Faster Diffusion Models via Higher-Order Approximation

Gen Li*† Yuchen Zhou*‡ Yuting Wei§ Yuxin Chen‡

June 30, 2025

Abstract

In this paper, we explore provable acceleration of diffusion models without any additional retraining. Focusing on the task of approximating a target data distribution in $\mathbb{R}^d$ to within ε total-variation distance, we propose a principled, training-free sampling algorithm that requires only the order of

$$d^{1+2/K}\varepsilon^{-1/K}$$

score function evaluations (up to log factor) in the presence of accurate scores, where K is an arbitrarily large fixed integer. This result applies to a broad class of target data distributions, without the need for assumptions such as smoothness or log-concavity. Our theory is robust vis-a-vis inexact score estimation, degrading gracefully as the score estimation error increases — without demanding higher-order smoothness on the score estimates as assumed in previous work. The proposed algorithm draws insight from high-order ODE solvers, leveraging high-order Lagrange interpolation and successive refinement to approximate the integral derived from the probability flow ODE.

Contents

1 Introduction **2**
1.1 Diffusion models and score function evaluations . . . 2
1.2 Training-free acceleration . . . 3
1.3 Main contributions . . . 3
1.4 Other related work . . . 4

2 Preliminaries **4**

3 Algorithm **5**

4 Main theory **7**
4.1 Assumptions . . . 7
4.2 Convergence guarantees . . . 8

5 Main analysis **8**
5.1 Additional notation . . . 8
5.2 Preliminary facts . . . 9
5.3 Main steps for proving Theorem 1 . . . 10

6 Discussion and future directions **14**

A Proof of Lemma 7 **18**

*The first two authors contributed equally.
†Department of Statistics, Chinese University of Hong Kong.
‡Department of Statistics, University of Illinois Urbana-Champaign.
§Department of Statistics and Data Science, the Wharton School, University of Pennsylvania.

B Proofs of technical lemmas **27**
B.1 Proof of Lemma 1 . . . 27
B.2 Proof of Lemma 3 . . . 28
B.3 Proof of Lemma 4 . . . 29
B.4 Proof of Lemma 6 . . . 30
B.5 Proof of Lemma 8 . . . 32
B.6 Proof of Lemma 9 . . . 39
B.7 Proof of Lemma 10 . . . 39
B.8 Proof of Lemma 11 . . . 40

1 Introduction

Diffusion models, which originally drew inspiration from nonequilibrium thermodynamics (Sohl-Dickstein et al., 2015), have emerged as a powerful driving force in the realm of generative modeling, reshaping the landscape of contemporary generative artificial intelligence (Croitoru et al., 2023; Yang et al., 2023). Yet, despite their incredible sample quality and enhanced stability, diffusion models face the challenge of slow data generation compared to alternatives like Generative Adversarial Networks (GANs) (Goodfellow et al., 2020) and Variational Autoencoders (VAEs) (Kingma and Welling, 2013). In comparison to GANs and VAEs that generate data in a single step (i.e., through a single forward pass of a neural network), diffusion models require a number of iterations. Each of these iterations, which oftentimes involves evaluating the pretrained score functions, necessitates at least one pass of a neural network or transformer, making the entire process more computationally intensive. In order to fully unleash the capability of diffusion models for real-time data generation, accelerating their sampling process without compromising sample quality is essential.

1.1 Diffusion models and score function evaluations

To set the stage, note that a diffusion model begins with a data-contaminating forward process:

$$\text{(forward process)} \qquad X_0 \sim p_{\mathsf{data}}, \quad X_t = \sqrt{1-\beta_t}\, X_{t-1} + \sqrt{\beta_t}\, W_t, \quad t = 1, \cdots, T, \tag{1a}$$

