the paper is organized as follows. We start with background on diffusion models in Section 2. In Section 3, we build the foundations for conditional diffusion guidance, leading to novel methodologies. In Section 4, we provide theoretical results of the proposed methodologies. Numerical experiments are reported in Section 5. We conclude with Section 6.

Notations: Below we collect a few notations that will be used throughout.

- $\mathbb{R}$ is the set of real numbers.
- For $x, y \in \mathbb{R}^d$, $x \cdot y$ denotes the scalar product of x and y, and $|x| := \sqrt{x \cdot x}$ is the Euclidean norm of x.
- For $A = (a_{ij})_{1 \le i,j \le d}$ a matrix, $|A|_F := \sqrt{\sum_{i,j=1}^d a_{ij}^2}$ is the Frobenius norm of A.
- For f a function on X, $|f|_\infty := \sup_X |f(x)|$ denotes its sup-norm.
- For $f : [0, \infty) \times \mathbb{R}^d \ni (t, x) \to \mathbb{R}$, $\partial_t f$ denotes its time derivative, ∇f is the gradient of f and $\partial_k f := \frac{\partial f}{\partial x_k}$ its k^{th} coordinate, and $\Delta f := \sum_{k=1}^d \frac{\partial^2 f}{\partial x_k^2}$ is the Laplacian of f.
- For Z a random variable, $\mathbb{E}Z$ denotes the expectation of Z.
- For $p(\cdot)$ and $q(\cdot)$ two probability distributions, $d_{TV}(p(\cdot), q(\cdot)) := \sup_A |p(A) - q(A)|$ is the total variation distance between $p(\cdot)$ and $q(\cdot)$; $d_{KL}(p(\cdot), q(\cdot)) := \int \log\left(\frac{dp}{dq}\right) dp$ is the KL divergence between $p(\cdot)$ and $q(\cdot)$; and $W_2(p(\cdot), q(\cdot)) := \sqrt{\inf_\gamma \mathbb{E}_{(X,Y)\sim\gamma} |X - Y|^2}$, where the infimum is taken over all couplings γ of $p(\cdot)$ and $q(\cdot)$, is the Wasserstein-2 distance between $p(\cdot)$ and $q(\cdot)$.

We use C for a generic constant whose values may change from line to line.

2. Background on diffusion models

This section provides preliminaries of score-based diffusion models. We follow the presentation of [52]. Diffusion models rely on a forward-backward procedure: the forward process transforms the target data to noise, and the backward process recovers the data from noise.

Let $p_{\text{data}}(\cdot)$ be the target data distribution. Fixing $T > 0$, the forward process $\{X_t\}_{0 \le t \le T}$ is governed by the stochastic differential equation (SDE):

$$dX_t = f(t, X_t)dt + g(t)dW_t, \quad X_0 \sim p_{\text{data}}(\cdot), \tag{2.1}$$

where $f : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$, $g : \mathbb{R}_+ \to \mathbb{R}_+$, and $\{W_t\}_{t \ge 0}$ is Brownian motion in $\mathbb{R}^d$. Some conditions on $f(\cdot, \cdot)$ and $g(\cdot)$ are required so that the SDE (2.1) is well-defined, and that X_t

[1] Pluralistic alignment refers to multi-objective alignment in an attempt to integrate complex, often conflicting, real-world values. Refer to `https://pluralistic-alignment.github.io/` for recent efforts in developing pluralistic alignment techniques.

has a smooth probability density $p(t,x) := \mathbb{P}(X_t \in dx)/dx$ (see [49]). By the time reversal formula [1, 25], let

$$d\overline{X}_t = \left(-f(T-t,\overline{X}_t) + g^2(T-t)\nabla \log p(T-t,\overline{X}_t)\right) dt + g(T-t)dB_t, \quad \overline{X}_0 \sim p(T,\cdot), \tag{2.2}$$

where $\{B_t\}_{t\geq 0}$ is an independent Brownian motion in $\mathbb{R}^d$. The processes $\{\overline{X}_t\}_{0\leq t\leq T}$ and $\{X_{T-t}\}_{0\leq t\leq T}$ have the same distribution, so the output $\overline{X}_T \sim p_{\text{data}}(\cdot)$.

The main obstacle in sampling the process $\overline{X}$ is that the *score function* $\nabla \log p(\cdot,\cdot)$, which depends on $p_{\text{data}}(\cdot)$, is not available. Moreover, the initialization $\overline{X}_0 \sim p(T,\cdot)$ also relies on $p_{\text{data}}(\cdot)$ in each sample generation. The idea of score-based diffusion models is to learn the score function via a parametrized family of functions $\{s_\theta(t,x)\}_\theta$ (e.g., neural networks), with a limited number of samples from $p_{\text{data}}(\cdot)$ (see [26, 46, 47]). The resulting backward process $\{Y_t\}_{0\leq t\leq T}$ for sampling is:

$$dY_t = \overline{f}(t,Y_t)dt + \overline{g}(t)dB_t, \quad Y_0 \sim p_{\text{noise}}(\cdot), \tag{2.3}$$

where $\overline{g}(t) := g(T-t)$ and $\overline{f}(t,y) := -f(T-t,y) + g(T-t)^2 s_{\theta_*}(T-t,y)$. Here,

- $p_{\text{noise}}(\cdot)$ is a proxy to $p(T,\cdot)$ for generating the target distribution from noise, which should *not* depend on $p_{\text{data}}(\cdot)$. The form of $p_{\text{noise}}(\cdot)$ is related to the design of the diffusion model, i.e., the pair $(f(\cdot,\cdot), g(\cdot))$. Popular examples include *variance exploding* (VE) model [34] with $f(t,x) = 0$, $g(t) = \sqrt{2t+\epsilon}$ for some (small) $\epsilon > 0$, and $p_{\text{noise}}(\cdot) = \mathcal{N}(0,T^2 I)$, and *variance preserving* (VP) model [47] with $f(t,x) = -\frac{1}{2}\left(a + \frac{(b-a)t}{T}\right)x$, $g(t) = \sqrt{a + \frac{(b-a)t}{T}}$ for some $b > a > 0$, and $p_{\text{noise}}(\cdot) = \mathcal{N}(0,I)$.
- $\{s_\theta(t,x)\}_\theta$ are function approximations to the score $\nabla \log p(t,x)$, which are trained by solving some stochastic optimization problem. This technique is known as *score matching*, and the learned $s_{\theta_*}(t,x)$ is the *score matching function*. There are several existing score matching methods, among which the most widely used one is the *denoising score matching* (DSM) [30, 57]:

$$\min_\theta \mathbb{E}_{t\sim \text{Unif}\,[0,T]} \left\{ \lambda(t)\, \mathbb{E}_{X_0 \sim p_{data}(\cdot)} \left[\mathbb{E}_{X_t|X_0} \left| s_\theta(t,X_t) - \nabla \log p(t,X_t|X_0)\right|^2 \right] \right\},$$

  where $\lambda : \mathbb{R}_+ \to \mathbb{R}_+$ is a weight function.

Most successful diffusion models require substantial training effort for score matching, and are therefore referred to as the *pretrained models* [2]. Let the process (2.3) be such a pretrained model. Denote by $P_{[0,T]}(\cdot)$ the distribution of the process $\{Y_t\}_{0\leq t\leq T}$ on the path space, and by $P_t(\cdot)$ its marginal distribution at Y_t. Write

$$p_{\text{pre}}(\cdot) := P_T(\cdot),$$

for the output of the pretrained model (2.3). A good model is expected to generate reliable samples such that $p_{\text{pre}}(\cdot) \approx p_{\text{data}}(\cdot)$. In fact, one can quantify how close $p_{\text{pre}}(\cdot)$ and $p_{\text{data}}(\cdot)$ are under suitable conditions on the model $\{f(\cdot,\cdot), g(\cdot), T, p_{\text{noise}}(\cdot), s_\theta(\cdot,\cdot)\}$, and the target distribution $p_{\text{data}}(\cdot)$. The following proposition bounds the total variation and the Wasserstein distance between $p_{\text{pre}}(\cdot)$ and $p_{\text{data}}(\cdot)$ for the aforementioned VE and VP models.

[2]It was shown in [34] that a wide class of diffusion models can be derived from the VE model by reparameterization. So practically, it suffices to pretrain a good VE model (i.e., learn the score function of the VE model suitably well).

Proposition 2.1. *Assume that* $\mathbb{E}_{p_{data}(\cdot)}|X|^2 < \infty$, *and that the score matching satisfies:*

$$\sup_{0 \le t \le T} \mathbb{E}_{p(t,\cdot)} |s_{\theta_*}(t, X) - \nabla \log p(t, X)|^2 \le \varepsilon^2_{\mathit{VP}} \quad (\mathit{resp.}\ \varepsilon^2_{\mathit{VE}}), \tag{2.4}$$

for the VE (resp. VP) model.

(1) *(Total variation)* [11, 52] *There are* $C_{VE}, C_{VP} > 0$ *(independent of* T*) such that*

$$d_{TV}\left(p_{pre}(\cdot), p_{data}(\cdot)\right) \le \begin{cases} C_{\mathit{VE}}\left(T^{-1}\sqrt{\mathbb{E}_{p_{data}(\cdot)}|X|^2} + \varepsilon_{\mathit{VE}}\sqrt{T}\right) & \text{for VE}, \\ C_{\mathit{VP}}\left(e^{-C_{\mathit{VP}}T}\sqrt{\mathbb{E}_{p_{data}(\cdot)}|X|^2} + \varepsilon_{\mathit{VP}}\sqrt{T}\right) & \text{for VP}. \end{cases} \tag{2.5}$$

(2) *(Wasserstein)* [22, 52] *Assume further that* $p_{data}(\cdot)$ *is* κ*-strongly log-concave* [3] *for* κ *sufficiently large. There are* $C_{VE}, C_{VP} > 0$ *(independent of* T*) such that*

$$W_2\left(p_{pre}(\cdot), p_{data}(\cdot)\right) \le \begin{cases} C_{\mathit{VE}}\left(T^{-1}\sqrt{\mathbb{E}_{p_{data}(\cdot)}|X|^2} + \varepsilon_{\mathit{VE}}T^2\right) & \text{for VE}, \\ C_{\mathit{VP}}\left(e^{-C_{\mathit{VP}}T}\sqrt{\mathbb{E}_{p_{data}(\cdot)}|X|^2} + \varepsilon_{\mathit{VP}}\right) & \text{for VP}. \end{cases} \tag{2.6}$$

The condition (2.4) is referred to as the blackbox score matching error, which was assumed in the most recent literature on the convergence of diffusion models [11, 22, 37, 39, 51]. There has also been a line of work [9, 24, 36, 59] on the quantitative rate of score matching, which requires a low-dimensional structure of $\{s_\theta(t, x)\}_\theta$. However, existing score-based diffusion models rely on very deep neural nets [34, 47]. So we adopt the blackbox score matching assumption (2.4) in this paper.