where p_{data} indicates the target data distribution in $\mathbb{R}^d$, T is the total number of steps, the W_t's are noise vectors independently drawn from $\mathcal{N}(0, I_d)$, and $\{\beta_t\}_{1\leq t\leq T} \subset (0,1)$ denotes some predetermined sequence that governs the variance of the additive noise at each step. In essence, this forward process first draws a sample X_0 from the distribution of interest, and progressively corrupts it by injecting independent Gaussian noise. When T is sufficiently large, it is often the case that the distribution of X_T becomes exceedingly close to $\mathcal{N}(0, I_d)$. The central task of diffusion generative modeling is to learn a time-reversal of the above forward process, enabling us to start with a pure noise distribution (e.g., $\mathcal{N}(0, I_d)$) and iteratively turn it back into the target distribution p_{data}. Mathematically, this can be described as the construction of a backward process

$$\text{(backward process)} \qquad Y_T \sim \mathcal{N}(0, I_d), \qquad Y_T \to Y_{T-1} \to \cdots \to Y_1 \to Y_0$$

that ensures the distributions of Y_t and X_t stay close for each step t, namely, $Y_t \overset{\mathrm{d}}{\approx} X_t$.

As already demonstrated by classical stochastic differential equation (SDE) literature (Anderson, 1982; Haussmann and Pardoux, 1986), time-reversal of the forward process (1) is made possible through the use of the (Stein) score functions $\{\nabla \log p_{X_t}(\cdot)\}$. Consequently, diffusion generative modeling involves an initial, extensive training phase to learn these score functions, which produces a collection of pretrained score estimates that can be readily invoked in data generation. For instance, two prominent score-based diffusion model paradigms — DDPM (Ho et al., 2020) and the probability flow ODE (or DDIM) (Song et al., 2021, 2020) — both rely upon these pretrained scores for iterative sampling. In practice, however, these mainstream approaches remain relatively slow due to their large iteration complexities; for example, generating high-quality samples with DDPM can take hundreds or even thousands of iterations, with each iteration requiring the evaluation of a neural network or a transformer. All this underscores the need to speed up the iterative sampling process while maintaining the outstanding sample quality diffusion models are known for.

1.2 Training-free acceleration

In this work, we explore principled, training-free acceleration of diffusion models (Lu et al., 2022; Zheng et al., 2023; Zhao et al., 2024; Zhang and Chen, 2023; Jolicoeur-Martineau et al., 2021; Xue et al., 2024). In a nutshell, the training-free approach wraps around and makes more efficient use of the pretrained score functions to expedite data generation, which stands in sharp contrast to the training-based approach — such as distillation or consistency models — that performs another round of resource-intensive training (Luhman and Luhman, 2021; Salimans and Ho, 2022; Meng et al., 2023; Song et al., 2023; Li et al., 2024b; Nichol and Dhariwal, 2021). Remarkably, training-free acceleration has already seen widespread adoption in practice, with a prominent example being the DPM-solver (Luhman and Luhman, 2021) that has achieved dramatic empirical speedup compared to the original probability flow ODE algorithm.

Nevertheless, the theoretical underpinnings for accelerated diffusion models remain largely elusive. It was not until recently that researchers began developing non-asymptotic analysis frameworks for the training-free approach, although significant technical hurdles remain that limit the theoretical performance. More concretely, consider training-free acceleration methods that come with provable guarantees, and suppose that we would like to generate a sample whose distribution is ε-close in total-variation (TV) distance to the target distribution of interest. Li et al. (2024a) proposed a deterministic accelerated approach, akin to the second-order ODE solver, achieving an iteration complexity of $\widetilde{O}(d^3/\sqrt{\varepsilon})$.[1] Huang et al. (2024a) analyzed the p-th order Runge-Kutta method for solving the probability flow ODE, and established an iteration complexity of $\widetilde{O}\big((LDd)^{1+1/p}/\varepsilon^{1/p}\big)$ for any constant order p, where D represents the radius of the data support and L bounds certain high-order derivatives of the score estimates. This represents a clear improvement over Li et al. (2024a) in terms of the ε dependency, although the stringent requirements on L and D limit the range of data distributions it can accommodate without compromising the acceleration effect. Meanwhile, stochastic training-free methods have been proposed and analyzed in recent literature as well (Li et al., 2024a; Wu et al., 2024; Li and Cai, 2024), with the state-of-the-art results achieving an iteration complexity of $\widetilde{O}(d^{5/4}/\sqrt{\varepsilon})$ (Li and Cai, 2024). While such stochastic acceleration methods often come with theoretical guarantees under minimal assumptions, their ε dependency remains considerably worse than that of the deterministic counterpart.

In summary, while the strand of work outlined above has made significant progress towards enriching the theoretical foundation for training-free acceleration, it still leaves room for improvement, given that all of these prior results fell short of optimality.

1.3 Main contributions

Motivated by the limitations of prior work as outlined above, the present paper seeks to strengthen the theoretical and algorithmic development for training-free acceleration of diffusion models. More concretely, we propose a novel training-free acceleration algorithm aimed at solving the probability flow ODE. Falling under the category of deterministic samplers (except for its initialization), the proposed algorithm exploits high-order Lagrange polynomial interpolation, in conjunction with suitable successive refinement, to approximately calculate the integral derived from the probability flow ODE. We prove that, to achieve ε-precision in total variation, it takes our algorithm $\widetilde{O}(\frac{d^{1+2/K}}{\varepsilon^{1/K}})$ iterations, or equivalently,

$$\widetilde{O}\bigg(\frac{d^{1+2/K}}{\varepsilon^{1/K}}\bigg) \text{ score function evaluations}$$

in the presence of accurate scores. Our theory does not require any smoothness or log-concavity assumptions on the target distribution, making it broadly applicable. When only inexact score estimates are available, we demonstrate that the sampling quality degrades gracefully as the score estimation error and the associated Jacobian errors increase. Our results outperform Li et al. (2024a); Wu et al. (2024); Li and Cai (2024) by offering better scaling in both ε and d, and improve upon Huang et al. (2024a) by accommodating a broader set of nonsmooth distributions without assuming higher-order smoothness.

[1] Throughout this paper, the standard notation $f(d,\varepsilon^{-1}) = O(g(d,\varepsilon^{-1}))$ or $f(d,\varepsilon^{-1}) \lesssim g(d,\varepsilon^{-1})$ means that there exists a numerical constant $c_1 > 0$ such that $|f(d,\varepsilon^{-1})| \le c_1 |g(d,\varepsilon^{-1})|$; $f(d,\varepsilon^{-1}) \gtrsim g(d,\varepsilon^{-1})$ means that $g(d,\varepsilon^{-1}) \lesssim f(d,\varepsilon^{-1})$; and $f(d,\varepsilon^{-1}) \asymp g(d,\varepsilon^{-1})$ means that both $f(d,\varepsilon^{-1}) \lesssim g(d,\varepsilon^{-1})$ and $f(d,\varepsilon^{-1}) \gtrsim g(d,\varepsilon^{-1})$ hold true. The notation $\widetilde{O}(\cdot)$ is defined analogously to $O(\cdot)$ except that the log dependency is hidden.

1.4 Other related work

The development of convergence theory for diffusion models — particularly DDPM and DDIM — has received much attention during the past few years. Partial examples include Chen et al. (2022); Liu et al. (2022); Lee et al. (2023); Chen et al. (2023a,c); Li et al. (2023); Cheng et al. (2023); Chen et al. (2023b); Benton et al. (2024); Tang and Zhao (2024); Liang et al. (2024); Li et al. (2025); Jiao and Li (2024); Cai and Li (2025); Li et al. (2024d); Gao and Zhu (2024); Li and Jiao (2024); Li and Yan (2024b); see also the references therein. The convergence analysis for the DDPM was carried out in Chen et al. (2022) by means of Girsanov's theorem, and this analysis framework was subsequently improved to accommodate nonsmooth data distributions Lee et al. (2023); Chen et al. (2023a), unveiling an iteration complexity with nearly linear d dependency Benton et al. (2024); Li and Yan (2024b). When it comes to the DDIM sampler or the probability flow ODE, the convergence analysis was originally provided in Chen et al. (2023c), with a set of subsequent work devoted to sharpening the iteration complexity (Li et al., 2024c; Huang et al., 2024a; Li et al., 2024d; Liang et al., 2025). In addition, a recent line of work Li and Yan (2024a); Huang et al. (2024b); Azangulov et al. (2024); Potaptchik et al. (2024); Liang et al. (2025); Tang and Yan (2025) explored how DDPM and DDIM adapt to unknown low-dimensional structures underlying the target data distribution, leading to substantially improved theoretical guarantees.