3. Diffusion guidance by conditioning

In this section, we build the foundations for *conditional diffusion guidance* in the context of the classifier guidance [14].

For a set $S \subset \mathbb{R}^d$ (which we call the *guidance set*), denote by $p^S_{\text{data}}(\cdot)$ the conditional target distribution on S. That is, for $\mathcal{Z} \sim p_{\text{data}}(\cdot)$,

$$(\mathcal{Z} \mid \mathcal{Z} \in S) \sim p^S_{\text{data}}(\cdot).$$

Typically, the set S corresponds to certain discrete labels or classifiers of the target data. For ease of presentation, we assume that the set S is nice enough (e.g., Borel and non-negligeable) so that $p^S_{\text{data}}(\cdot)$ is well-defined.

The problem of interest is to exploit the pretrained models to generate samples that approximate the conditional distribution $p^S_{\text{data}}(\cdot)$. Recall that $p_{\text{pre}}(\cdot)$ is the output distribution of the pretrained model (2.3), and denote by $p^S_{\text{pre}}(\cdot)$ its conditioning on S. According to the discussions in Section 2, a good pretrained model yields $p_{\text{pre}}(\cdot) \approx p_{\text{data}}(\cdot)$. This would imply $p^S_{\text{pre}}(\cdot) \approx p^S_{\text{data}}(\cdot)$ under mild assumptions on the guidance set S (see Section 4). So our focus is to use the pretrained model to sample $p^S_{\text{pre}}(\cdot)$, or more precisely, to sample

$$\{Y^S_t\}_{0 \le t \le T} \sim P_{[0,T]}(Y \mid Y_T \in S). \tag{3.1}$$

[3] A smooth function $\ell : \mathbb{R}^d \to \mathbb{R}$ is κ-strongly log-concave if $(\nabla \log \ell(x) - \nabla \log \ell(y)) \cdot (x - y) \le -\kappa |x - y|^2$ for all x, y.

3.1. **Conditional diffusion sampling.** We start by discussing how the process $\{Y_t^S\}_{0\le t\le T}$ is generated. For $t \ge 0$, let

$$h(t,y) := P_{[0,T]}(Y_T \in S \,|\, Y_t = y). \tag{3.2}$$

By the Bayes rule, we have for $s > 0$,

$$\begin{aligned} P_{[0,T]}(Y_{t+s} = y' \,|\, Y_t = y,\, Y_T \in S) &= \frac{P_{[0,T]}(Y_{t+s} = y', Y_T \in S \,|\, Y_t = x)}{P_{[0,T]}(Y_T \in S \,|\, Y_t = x)} \\ &= \frac{h(t+s,y')}{h(t,y)} P_{[0,T]}(Y_{t+s} = y' \,|\, Y_t = y), \end{aligned} \tag{3.3}$$

which is known as *Doob's h-transform* [4]. The h-transform (3.3) reveals a change of measure between the processes $\{Y_t\}_{0\le t\le T}$ and $\{Y_t^S\}_{0\le t\le T}$. As a result, $\{Y_t^S\}_{0\le t\le T}$ is also a diffusion process, whose dynamics is given in the following proposition.

Proposition 3.1. [45, IV.39] *The distribution of $\{Y_t^S\}_{0\le t\le T}$ is governed by:*

$$dY_t^S = \left(\overline{f}(t, Y_t^S) + \overline{g}(t)^2 \nabla \log h(t, Y_t^S)\right) dt + \overline{g}(t) dB_t, \quad Y_0^S \sim p_{noise}(\cdot). \tag{3.4}$$

The proof of this proposition is a direct application of (3.3) and Girsanov's theorem. As pointed out by [55, Appendix B], the dynamics (3.4) can be viewed as the classifier-guided diffusion sampling [14] in continuous time. Denote by $P^S_{[0,T]}(\cdot)$ the distribution of the process $\{Y_t^S\}_{0\le t\le T}$, and $P_t^S(\cdot)$ its marginal distribution at Y_t^S. Now by running the process (3.4), we get a sample $Y_T^S \sim P_T^S(\cdot) := p^S_{\text{pre}}(\cdot)$. The terms $\overline{f}(\cdot,\cdot), \overline{g}(\cdot)$ are given by the pretrained model (2.3). Only the term $\nabla \log h(\cdot,\cdot)$ in (3.4) is unknown, which requires to be learned.

3.2. **Learning h function.** As explained in Section 3.1, the key (and obstacle) in sampling $\{Y_t^S\}_{0\le t\le T}$ is to learn the h function, or more concretely, its logarithmic derivative $\nabla \log h$. Here our goal is to develop principled approaches to learn the h function, which combined with the dynamics (3.4) yields an implementation of the diffusion guidance $\{Y_t^S\}_{0\le t\le T}$. The main tools are from stochastic analysis: martingale theory and quadratic variation of stochastic processes.

3.2.1. *Learning h via martingale loss.* First, as mentioned in the footnote [4], $h(t,y) := P_{[0,T]}(Y_T \in S \,|\, Y_t = y)$ is harmonic with respect to the generator of $\{Y_t\}_{0\le t\le T}$:

$$\partial_t h + \overline{f}(t,y) \cdot \nabla h + \frac{1}{2}\overline{g}(t)^2 \Delta h = 0 \quad (\text{and } \; h(T,\cdot) = 1(\cdot \in S)). \tag{3.5}$$

By applying Itô's formula to $h(t, Y_t)$, we get:

$$dh(t, Y_t) = \overline{g}(t) \nabla h(t, Y_t) \cdot dB_t, \tag{3.6}$$

which leads to the following classical result.

Proposition 3.2. *Let $\{Y_t\}_{0\le t\le T}$ be the pretrained model defined by* (2.3)*, and $h(\cdot,\cdot)$ be defined by* (3.2)*. Then the process $\{h(t,Y_t)\}_{0\le t\le T}$ is a local martingale.*

[4]Doob's h-transform is a general concept of conditioning a Markov process, provided that h is harmonic with respect to the Markov generator. In our setting, the h function (3.2) arises as a special case by conditioning on the terminal data. The "bridge" calculation in (3.3) can be understood in terms of transition densities, see [20, Section 2] for a justification.

A natural idea is to exploit the (local) martingale property [5] of $\{h(t, Y_t)\}_{0\le t\le T}$ to learn the function h. Assume that $\{h(t, Y_t)\}_{0\le t\le T}$ is a (true) martingale. By the L^2 projection of conditional expectations, the h function (3.2) uniquely solves the optimization problem:

$$\min_{\ell(\cdot,\cdot)} \mathbb{E}_{[0,T]} \int_0^T (\ell(t, Y_t) - 1(Y_T \in S))^2 dt.$$

Now we restrict our search within a class of parametrized functions $\{h_\phi(t, y)\}_\phi$ to approximate $h(t, y)$. This leads to the *martingale loss* objective to learn the h function:

$$\min_{\phi} \mathbb{E}_{[0,T]} \int_0^T (h_\phi(t, Y_t) - 1(Y_T \in S))^2 dt = \mathbb{E}_{t\sim \mathrm{Unif}\,[0,T]} \left\{\mathbb{E}_{[0,T]}(h_\phi(t, Y_t) - 1(Y_T \in S))^2\right\}. \tag{3.7}$$

Denote by $h_{\phi_*}(t, y)$ the learned h function by solving the stochastic optimization problem (3.7). The algorithm for *conditional diffusion guidance via martingale loss* (CDG-ML) is summarized as follows.

Algorithm A Conditional diffusion guidance via martingale loss (CDG-ML)

Input: pretrained model $\{Y_t\}_{0\le t\le T}$ (2.3), guidance set S, parametrized family $\{h_\phi(t, y)\}_\phi$

Step 1. Solve the stochastic optimization problem (3.7) that outputs ϕ_*.

Step 2. Sample

$$\begin{aligned} dY_t^S &= \left(\overline{f}(t, Y_t^S) + \overline{g}(t)^2 \nabla \log h_{\phi_*}(t, Y_t^S)\right) dt + \overline{g}(t) dB_t \\ &= \left(\overline{f}(t, Y_t^S) + \overline{g}(t)^2 \frac{\nabla h_{\phi_*}(t, Y_t^S)}{h_{\phi_*}(t, Y_t^S)}\right) dt + \overline{g}(t) dB_t, \quad Y_0^S \sim p_{\mathrm{noise}}(\cdot). \end{aligned}$$

Output: Y_T^S.

The logic of Algorithm A is that if h_{ϕ_*} is a good approximation to h, then so is $\nabla \log h_{\phi_*}$ to $\nabla \log h$. This is not always true mathematically, but can still serve as a simple computational proxy to $\nabla \log h$.

3.2.2. *Learning ∇h via quadratic variation.* Scrutinizing the conditional guided process (3.4), it is the logarithmic derivative $\nabla \log h$ (rather than the h function itself) that needs to be learned. Noting that $\nabla \log h = \frac{\nabla h}{h}$, a more reasonable approach is to learn the numerator ∇h and the denominator h separately. We have seen that the denominator h can be learned by means of the martingale loss (3.7). Now we explain how to learn its gradient ∇h by exploiting the quadratic variation of $\{h(t, Y_t)\}_{0\le t\le T}$. Recall from (3.6) that

$$dh(t, Y_t) = \overline{g}(t) \sum_{k=1}^d \partial_k h(t, Y_t) dB_t^k,$$

where $\{B_t^k\}_{0\le t\le T}$ is the k-th coordinate of $\{B_t\}_{0\le t\le T}$. Also denote by $\{Y_t^k\}_{0\le t\le T}$ the k-th coordinate of the pretrained model $\{Y_t\}_{0\le t\le T}$. For each $k = 1, \dots, d$, the covariation of

[5]It is well known that a uniformly integrable local martingale is a (true) martingale, see [44, Chapter II]. In our setting, a sufficient condition for $\{h(t, Y_t)\}_{0\le t\le T}$ to be a (true) martingale is that $\overline{g}(\cdot)$, $|\nabla h(\cdot,\cdot)|$ are bounded.

$h(t, Y_t)$ and Y_t^k is:

$$d[h, Y^k]_t = \overline{g}(t)^2 \partial_k h(t, Y_t) dt, \tag{3.8}$$

see [44, Chapter IV] [6]. Put it compactly,

$$d[h, Y]_t = \overline{g}(t)^2 \nabla h(t, Y_t) dt. \tag{3.9}$$

By substituting h on the right side of (3.9) with h_{ϕ_*}, and approximating $\nabla h(t, y)$ by a class of parametrized functions $\{q_\psi(t, y)\}_\psi$, we derive the *covariation loss* objective:

$$\min_\psi \mathbb{E}_{t \sim \text{Unif}\,[0,T]} \left\{ \mathbb{E}_{[0,T]} \left(\frac{1}{\overline{g}(t)^2} \frac{d[h_{\phi_*}, Y]_t}{dt} - q_\psi(t, Y_t) \right)^2 \right\}. \tag{3.10}$$

Denote by $q_{\psi_*}(t, y)$ the learned gradient of h by solving the stochastic optimization problem (3.10). The algorithm for conditional diffusion guidance via martingale-covariation loss (CDG-MCL) is summarized as follows.

Algorithm B Conditional diffusion guidance via martingale-covariation loss (CDG-MCL)

Input: pretrained model $\{Y_t\}_{0 \le t \le T}$ (2.3), guidance set S, parametrized families $\{h_\phi(t, y)\}_\phi$, $\{q_\psi(t, y)\}_\psi$

Step 1. Solve the stochastic optimization problem (3.7) that outputs ϕ_*.

Step 2. Solve the stochastic optimization problem (3.10) that outputs ψ_*.

Step 3. Sample

$$dY_t^S = \left(\overline{f}(t, Y_t^S) + \overline{g}(t)^2 \frac{q_{\psi_*}(t, Y_t^S)}{h_{\phi_*}(t, Y_t^S)} \right) dt + \overline{g}(t) dB_t, \quad Y_0^S \sim p_{\text{noise}}(\cdot).$$

Output: Y_T^S.

In Algorithm B, we write $\nabla \log h = \frac{\nabla h}{h}$, and estimate the numerator ∇h and the denominator h separately. We first learn the h function via the martingale property of $\{h(t, Y_t)\}_{0 \le t \le T}$, and then learn its gradient by using the (quadratic) covariation of $\{h(t, Y_t)\}_{0 \le t \le T}$ and $\{Y_t\}_{0 \le t \le T}$. Denote by $\widetilde{p}^S_{\text{pre}}(\cdot)$ the distribution of the output Y_T^S in Algorithm A or B.

3.3. **Practical considerations.** Now we provide a few practical variants for implementing the conditional diffusion guidance algorithms.

Labeling. Typically, the guidance set S is defined by a *label* or *reward function* $r : \mathbb{R}^d \to \mathbb{R}^m$. To be more precise,

$$Y_T \in S \Longleftrightarrow r(Y_T) \in S_r,$$

[6]The equation (3.8) can be understood as:

$$\frac{(h(t+\delta, Y_{t+\delta}) - h(t, Y_t))(Y_{t+\delta}^k - Y_t^k)}{\overline{g}(t)^2 \delta} \approx \partial_k h(t, Y_t), \quad \text{for sufficiently small } \delta > 0. \tag{$\star$}$$

There have been a body of works [4, 12, 17, 28, 29, 32] on the statistical estimation of quadratic variation. These papers consider how to estimate the quadratic variation of a process from a single trajectory in the context of (financial) time series. Here we have a different scenario, where the pretrained model $\{Y_t\}_{0 \le t \le T}$ under $P_{[0,T]}(\cdot)$ can be sampled repeatedly. The relation ($\star$) naturally provides an approximation to $\partial_k h$ by regressing the left side term over (t, Y_t).

where $S_r \subset \mathbb{R}^m$ is the guidance set in the label/reward space [7]. The corresponding martingale loss to learn $h(t,y) = \mathbb{P}_{[0,T]}(r(Y_T) \in S_r \,|\, Y_t = y)$ is specified by

$$\min_{\phi} \mathbb{E}_{t\sim \mathrm{Unif}\,[0,T]} \left\{ \mathbb{E}_{[0,T]}(h_\phi(t,Y_t) - 1(r(Y_T) \in S_r))^2 \right\}. \tag{3.7'}$$

ODE sampling. It is known in the diffusion folklore that the deterministic or ODE samplers are more efficient than the stochastic counterparts. The ODE sampler for the pretrained model (2.3) is:

$$\frac{dY_t}{dt} = -f(t,Y_t) + \frac{1}{2} g(T-t)^2 s_{\theta_*}(T-t,Y_t), \quad Y_0 \sim p_{\mathrm{noise}}(\cdot). \tag{3.11}$$

Denote by $P^o_{[0,T]}(\cdot)$ the distribution of the (true) time reversal (2.2) of $\{X_t\}_{0\le t\le T}$. Let $\mathring{h}(t,y) := \mathbb{P}^o_{[0,T]}(Y_T \in S \,|\, Y_t = y)$ be the associated h function. The following proposition is key to the ODE sampling of $p^S_{\mathrm{data}}(\cdot)$.

Proposition 3.3. *Let $\{X_t\}_{0\le t\le T}$ be defined by the SDE $dX_t = f(t,X_t)dt + g(t)dW_t$, $X_0 \sim p^S_{data}(\cdot)$. Let $\{X'_t\}_{0\le t\le T}$ be defined by the ODE:*

$$\frac{dX'_t}{dt} = f(t,X'_t) - \frac{1}{2}\left(\nabla \log p(t,X'_t) + \nabla \log \mathring{h}(t,X'_t)\right), \quad X_0 \sim p^S_{data}(\cdot). \tag{3.12}$$

Then X_t and X'_t have the same distribution for each t.

Proof. Let $p(t,x;S) := \frac{\mathbb{P}(X_t \in dx, X_0 \in S)/dx}{p_{\mathrm{data}}(S)}$, and note that $\nabla \log p(t,x;S) = \nabla \log p(t,x) + \nabla \log \mathring{h}(t,x)$. Thus, $\frac{dX'_t}{dt} = f(t,X'_t) - \frac{1}{2}\nabla \log p(t,X'_t;S)$. The conclusion follows from [52, Theorem 6.1]. □

It is expected that $\nabla \log \mathring{h}(t,y) \approx \nabla \log h(t,y)$. That is, the h functions are close under $P^o_{[0,T]}(\cdot)$ and $P_{[0,T]}(\cdot)$. Denote by $\mu_{\phi_*}(t,y)$ the function approximation for $\nabla \log h(t,y)$ in Algorithm A or B (i.e., $\nabla \log h_{\phi_*}(t,y)$ in Algorithm A and $\frac{q_{\psi_*}(t,y)}{h_{\phi_*}(t,y)}$ in Algorithm B). The ODE sampler of $p^S_{\mathrm{data}}(\cdot)$ is:

$$\frac{dY^S_t}{dt} = -f(t,Y_t) + \frac{1}{2} g(T-t)^2 \left(s_{\theta_*}(T-t,Y^S_t) + \mu_{\phi_*}(t,Y^S_t)\right), \quad Y_0 \sim p_{\mathrm{noise}}(\cdot). \tag{3.13}$$

Algorithm A and B can be easily adapted to labeling and ODE sampling, which are summarized as follows.

Note that Algorithm B' also requires the SDE sampler (2.3), which is used to estimate the covariation $\frac{d[h_{\phi_*},Y]_t}{dt}$ in Step 2. This is because the ODE and the SDE sampler only agree in distribution marginally, but not at the level of the process that is needed to approximate the quadratic variation.

---

[7] Here we consider the diffusion guidance by conditioning $(Y_T \,|\, r(Y_T) \in S_r)$. An alternative (popular) approach to fine-tuning the diffusion process is to solve:

$$\max \mathbb{E}[r_v(Y_T)] \quad \text{or} \quad \max \mathbb{E}_{p(\cdot)}[r_v(Y)] - \beta d_{KL}(p(\cdot), p_{\mathrm{pre}}(\cdot)),$$

where $r_v : \mathbb{R}^d \to \mathbb{R}$ ($m = 1$) is the real-valued reward, and $\beta > 0$ is the level of exploration (see [50, 55, 56]). In the latter case, the guided distribution is $\propto e^{r_v(y)/\beta} p_{\mathrm{pre}}(y)dy$, which is referred to as the *soft conditioning*. By taking $r_v(y) = \begin{cases} 1 & \text{if} \quad y \in S \\ -\infty & \text{if} \quad y \notin S \end{cases}$, we get the guidance by conditioning.

Algorithm A' Conditional diffusion guidance via martingale loss (CDG-ML)

Input: pretrained model $\{Y_t\}_{0\le t\le T}$ (3.11), label $r(y)$, guidance set S_r, parametrized family $\{h_\phi(t,y)\}_\phi$

Step 1. Solve the stochastic optimization problem (3.7') that outputs ϕ_*.

Step 2. Sample

$$\frac{dY_t^S}{dt} = -f(t,Y_t) + \frac{1}{2}\overline{g}(t)^2 s_{\theta_*}(T-t, Y_t^S) + \frac{1}{2}\overline{g}(t)^2 \nabla \log h_{\phi_*}(t, Y_t^S), \quad Y_0 \sim p_{\text{noise}}(\cdot).$$

Output: Y_T^S.

Algorithm B' Conditional diffusion guidance via martingale-covariation loss (CDG-MCL)

Input: pretrained model $\{Y_t\}_{0\le t\le T}$ (2.3) and (3.11), label $r(y)$, guidance set S_r, parametrized families $\{h_\phi(t,y)\}_\phi$, $\{q_\psi(t,y)\}_\psi$

Step 1. Solve the stochastic optimization problem (3.7') that outputs ϕ_*.

Step 2. Use the ODE (3.11) to sample Y_t, and then the SDE (2.3) to estimate $\frac{d[h_{\phi_*}, Y]_t}{dt}$. Solve the stochastic optimization problem (3.10) that outputs ψ_*.

Step 3. Sample

$$\frac{dY_t^S}{dt} = -f(t,Y_t) + \frac{1}{2}\overline{g}(t)^2 s_{\theta_*}(T-t, Y_t^S) + \frac{1}{2}\overline{g}(t)^2 \frac{q_{\psi_*}(t, Y_t^S)}{h_{\phi_*}(t, Y_t^S)}, \quad Y_0 \sim p_{\text{noise}}(\cdot).$$

Output: Y_T^S.

We also point out various numerical schemes to discretize the above continuous-time samplers, see [52, Section 5.3] and [62] for the references. Since we rely on the pretrained model for sampling, we will simply follow its built-in schemes (so we do not pursue this direction here).

4. Theoretical results

In this section, we provide theoretical results of the conditional diffusion guidance introduced in Section 2. The total variation and Wasserstein distance between the conditional target distribution $p^S_{\text{data}}(\cdot)$, and the diffusion guidance $\widetilde{p}^S_{\text{pre}}(\cdot)$ output by Algorithm A or B are studied in Section 4.1 and 4.2 respectively. In Section 4.3, we explore the convergence of the stochastic optimization algorithms to learn the h function.

4.1. **Total variation bounds.**

This part studies the total variation distance between $p^S_{\text{data}}(\cdot)$ and $\widetilde{p}^S_{\text{pre}}(\cdot)$.

Recall from Section 2 that $p(t,x)$ is the probability density of the forward process (2.1), and $s_{\theta_*}(t,x)$ is the score matching function of the pretrained model. Also recall from Section 3.1 that $P_t^S(\cdot)$ is the (marginal) distribution of Y_t^S defined by (3.4). Below we present a few assumptions.

Assumption 4.1.

(i) $d_{TV}(p(T,\cdot), p_{noise}(\cdot)) < \infty$.

(ii) The score matching satisfies: $\sup_{0\le t\le T} \mathbb{E}_{p(t,\cdot)} |s_{\theta_*}(t,X) - \nabla \log p(t,X)|^2 \le \varepsilon^2$.

(iii) *There is $\rho > 0$ such that $p_{data}(S) \ge \rho$.*

(iv) *There is $\eta > 0$ such that $|\nabla \log h - \nabla \log h_{\phi_*}|_\infty \le \eta$ for Algorithm A, or $\left|\nabla \log h - \frac{q_{\psi_*}}{h_{\phi_*}}\right|_\infty < \infty$ for Algorithm B.*

The assumptions (i)-(ii) ensure that $p_{\text{data}}(\cdot)$ and $p_{\text{pre}}(\cdot)$ are close. The assumption (iii) indicates that the guidance set S is non-negligible, so the conditional distributions on S are well-defined. The assumption (iv) provides blackbox errors for learning $\nabla \log h$ [8], which will be further developed in Section 4.3.

First, we bound the total variation distance between $p^S_{\text{data}}(\cdot)$ and $p^S_{\text{pre}}(\cdot)$.

Lemma 4.2. *Let Assumption 4.1 (iii) hold. We have:*

$$d_{TV}(p^S_{pre}(\cdot),\, p^S_{data}(\cdot)) \le \frac{3}{2\rho} d_{TV}(p_{pre}(\cdot),\, p_{data}(\cdot)). \tag{4.1}$$

Proof. Without loss of generality, assume that $p_{\text{pre}}(\cdot)$, $p_{\text{data}}(\cdot)$ have densities. We have:

$$\begin{aligned} d_{TV}(p^S_{\text{pre}}(\cdot),\, p^S_{\text{data}}(\cdot)) &= \frac{1}{2}\int |p^S_{\text{pre}}(x) - p^S_{\text{data}}(x)|dx \\ &= \frac{1}{2}\int_S \left|\frac{p_{\text{pre}}(x)}{p_{\text{pre}}(S)} - \frac{p_{\text{data}}(x)}{p_{\text{data}}(S)}\right| dx \\ &\le \frac{1}{2p_{\text{data}}(S)}\left(|p_{\text{pre}}(S) - p_{\text{data}}(S)| + \int_S |p_{\text{pre}}(x) - p_{\text{data}}(x)|dx\right) \\ &\le \frac{3}{2p_{\text{data}}(S)} d_{TV}(p_{\text{pre}}(\cdot),\, p_{\text{data}}(\cdot)), \end{aligned} \tag{4.2}$$

where the third inequality follows the triangle inequality $|p_{\text{pre}}(x)p_{\text{data}}(S) - p_{\text{data}}(x)p_{\text{pre}}(S)| \le p_{\text{pre}}(x)|p_{\text{pre}}(S) - p_{\text{data}}(S)| + p_{\text{pre}}(S)|p_{\text{pre}}(x) - p_{\text{data}}(x)|$. Combining (4.2) with the fact that $p_{\text{data}}(S) \ge \rho$ yields the bound (4.1). □

The next lemma bounds the total variation distance between the conditional pretrained distribution $p^S_{\text{pre}}(\cdot)$, and $\widetilde{p}^S_{\text{pre}}(\cdot)$ output by Algorithm A or B.

Lemma 4.3. *Let Assumption 4.1 (iv) hold. We have:*

$$d_{TV}(p^S_{pre}(\cdot),\, \widetilde{p}^S_{pre}(\cdot)) \le \eta\sqrt{\frac{T}{2}}. \tag{4.3}$$

---

[8] The assumption (iv) means that the function $\nabla \log h$ can be learned pointwise. Note that in Algorithm A and B, the h function is learned by solving stochastic optimization problems using the pretrained samples under $P_{[0,T]}(\cdot)$. So a more "reasonable" hypothesis is that $\nabla \log h$ can be learned under the pretrained distribution:

$$\sup_{0\le t\le T} \mathbb{E}_t|\nabla \log h(t,Y) - \nabla \log h_{\phi_*}(t,Y)| \le \eta^2 \quad \text{or} \quad \sup_{0\le t\le T} \mathbb{E}_t\left|\nabla \log h(t,Y) - \frac{q_{\psi_*}(t,Y)}{h_{\phi_*}(t,Y)}\right|^2 \le \eta^2, \tag{$\star\star$}$$

which will be studied in Section 4.3. As will be clear in the proof of Lemma 4.3, we need (technically):

$$\sup_{0\le t\le T} \mathbb{E}^S_t\,|\nabla \log h(t,Y) - \nabla \log h_{\phi_*}(t,Y)|^2 \le \eta^2 \ \text{or} \ \sup_{0\le t\le T} \mathbb{E}^S_t\left|\nabla \log h(t,Y) - \frac{q_{\psi_*}(t,Y)}{h_{\phi_*}(t,Y)}\right|^2 \le \eta^2.$$

to establish the total variation bound. That is, $\nabla \log h$ can be learned under the conditional guided distribution $P^S_{[0,T]}(\cdot)$. Our conjecture is that using sufficiently rich function approximations, the function $\nabla \log h$ can be learned pointwise, so it does not matter whether the evaluation is under $P_{[0,T]}(\cdot)$ or $P^S_{[0,T]}(\cdot)$.

Proof. Recall that $\mu_{\phi_*}(t, y)$ denotes the function approximation for $\nabla \log h(t, y)$ in Algorithm A or B (i.e., $\nabla \log h_{\phi_*}(t, y)$ in Algorithm A and $\frac{q_{\psi_*}(t,y)}{h_{\phi_*}(t,y)}$ in Algorithm B). Note that

$$\begin{aligned} d_{KL}(p^S_{\mathrm{pre}}(\cdot),\, \tilde{p}^S_{\mathrm{pre}}(\cdot)) &\le d_{KL}(Y^S_T,\, \widetilde{Y}^S_T) \\ &= \mathbb{E}\int_0^T \left|\nabla \log h(t, Y^S_t) - \mu_{\phi_*}(t, Y^S_t)\right|^2 dt \le \eta^2 T, \end{aligned} \tag{4.4}$$

where the first inequality follows the data processing inequality, and the second identity is a consequence of Girsanov's theorem. Further by Pinsker's inequality, we get the bound (4.3). ☐

Combining the above two lemmas yields the following result on the total variation distance between $\widetilde{p}^S_{\mathrm{data}}(\cdot)$ and $\widetilde{p}^S_{\mathrm{pre}}(\cdot)$.

Theorem 4.4. *Let Assumption 4.1 hold. We have:*

$$d_{TV}(p^S_{data}(\cdot),\, \widetilde{p}^S_{pre}(\cdot)) \le \frac{3}{2\rho} d_{TV}(p(T,\cdot),\, p_{noise}(\cdot)) + \left(\frac{3\varepsilon}{2\rho} + \eta\right)\sqrt{\frac{T}{2}}. \tag{4.5}$$

In particular, assuming that $\mathbb{E}_{p_{data}(\cdot)}|X|^2 < \infty$, there are $C_{VE}, C_{VP} > 0$ (independent of T) such that

$$d_{TV}(p^S_{data}(\cdot),\, \widetilde{p}^S_{pre}(\cdot)) \le \begin{cases} \frac{C_{VE}}{\rho} T^{-1}\sqrt{\mathbb{E}_{p_{data}(\cdot)}|X|^2} + \left(\frac{3\varepsilon_{VE}}{2\rho} + \eta\right)\sqrt{\frac{T}{2}} & \text{for VE,} \\ \frac{C_{VP}}{\rho} e^{-C_{VP} T}\sqrt{\mathbb{E}_{p_{data}(\cdot)}|X|^2} + \left(\frac{3\varepsilon_{VP}}{2\rho} + \eta\right)\sqrt{\frac{T}{2}} & \text{for VP.} \end{cases} \tag{4.6}$$

Proof. It follows from [52, Theorem 5.2] that $d_{TV}(p_{\mathrm{pre}}(\cdot),\, p_{\mathrm{data}}(\cdot)) \le d_{TV}(p(T,\cdot),\, p_{\mathrm{noise}}(\cdot)) + \varepsilon\sqrt{\frac{T}{2}}$. Combining this with Lemma 4.2 and 4.3 yields the bound (4.5). The rest of the theorem follows from (4.5) and Proposition 2.1. ☐

4.2. **Wasserstein bounds.** Here we consider the Wasserstein-2 distance between $p^S_{\mathrm{data}}(\cdot)$ and $\widetilde{p}^S_{\mathrm{pre}}(\cdot)$, which is more involved than the total variation bounds. Note that we cannot bound the Wasserstein distance between $p^S_{\mathrm{data}}(\cdot)$ and $p^S_{\mathrm{pre}}(\cdot)$ in terms of that between $p_{\mathrm{data}}(\cdot)$ and $p_{\mathrm{data}}(\cdot)$ as in Lemma 4.2. Our analysis relies on coupling and Malliavin calculus.

Recall from Section 3.3 that $P^o_{[0,T]}(\cdot)$ denotes the distribution of the (true) time reversal process of $\{X_t\}_{0 \le t \le T}$, whose drift is:

$$\overline{f}^o(t, y) = -f(T-t, y) + \overline{g}(t)^2 \nabla \log p(T-t, y).$$

Let $P^{o,S}_{[0,T]}(\cdot)$ be the conditional distribution of $P^o_{[0,T]}(\cdot)$ on $\{Y_T \in S\}$, and $P^{o,S}_t(\cdot)$ be its marginal distribution at Y_t. We need the following assumptions.

Assumption 4.5.

(i) $W_2(p(T,\cdot),\, p_{noise}(\cdot)) < \infty$.
(ii) There is $r > 0$ *such that* $(x-y)\cdot(f(t,x) - f(t,y)) \ge \alpha|x-y|^2$ *for all* t, x, y.
(iii) There is $\kappa_1 > 0$ *such that* $p(t,\cdot)$ *is* κ_1*-strongly log-concave for all* t.
(iv) There is $\varepsilon > 0$ *sufficiently small:* $|s_{\theta_*} - \nabla \log p|_\infty \le \varepsilon$.
(v) There is $\kappa_2 > 0$ *such that* $h(t,\cdot)$ *is* κ_2*-strongly log-concave for all* t.

(vi) *There is* $G:\mathbb{R}^d \to \mathbb{R}_+$ *such that* $|\nabla \log h(t,y)| \le \frac{G(y)}{T-t}$ *for all* t, y.

(vii) *There is* $\eta > 0$ *such that* $|\nabla \log h - \nabla \log h_{\phi_*}|_\infty \le \eta$ *for Algorithm A, or* $\left|\nabla \log h - \frac{q_{\psi_*}}{h_{\phi_*}}\right|_\infty < \infty$ *for Algorithm B.*

(viii) *There is* $F > 0$ *such that* $\mathbb{E}^o_{[0,T]}(\int_t^T |e^{\int_t^u \nabla \overline{f}(r,Y_r)dr}|_F^2 du \,|\, Y_t = y) \le F^2$ *for all* t, y.

(ix) *There is* $\gamma > 0$ *such that* $\mathbb{E}^o_{[0,T]}(\int_t^T |e^{\int_t^u \nabla \overline{f}(r,Y_r)dr} - e^{\int_t^u \nabla \overline{f}^o(r,Y_r)dr}|_F^2 du \,|\, Y_t = y) \le \gamma^2$ *for all* t, y.

(x) *There is* $K > 0$ *such that* $\mathbb{E}^{o,S}_t\left[\mathring{h}(t,Y)^{-\frac{3}{2}}\right]$, $\mathbb{E}^{o,S}_t\left[G(Y)^2\mathring{h}(t,Y)^{-\frac{3}{2}}\right] \le K$ *for all* t.

Before stating our result, we make several comments on Assumption 4.5. The conditions (i)–(iv) can be used to bound $W_2(p_{\text{data}}(\cdot), p_{\text{pre}}(\cdot))$, which together with (vii) yields an estimate of $W_2(p^S_{\text{data}}(\cdot), p^S_{\text{pre}}(\cdot))$ involving a perturbation bound on $\nabla \log h$. The conditions (v)–(x) are required for the perturbation analysis of $\nabla \log h$ via Malliavin calculus. Note that the condition (iii) holds for the VE and VP models, if $p_{\text{data}}(\cdot)$ is strongly log-concave. The condition (iv) is stronger than Assumption 4.1 (ii) for the same reason as explained in the footnote [8]. (In fact, it suffices to assume an L^2 bound under the guided distribution in Algorithm A or B.) Finally, the condition (vi) is reasonable, since it holds for heat(-like) kernels.

The following theorem provides a bound on $W_2(p^S_{\text{data}}(\cdot), \widetilde{p}^S_{\text{pre}}(\cdot))$.

Theorem 4.6. *Let Assumption 4.5 hold, and set* $\Lambda := \alpha + (\kappa_1 + \kappa_2)g^2_{\max}$. *We have:*

$$W_2(p^S_{data}(\cdot), \widetilde{p}^S_{pre}(\cdot)) \le e^{-\Lambda T} W_2(p(T,\cdot), p_{noise}(\cdot)) + C(\varepsilon + \eta + \gamma), \tag{4.7}$$

for some $C > 0$ *(independent of* T*). In particular, assuming that* $\mathbb{E}_{p_{data}(\cdot)}|X|^2 < \infty$, $p_{data}(\cdot)$ *is* κ*-strongly log-concave for* κ *sufficiently large and Assumption 4.5 (iv)–(x) holds, there are* $C_{VE}, C_{VP} > 0$ *(independent of* T*) such that*

$$W_2(p^S_{data}(\cdot), \widetilde{p}^S_{pre}(\cdot)) \le \begin{cases} C_{VE}\left(e^{-C_{VE}T}\sqrt{\mathbb{E}_{p_{data}(\cdot)}|X|^2} + \varepsilon + \eta + \gamma\right) & \text{for VE,} \\ C_{VP}\left(e^{-C_{VP}T}\sqrt{\mathbb{E}_{p_{data}(\cdot)}|X|^2} + \varepsilon + \eta + \gamma\right) & \text{for VP.} \end{cases} \tag{4.8}$$

Proof. The proof is split into three steps.

Step 1. We start by establishing a coupling bound on $W_2(p^S_{\text{data}}(\cdot), \widetilde{p}^S_{\text{pre}}(\cdot))$. Recall that $\mu_{\phi_*}(t,y)$ denotes the function approximation for $\nabla \log h(t,y)$ in Algorithm A or B (i.e., $\nabla \log h_{\phi_*}(t,y)$ in Algorithm A and $\frac{q_{\psi_*}(t,y)}{h_{\phi_*}(t,y)}$ in Algorithm B).

Consider the coupled equations:

$$\begin{cases} dU_t = \left(\overline{f}^o(t,U_t) + \overline{g}(t)^2 \nabla \log \mathring{h}(t,U_t)\right)dt + \overline{g}(t)dB_t, \\ dV_t = \left(\overline{f}(t,V_t) + \overline{g}(t)^2 \mu_{\phi^*}(t,V_t)\right)dt + \overline{g}(t)dB_t, \end{cases}$$

where (U_0, V_0) are coupled to achieve $W_2(p(T,\cdot), p_{\text{noise}}(\cdot))$. Note that $W_2^2(p^S_{\text{data}}(\cdot), \widetilde{p}^S_{\text{pre}}(\cdot)) \le \mathbb{E}|U_T - V_T|^2$, so our goal is to bound $\mathbb{E}|U_T - V_T|^2$. By Itô's formula, we get:

$$\begin{aligned} d|U_t - V_t|^2 = 2(U_t - V_t)\cdot(&-f(T-t,U_t) + \overline{g}(t)^2\nabla \log p(T-t,U_t) + \overline{g}(t)^2 \nabla \log \mathring{h}(t,U_t) \\ &+ f(T-t,V_t) - \overline{g}(t)^2 s_{\theta_*}(T-t,V_t) - \overline{g}(t)^2\mu_{\phi_*}(t,V_t))dt, \end{aligned}$$

which implies that

$$\begin{aligned}\frac{1}{2}\frac{d\mathbb{E}|U_t-V_t|^2}{dt} = &-\underbrace{\mathbb{E}[(U_t-V_t)\cdot(f(T-t,U_t)-f(T-t,V_t))]}_{(a)} \\ &+\overline{g}(t)^2\underbrace{\mathbb{E}[(U_t-V_t)\cdot(\nabla\log p(T-t,U_t)-s_{\theta_*}(T-t,V_t))]}_{(b)} \\ &+\overline{g}(t)^2\underbrace{\mathbb{E}[(U_t-V_t)\cdot(\nabla\log\mathring{h}(t,U_t)-\mu_{\phi_*}(t,V_t))]}_{(c)}.\end{aligned} \tag{4.9}$$

By Assumption 4.5 (ii), the term $(a)\geq\alpha\mathbb{E}|U_t-V_t|^2$. For the term (b), we get:

$$\begin{aligned}(b) &= \mathbb{E}[(U_t-V_t)\cdot(\nabla\log p(T-t,U_t)-\nabla\log p(T-t,V_t))] \\ &\quad+\mathbb{E}[(U_t-V_t)\cdot(\nabla\log p(T-t,V_t)-s_{\theta_*}(T-t,V_t))] \\ &\leq -\kappa_1\mathbb{E}|U_t-V_t|^2+\varepsilon\sqrt{\mathbb{E}|U_t-V_t|^2},\end{aligned} \tag{4.10}$$

which follows from Assumption 4.5 (iii) and (iv). Similarly, we have:

$$\begin{aligned}(c) &= \mathbb{E}[(U_t-V_t)\cdot(\nabla\log\mathring{h}(t,U_t)-\nabla\log h(t,U_t))] \\ &\quad+\mathbb{E}[(U_t-V_t)\cdot(\nabla\log h(t,U_t)-\nabla\log h(t,V_t))] \\ &\quad+\mathbb{E}[(U_t-V_t)\cdot(\nabla\log h(t,V_t)-\mu_{\phi_*}(t,V_t))] \\ &\leq -\kappa_2\mathbb{E}|U_t-V_t|^2+\left(\eta+\sqrt{\mathbb{E}|\nabla\log\mathring{h}(t,U_t)-\nabla\log h(t,U_t))|^2}\right)\sqrt{\mathbb{E}|U_t-V_t|^2}\end{aligned} \tag{4.11}$$

which follows from Assumption 4.5 (v) and (vii). Combining (4.9), (4.10) and (4.11) yields:

$$\begin{aligned}\frac{d\mathbb{E}|U_t-V_t|^2}{dt} &\leq -2\Lambda\,\mathbb{E}|U_t-V_t|^2 \\ &\quad+2g_{\max}^2\left(\sqrt{\mathbb{E}|\nabla\log\mathring{h}(t,U_t)-\nabla\log h(t,U_t))|^2}+\varepsilon+\eta\right)\sqrt{\mathbb{E}|U_t-V_t|^2}.\end{aligned} \tag{4.12}$$

Step 2. Now we apply Malliavin calculus to bound $|\nabla\log h(t,y)-\nabla\log\mathring{h}(t,y)|$. First we consider $|h(t,y)-\mathring{h}(t,y)|$. It follows from [21, Proposition 3.1] that for $\varepsilon>0$ sufficiently small,

$$\begin{aligned}&|h(t,y)-\mathring{h}(t,y)| \\ &\leq C\left|\mathbb{E}^o_{[0,T]}\left(1(Y_T\in S)\int_t^T\overline{g}(u)(\nabla\log p(T-u,Y_u)-s_\theta(T-u,Y_u))dB_u\,\middle|\,Y_t=y\right)\right| \\ &\leq Cg_{\max}\sqrt{\mathring{h}(t,y)}\sqrt{\mathbb{E}^o_{[0,T]}\left(\int_t^T|\nabla\log p(T-u,Y_u)-s_\theta(T-u,Y_u))|^2du\,\middle|\,Y_t=y\right)} \\ &\leq Cg_{\max}\varepsilon\sqrt{T-t}\sqrt{\mathring{h}(t,y)},\end{aligned} \tag{4.13}$$

where the second inequality is by the Cauchy–Schwarz inequality, and the last inequality is due to Assumption 4.5 (iv).

Next we bound $|\nabla h(t,y) - \nabla \mathring{h}(t,y)|$. Introduce the first variation process $\{Z_u\}_{t\le u\le T}$ that solves:

$$dZ_u = \nabla \overline{f}(u, Y_u) Z_u du, \quad Z_t = I.$$

So $Z_u = \exp\left(\int_t^u \nabla \overline{f}(r, Y_r)dr\right)$ (here $\nabla \overline{f}$ is a matrix.) By [21, Proposition 3.2],

$$\nabla h(t,y) = \mathbb{E}_{[0,T]}\left(\frac{1(Y_T\in S)}{T-t}\int_t^T \frac{Z_u}{\overline{g}(u)}dB_u\right) = \mathbb{E}_{[0,T]}\left(\frac{1(Y_T\in S)}{T-t}\int_t^T \frac{e^{\int_t^u \nabla \overline{f}(r,Y_r)dr}}{\overline{g}(u)}dB_u\right).$$

A similar argument as before shows that

$$\begin{aligned}
&|\nabla h(t,y) - \nabla \mathring{h}(t,y)| \le (d) + (e), \text{where}\\
&(d) = C\left|\mathbb{E}^o_{[0,T]}\left(\frac{1(Y_T\in S)}{T-t}\int_t^T \frac{e^{\int_t^u \nabla \overline{f}(r,Y_r)dr}}{\overline{g}(u)}dB_u\right.\right.\\
&\qquad\qquad \left.\left.\int_t^T \overline{g}(u)(\nabla \log p(T-u, Y_u) - s_\theta(T-u, Y_u))dB_u \,\middle|\, Y_t = y\right)\right| \qquad (4.14)\\
&(e) = \left|\mathbb{E}^o_{[0,T]}\left(\frac{1(Y_T\in S)}{T-t}\int_t^T \frac{e^{\int_t^u \nabla \overline{f}(r,Y_r)dr} - e^{\int_t^u \nabla \overline{f}^o(r,Y_r)dr}}{\overline{g}(u)}dB_u \,\middle|\, Y_t = y\right)\right|.
\end{aligned}$$

For the term (d), we have:

$$\begin{aligned}
(d) &\le \frac{C}{(T-t)g_{\min}}\mathring{h}(t,y)^{\frac14}\left\{\mathbb{E}^o_{[0,T]}\left(\int_t^T |e^{\int_t^u \nabla \overline{f}(r,Y_r)dr}|_F^2 du \,\middle|\, Y_t = y\right)\right\}^{\frac12}\\
&\qquad \left\{\mathbb{E}^o_{[0,T]}\left(\int_t^T \overline{g}(u)(\nabla \log p(T-u, Y_u) - s_\theta(T-u, Y_u))dB_u \,\middle|\, Y_t = y\right)^4\right\}^{\frac14}\\
&\le \frac{C}{(T-t)g_{\min}}\varepsilon \mathring{h}(t,y)^{\frac14}\left\{\mathbb{E}^o_{[0,T]}\left(\int_t^T |e^{\int_t^u \nabla \overline{f}(r,Y_r)dr}|_F^2 du \,\middle|\, Y_t = y\right)\right\}^{\frac12}\left(g_{\max}\varepsilon\sqrt{T-t}\right)\\
&\le \frac{C g_{\max} F}{g_{\min}}\varepsilon \mathring{h}(t,y)^{\frac14}, \qquad (4.15)
\end{aligned}$$

where the first inequality is due to Hölder's inequality, the second inequality follows from the moment inequality (see [44, Chapter IV, §4]) and Assumption 4.5 (iv), and the final inequality is by Assumption 4.5 $(viii)$. For the term (e), we have:

$$\begin{aligned}
(e) &\le \frac{1}{(T-t)g_{\min}}\sqrt{\mathring{h}(t,y)}\sqrt{\mathbb{E}^o_{[0,T]}\left(\int_t^T |e^{\int_t^u \nabla \overline{f}(r,Y_r)dr} - e^{\int_t^u \nabla \overline{f}^o(r,Y_r)dr}|_F^2 du \,\middle|\, Y_t = y\right)}\\
&\le \frac{1}{g_{\min}}\frac{\gamma}{\sqrt{T-t}}\sqrt{\mathring{h}(t,y)},
\end{aligned} \qquad (4.16)$$

where the first inequality is by the Cauchy–Schwarz inequality, and the second inequality follows from Assumption 4.5 (ix). Injecting (4.15) and (4.16) into (4.14) yields:

$$|\nabla h(t,y) - \nabla \mathring{h}(t,y)| \le C\left(F\varepsilon + \frac{\gamma}{\sqrt{T-t}}\right)\mathring{h}(t,y)^{\frac14}, \quad \text{for some } C > 0. \qquad (4.17)$$

Combining (4.13), (4.17) and Assumption 4.5 (vi) leads to:

$$\begin{aligned}|\nabla \log h(t,y) - \nabla \log \mathring{h}(t,y)| &\le \left|\frac{\mathring{h}(t,y)-h(t,y)}{\mathring{h}(t,y)}\nabla \log h(t,y)\right| + \left|\frac{\nabla h(t,y)-\nabla \mathring{h}(t,y)}{\mathring{h}(t,y)}\right| \\ &\le C\left\{\left(F+\frac{G(y)}{\sqrt{T-t}}\right)\varepsilon + \frac{\gamma}{\sqrt{T-t}}\right\}\mathring{h}(t,y)^{-\frac{3}{4}}.\end{aligned} \tag{4.18}$$

Step 3. Observe that $\{U_t\}_{0\le t\le T}$ is distributed by $P^{o,S}_{[0,T]}(\cdot)$. By (4.12), (4.18) and Assumption 4.5 (x), we have:

$$\frac{d\mathbb{E}|U_t-V_t|^2}{dt} \le -\Lambda\,\mathbb{E}|U_t-V_t|^2 + C\left(\varepsilon+\eta+\frac{\varepsilon+\gamma}{\sqrt{T-t}}\right)\sqrt{\mathbb{E}|U_t-V_t|^2}. \tag{4.19}$$

By Grönwall's inequality (see [15, Theorem 21]), we get:

$$\begin{aligned}\mathbb{E}|U_T-V_T|^2 &\le \left(e^{-\Lambda T}W_2(p(T,\cdot),\, p_{\text{noise}}(\cdot)) + C\int_0^T\left(\varepsilon+\eta+\frac{\varepsilon+\gamma}{\sqrt{T-t}}\right)e^{-\Lambda(T-t)}dt\right)^2 \\ &\le \left(e^{-\frac{1}{2}\Lambda T}W_2(p(T,\cdot),\, p_{\text{noise}}(\cdot)) + C(\varepsilon+\eta+\gamma)\right)^2,\end{aligned}$$

which leads to the bound (4.7).

The rest of the theorem follows from the fact that $W_2^2(p(T,\cdot),\, p_{\text{noise}}(\cdot)) \le \mathbb{E}_{p_{\text{data}}(\cdot)}|X|^2$ for VE, and $W_2(p(T,\cdot),\, p_{\text{noise}}(\cdot)) \le e^{-CT}\mathbb{E}_{p_{\text{data}}(\cdot)}|X|^2$ with $C>0$ for VP. □

4.3. **Learning the h function.** In this part, we study the convergence of the stochastic optimization problems outlined in Section 3.2 (Alogrithm A and B) to learn the function $\nabla \log h$. Our approach is generic, and does not require any explicit structure of function approximations.

4.3.1. *Learning h.* We first consider the convergence of the stochastic optimization problem (3.7). The stochastic approximation to the martingale loss is given by

$$\phi_{n+1} = \phi_n + \delta_n \mathcal{V}(\phi_n, \tau^{(n)}, Y^{(n)}), \tag{4.20}$$

where $\delta_n > 0$ is the step size, $\tau^{(n)} \sim \text{Unif}\,[0,T]$, $Y^{(n)} = \{Y^{(n)}_t\}_{0\le t\le T}$ is a copy of the pretrained model $P_{[0,T]}(\cdot)$, and

$$\mathcal{V}(\phi,\tau,Y) := -2\partial_\phi h_\phi(\tau,Y_\tau)(h_\phi(\tau,Y_\tau) - 1(Y_T\in S)). \tag{4.21}$$

Our goal is to provide a quantitative bound on $|h_{\phi_n}(t,y) - h(t,y)|$ (in some weak sense). Our idea follows from [53, Section 4], which relies on [5] for stochastic approximations. Set

$$V(\phi) := \mathbb{E}_{\tau\sim\text{Unif}\,[0,T]}\{\mathbb{E}_{[0,T]}[\mathcal{V}(\phi,\tau,Y)]\}. \tag{4.22}$$

We need the following assumptions.

Assumption 4.7.

(i) The ODE $\phi'(t) = V(\phi(t))$ has a unique stable equilibrium ϕ_* [9]*.*

(ii) There is $C>0$ such that $\mathbb{E}_{[0,T]}[\mathcal{V}(\phi_{n+1},Y)\,|\,\phi_n] \le C(1+\phi_n^2)$.

(iii) There is $\ell>0$ such that $(\phi-\phi_*)V(\phi) \le -\ell|\phi-\phi_*|^2$.

[9] ϕ_* is the unique stable equilibrium means that $V(\phi)=0$ has a unique root ϕ_*, and $V'(\phi_*)<0$.

(iv) There is a function $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\omega(r)/r^\nu$ is bounded for some $\nu \le 2$, and $|h_\phi - h_{\phi'}|_\infty \le \omega(|\phi - \phi'|)$ for all ϕ, ϕ'.

The assumptions (i)–(iii) guarantees that the stochastic approximation (4.20) converges, and the assumption (iv) quantifies the sensitivity of the function approximation $\{h_\phi(t,x)\}_\phi$ with respect to the parameter.

Theorem 4.8. *Let Assumption 4.7 hold, and $\delta_n = \frac{A}{n^\zeta + B}$ for some $\zeta \le 1$, $A > \frac{\zeta}{2\ell}$ and $B > 0$. We have:*

$$\mathbb{E}_{[0,T]}|h_{\phi_n} - h|_\infty \le |h - h_{\phi_*}|_\infty + Cn^{-\frac{\zeta\nu}{2}}. \tag{4.23}$$

Proof. It follows from [5, Theorem 22] that under Assumption 4.7 (i)–(iii) and $\delta_n = \frac{A}{n^\zeta + B}$,

$$\mathbb{E}_{[0,T]}|\phi_n - \phi_*|^2 \le Cn^{-\zeta}. \tag{4.24}$$

As a result,

$$\begin{aligned}\mathbb{E}_{[0,T]}|h_{\phi_n} - h|_\infty &\le |h - h_{\phi_*}|_\infty + \mathbb{E}_{[0,T]}|h_{\phi_n} - h_{\phi_*}|_\infty \\ &\le |h - h_{\phi_*}|_\infty + \mathbb{E}_{[0,T]}[\omega(|\phi_n - \phi_*|)] \\ &\le |h - h_{\phi_*}|_\infty + C(\mathbb{E}|\phi_n - \phi_*|^2)^{\frac{\nu}{2}},\end{aligned} \tag{4.25}$$

where the first inequality is from the triangle inequality, the second inequality is due to Assumption 4.7 (iv), and the last inequality is by Cauchy-Schwarz inequality. Combining (4.24) and (4.25) yields the bound (4.23). □

The first term $|h - h_{\phi_*}|_\infty$ on the right side of (4.23) quantifies how well the family $\{h_\phi(t,y)\}_\phi$ approximates the h function, and the second term $n^{-\frac{\zeta\nu}{2}}$ gives the convergence rate of the stochastic approximation (4.20). In particular, if the family $\{h_\phi(t,y)\}_\phi$ is rich enough to contain the h function (i.e., $|h - h_{\phi_*}|_\infty = 0$), and $\{h_\phi(t,y)\}_\phi$ is Lipschitz in ϕ (i.e, $\nu = 1$), then h_{ϕ_n} converges to h at a rate $n^{-\frac{1}{2}}$ by taking the step size $1/n$.

4.3.2. *Learning ∇h.* Now we establish similar results for the stochastic optimization problem (3.10). Fixing $n > 0$, we use h_{ϕ_n} to approximate the covariation, so the stochastic approximation to the covariation loss is:

$$\psi_{m+1} = \psi_m + \delta_m \mathcal{U}_n(\psi_m, \tau^{(m)}, Y^{(m)}), \tag{4.26}$$

where

$$\mathcal{U}_n(\psi, \tau, Y) := -2\partial_\psi q_\phi(\tau, Y_\tau)\left(q_\psi(\tau, Y_\tau) - \frac{1}{\overline{g}(\tau)^2}\frac{d[h_{\phi_n}, Y]_t}{dt}|_{t=\tau}\right). \tag{4.27}$$

Also set

$$U_n(\psi) := \mathbb{E}_{\tau\sim\mathrm{Unif}\,[0,T]}\{\mathbb{E}_{[0,T]}[\mathcal{U}_n(\psi, \tau, Y)]\}. \tag{4.28}$$

We need the following assumptions.

Assumption 4.9.

(i) The ODE $\psi'(t) = U_n(\psi(t))$ has a unique stable equilibrium ψ_{*n}.

(ii) There is $C > 0$ such that $\mathbb{E}_{[0,T]}[\mathcal{U}_n(\psi_{m+1}, Y) \,|\, \psi_m] \le C(1 + \psi_m^2)$.

(iii) There is $\ell > 0$ such that $(\psi - \psi_*)U_n(\psi) \le -\ell|\psi - \psi_*|^2$.

(iv) There is a function $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\omega(r)/r^{\nu'}$ is bounded for some $\nu' \le 2$, and $|q_\psi - q_{\psi'}|_\infty \le \omega(|\psi - \psi'|)$ for all ψ, ψ'.

Theorem 4.10. *Let Assumption 4.9 hold, and $\delta_n = \frac{A}{n^{\zeta'}+B}$ for some $\zeta' \le 1$, $A > \frac{\zeta'}{2\ell}$ and $B > 0$. We have:*

$$\mathbb{E}_{[0,T]}|q_{\psi_m} - \nabla h|_\infty \le \mathbb{E}_{[0,T]}\left|\frac{1}{\overline{g}(t)^2}\frac{d[h_{\phi_n}, Y]_t}{dt} - \nabla h(t, Y_t)\right| + \mathbb{E}_{[0,T]}|\nabla h_{\phi_n} - q_{\psi_{*n}}|_\infty + Cm^{-\frac{\zeta'\nu'}{2}}. \tag{4.29}$$

The proof of the theorem is in the same vein as that of Theorem 4.8. The first term $\mathbb{E}_{[0,T]}\left(\left|\frac{1}{\overline{g}(t)^2}\frac{d[h_{\phi_n},Y]_t}{dt} - \nabla h(t,Y_t)\right|\right)$ on the right side of (4.29) quantifies how close the covariation $\frac{d[h_{\phi_n},Y]_t}{dt}$ is to $\frac{d[h,Y]_t}{dt}$. However, it is generally hard to provide an explicit bound on this term [10], and we simply denote it by $\theta(n)$. Also note that the estimation of $\frac{d[h_{\phi_n},Y]_t}{dt}$ also incurs a sample error [11], which we do not pursue here. The second term $|\nabla h_{\phi_n} - q_{\psi_{*n}}|_\infty$ measures how well the family $\{q_\psi(t,y)\}_\psi$ approximates ∇h_{ϕ_n}, and the third term $m^{-\frac{\zeta\nu}{2}}$ is the convergence rate of the stochastic approximation (4.26).

Combining Theorem 4.8 and 4.10, we have (at least heuristically) that the learning error η of $\nabla \log h$ is of order:

$$\theta(n) + n^{-\frac{\zeta\nu}{2}} + m^{-\frac{\zeta'\nu'}{2}} + \text{discrepancy of approximations } \{h_\phi(t,y)\}_\phi, \{q_\psi(t,y)\}_\psi. \tag{4.30}$$

Again if the families $\{h_\phi(t,y)\}_\phi$, $\{q_\psi(t,y)\}_\psi$ are rich enough and Lipschitz in the parameter (i.e., $\nu = \nu' = 1$), then η is of order $\theta(n) + n^{-\frac{1}{2}} + m^{-\frac{1}{2}}$ by taking the step size $\delta_n = 1/n$.

5. Numerical experiments

5.1. Synthetic examples.

6. Conclusion

Acknowledgment: We thank David Yao for helpful discussions. Tang acknowledges financial support by NSF grant DMS-2206038, the Innovation Hub grant, and the Tang family Assistant Professorship.

References

[1] B. D. O. Anderson. Reverse-time diffusion equation models. *Stochastic Process. Appl.*, 12(3):313–326, 1982.

[2] S. r. Asmussen and P. W. Glynn. *Stochastic simulation: algorithms and analysis*, volume 57 of *Stochastic Modelling and Applied Probability*. Springer, New York, 2007.

[3] A. Bansal, H.-M. Chu, A. Schwarzschild, S. Sengupta, M. Goldblum, J. Geiping, and T. Goldstein. Universal guidance for diffusion models. In *CVPR*, pages 843–852, 2023.

[10] It follows from [33, Chapter VI, §6c] (and also [31]) that if a diffusion process $\{Z_t^n\}_{0\le t\le T}$ converges in distribution to $\{Z_t\}_{0\le t\le T}$, and under very technical conditions, the quadratic variation of Z^n converges in distribution to that of Z. So $\frac{d[h_{\phi_n},Y]_t}{dt}$ and $\frac{d[h,Y]_t}{dt}$ are expected to be close, because $h_{\phi_n} \approx h$ by Theorem 4.8. However, the proofs in [31, 33] rely on soft measure-theoretic arguments, and it seems to be a challenging task to provide an explicit convergence rate of quadratic variation.

[11] As mentioned in the footnote [6], the quadratic variation can be estimated by sampling the pretrained model repeatedly. So the central limit theorem implies that the sample error of estimating the covariation $\frac{d[h_{\phi_n},Y]_t}{dt}$ is of order $1/M$, with M the sample size.

[4] O. E. Barndorff-Nielsen and N. Shephard. Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica*, 72(3):885–925, 2004.
[5] A. Benveniste, M. Métivier, and P. Priouret. *Adaptive algorithms and stochastic approximations*, volume 22 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1990.
[6] J. Blanchet and H. Lam. Rare event simulation techniques. In *WSC*, pages 146–160, 2011.
[7] J. Blanchet and H. Lam. State-dependent importance sampling for rare-event simulation: An overview and recent advances. *Surveys in Operations Research and Management Science*, 17(1):38–59, 2012.
[8] J. A. Bucklew. *Introduction to rare event simulation*. Springer Series in Statistics. Springer-Verlag, New York, 2004.
[9] M. Chen, K. Huang, T. Zhao, and M. Wang. Score approximation, estimation and distribution recovery of diffusion models on low-dimensional data. In *ICML*, volume 40, pages 4672–4712, 2023.
[10] M. Chen, S. Mei, J. Fan, and M. Wang. An overview of diffusion models: Applications, guided generation, statistical rates and optimization. 2024. arXiv:2404.07771.
[11] S. Chen, S. Chewi, J. Li, Y. Li, A. Salim, and A. R. Zhang. Sampling is as easy as learning the score: theory for diffusion models with minimal data assumptions. In *ICLR*, 2023.
[12] K. Christensen, M. S. Nielsen, and M. Podolskij. High-dimensional estimation of quadratic variation based on penalized realized variance. *Stat. Inference Stoch. Process.*, 26(2):331–359, 2023.
[13] A. Denker, F. Vargas, S. Padhy, K. Didi, S. V. Mathis, R. Barbano, V. Dutordoir, E. Mathieu, U. J. Komorowska, and P. Lio. Deft: Efficient fine-tuning of diffusion models by learning the generalised h-transform. In *Neurips*, volume 37, 2024.
[14] P. Dhariwal and A. Nichol. Diffusion models beat GANs on image synthesis. In *Neurips*, volume 34, pages 8780–8794, 2021.
[15] S. S. Dragomir. *Some Gronwall type inequalities and applications*. Nova Science Publishers, Inc., Hauppauge, NY, 2003.
[16] Y. Du, M. Plainer, R. Brekelmans, C. Duan, F. Noé, C. P. Gomes, A. Aspuru-Guzik, and K. Neklyudov. Doob's lagrangian: A sample-efficient variational approach to transition path sampling. In *Neurips*, volume 37, 2024.
[17] J. Fan. A selective overview of nonparametric methods in financial econometrics. *Stat. Sci.*, pages 317–337, 2005.
[18] Y. Fan and K. Lee. Optimizing DDPM sampling with shortcut fine-tuning. 2023. arXiv:2301.13362.
[19] Y. Fan, O. Watkins, Y. Du, H. Liu, M. Ryu, C. Boutilier, P. Abbeel, M. Ghavamzadeh, K. Lee, and K. Lee. DPOK: Reinforcement learning for fine-tuning text-to-image diffusion models. In *Neurips*, volume 36, 2023.
[20] P. Fitzsimmons, J. Pitman, and M. Yor. Markovian bridges: construction, Palm interpretation, and splicing. In *Seminar on Stochastic Processes, 1992 (Seattle, WA, 1992)*, volume 33 of *Progr. Probab.*, pages 101–134. Birkhäuser Boston, Boston, MA, 1993.
[21] E. Fournié, J.-M. Lasry, J. Lebuchoux, P.-L. Lions, and N. Touzi. Applications of Malliavin calculus to Monte Carlo methods in finance. *Finance Stoch.*, 3(4):391–412, 1999.
[22] X. Gao, H. M. Nguyen, and L. Zhu. Wasserstein convergence guarantees for a general class of score-based generative models. 2023. arXiv:2311.11003.
[23] X. Gao, J. Zha, and X. Y. Zhou. Reward-directed score-based diffusion models via q-learning. 2024. arXiv:2409.04832.
[24] Y. Han, M. Razaviyayn, and R. Xu. Neural network-based score estimation in diffusion models: Optimization and generalization. In *ICLR*, 2024.
[25] U. G. Haussmann and E. Pardoux. Time reversal of diffusions. *Ann. Probab.*, 14(4):1188–1205, 1986.
[26] J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. In *Neurips*, volume 33, pages 6840–6851, 2020.
[27] J. Ho and T. Salimans. Classifier-free diffusion guidance. In *NeurIPS Workshop on Deep Generative Models and Downstream Applications*, 2021.
[28] M. Hoffmann. Adaptive estimation in diffusion processes. *Stochastic Process. Appl.*, 79(1):135–163, 1999.
[29] M. Hoffmann. L_p estimation of the diffusion coefficient. *Bernoulli*, 5(3):447–481, 1999.
[30] A. Hyvärinen. Estimation of non-normalized statistical models by score matching. *J. Mach. Learn. Res.*, 6:695–709, 2005.

[31] J. Jacod. Convergence en loi de semimartingales et variation quadratique. In *Seminar on Probability, XV (Univ. Strasbourg, Strasbourg, 1979/1980) (French)*, volume 850 of *Lecture Notes in Math.*, pages 547–560. Springer, Berlin, 1981.

[32] J. Jacod. Asymptotic properties of realized power variations and related functionals of semimartingales. *Stochastic Process. Appl.*, 118(4):517–559, 2008.

[33] J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 2003.

[34] T. Karras, M. Aittala, T. Aila, and S. Laine. Elucidating the design space of diffusion-based generative models. In *Neurips*, volume 35, pages 26565–26577, 2022.

[35] S. Khalafi, D. Ding, and A. Ribeiro. Constrained diffusion models via dual training. In *Neurips*, volume 37, 2024.

[36] F. Koehler, A. Heckett, and A. Risteski. Statistical efficiency of score matching: The view from isoperimetry. In *ICLR*, 2023.

[37] H. Lee, J. Lu, and Y. Tan. Convergence for score-based generative modeling with polynomial complexity. In *Neurips*, volume 35, pages 22870–22882, 2022.

[38] K. Lee, H. Liu, M. Ryu, O. Watkins, Y. Du, C. Boutilier, P. Abbeel, M. Ghavamzadeh, and S. S. Gu. Aligning text-to-image models using human feedback. 2023. arXiv:2302.12192.

[39] G. Li, Y. Wei, Y. Chen, and Y. Chi. Towards faster non-asymptotic convergence for diffusion-based generative models. In *ICLR*, 2024.

[40] Y. Liang, J. He, G. Li, P. Li, A. Klimovskiy, N. Carolan, J. Sun, J. Pont-Tuset, S. Young, and F. Yang. Rich human feedback for text-to-image generation. In *CVPR*, pages 19401–19411, 2024.

[41] H. Liu, T. Zhu, N. Jia, J. He, and Z. Zheng. Learning to simulate from heavy-tailed distribution via diffusion model. 2024. SSRN 4975931.

[42] Z. Ma, Y. Zhang, G. Jia, L. Zhao, Y. Ma, M. Ma, G. Liu, K. Zhang, J. Li, and B. Zhou. Efficient diffusion models: A comprehensive survey from principles to practices. 2024. arXiv:2410.11795.

[43] A. Q. Nichol, P. Dhariwal, A. Ramesh, P. Shyam, P. Mishkin, B. Mcgrew, I. Sutskever, and M. Chen. Glide: Towards photorealistic image generation and editing with text-guided diffusion models. In *ICML*, volume 39, pages 16784–16804, 2022.

[44] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.

[45] L. C. G. Rogers and D. Williams. *Diffusions, Markov processes, and martingales. Vol. 2*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000. Itô calculus, Reprint of the second (1994) edition.

[46] Y. Song and S. Ermon. Generative modeling by estimating gradients of the data distribution. In *Neurips*, volume 32, 2019.

[47] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole. Score-based generative modeling through stochastic differential equations. In *ICLR*, 2021.

[48] T. Sorensen, J. Moore, J. Fisher, M. L. Gordon, N. Mireshghallah, C. M. Rytting, A. Ye, L. Jiang, X. Lu, N. Dziri, T. Althoff, and Y. Choi. Position: A roadmap to pluralistic alignment. In *ICML*, volume 41, 2024.

[49] D. W. Stroock and S. R. S. Varadhan. *Multidimensional diffusion processes*, volume 233 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, 1979.

[50] W. Tang. Fine-tuning of diffusion models via stochastic control: entropy regularization and beyond. 2024. arXiv:2403.06279.

[51] W. Tang and H. Zhao. Contractive diffusion probabilistic models. 2024. arXiv:2401.13115.

[52] W. Tang and H. Zhao. Score-based diffusion models via stochastic differential equations–a technical tutorial. 2024. arXiv:2402.07487.

[53] W. Tang and X. Y. Zhou. Regret of exploratory policy improvement and q-learning. 2024. arXiv:2411.01302.

[54] M. Uehara, Y. Zhao, T. Biancalani, and S. Levine. Understanding reinforcement learning-based fine-tuning of diffusion models: A tutorial and review. 2024. arXiv:2407.13734.

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

[56] M. Uehara, Y. Zhao, K. Black, E. Hajiramezanali, G. Scalia, N. L. Diamant, A. M. Tseng, S. Levine, and T. Biancalani. Feedback efficient online fine-tuning of diffusion models. In *ICML*, volume 41, pages 48892–48918, 2024.

[57] P. Vincent. A connection between score matching and denoising autoencoders. *Neural Comput.*, 23(7):1661–1674, 2011.

[58] B. Wallace, M. Dang, R. Rafailov, L. Zhou, A. Lou, S. Purushwalkam, S. Ermon, C. Xiong, S. Joty, and N. Naik. Diffusion model alignment using direct preference optimization. In *CVPR*, pages 8228–8238, 2024.

[59] Y. Wang, Y. He, and M. Tao. Evaluating the design space of diffusion-based generative models. 2024. arXiv:2406.12839.

[60] G. I. Winata, H. Zhao, A. Das, W. Tang, D. D. Yao, S.-X. Zhang, and S. Sahu. Preference tuning with human feedback on language, speech, and vision tasks: A survey. 2024. arXiv:2409.11564.

[61] Y. Wu, M. Chen, Z. Li, M. Wang, and Y. Wei. Theoretical insights for diffusion guidance: A case study for Gaussian mixture models. In *ICML*, volume 41, 2024.

[62] Y. Wu, Y. Chen, and Y. Wei. Stochastic Runge-Kutta methods: Provable acceleration of diffusion models. 2024. arXiv:2410.04760.

[63] K. Yang, J. Tao, J. Lyu, C. Ge, J. Chen, W. Shen, X. Zhu, and X. Li. Using human feedback to fine-tune diffusion models without any reward model. In *CVPR*, pages 8941–8951, 2024.

[64] H. Yuan, K. Huang, C. Ni, M. Chen, and M. Wang. Reward-directed conditional diffusion: Provable distribution estimation and reward improvement. In *Neurips*, volume 36, 2023.

[65] H. Zhao, H. Chen, D. D. Yao, and W. Tang. Towards RLHF for deterministic samplers in diffusion models. 2024. Working paper.

[66] H. Zhao, H. Chen, J. Zhang, D. D. Yao, and W. Tang. Scores as Actions: a framework of fine-tuning diffusion models by continuous-time reinforcement learning. 2024. arXiv:2409.08400.

Department of Industrial Engineering and Operations Research, Columbia University.
Email address: `wt2319@columbia.edu`

Department of Finance and Risk Engineering, New York University.
Email address: `rx2364@nyu.edu`