2 Preliminaries

Before proceeding to the description of our algorithm, let us start by gathering several preliminary facts.

The forward process. Recall the forward process (1). By taking

$$\alpha_t := 1 - \beta_t \qquad \text{and} \qquad \overline{\alpha}_t := \prod_{i=1}^{t} \alpha_i, \tag{2}$$

we see that (1) admits the following distributional characterization:

$$X_t = \sqrt{\overline{\alpha}_t} X_0 + \sqrt{1-\overline{\alpha}_t}\, \overline{W}_t \qquad \text{with } \overline{W}_t \sim \mathcal{N}(0, I_d). \tag{3}$$

Moreover, we find it useful to introduce a continuous-time generalization $(\overline{X}_\tau)$ of the above process. Specifically, for every $\tau \in [0,1]$, construct $\overline{X}_\tau$ such that

$$\overline{X}_\tau = \sqrt{1-\tau} X_0 + \sqrt{\tau} Z \qquad \text{with } \; X_0 \sim p_{\mathsf{data}} \; \text{ and } \; Z \sim \mathcal{N}(0, I_d), \tag{4}$$

where X_0 and Z are independently generated. Clearly, with this construction we have

$$\overline{X}_{1-\overline{\alpha}_t} \overset{\mathrm{d}}{=} X_t. \tag{5}$$

In the sequel, we will often refer to τ as the time variable too as long as it is clear from the context, and will develop and describe our algorithm mainly based on this continuous-time forward process (4).

The score function and Tweedie's formula. As mentioned previously, a key object that plays a pivotal role in the sampling procedure is the (Stein) score function, defined as the gradient of the log marginal density of the forward process. To be precise, the score function associated with $(\overline{X}_\tau)$ is defined and denoted by

$$s_\tau^\star(X) := \nabla \log p_{\overline{X}_\tau}(X) \tag{6}$$

for each $\tau \in [0,1]$. The celebrated Tweedie formula (Efron, 2011) tells us that

$$s_\tau^\star(x) = -\frac{1}{\tau}\Big\{ x - \sqrt{1-\tau}\, \mathbb{E}\big[X_0 \mid \overline{X}_\tau = x\big] \Big\} \tag{7}$$

for any $x \in \mathbb{R}^d$, which implies that

$$s_\tau^\star(\sqrt{1-\tau}x) = -\frac{\sqrt{1-\tau}}{\tau} \int_{x_0} p_{X_0 \mid \overline{X}_\tau}(x_0 \mid \sqrt{1-\tau}x)(x - x_0) \mathrm{d}x_0. \tag{8}$$

Algorithm 1: HEROISM (**H**igh-ord**ER** **O**de-based d**I**ffusion **S**a**M**pler)

input: $\tau_{t,0} = \tau_{t+1,K-1} = 1 - \overline{\alpha}_t$ for all $1 \le t \le T-1$, and $\tau_{T,0} = 1 - \overline{\alpha}_T$. Choose $\tau_{t,i}$ as in (11).
initialization: $t = T, \ldots, 2$, $x_{\tau_{T,0}} \leftarrow Y_T \sim \mathcal{N}(0, I_d)$.
for $t > 1$ **do**
  $x_{\tau_{t,0}} \leftarrow Y_t$ and $x_{\tau_{t,i}}^{(0)} \leftarrow Y_t$ for all $i = 0, \ldots, K-1$
  for $n = 0, \ldots, N-1$ **do**
    compute $x_{\tau_{t,i}}^{(n+1)}$ via Eqn. (14) for all $i = 0, \ldots, K-1$
  $Y_{t-1} \leftarrow x_{\tau_{t,K-1}}^{(N)}$
output: Y_1

The probability flow ODE. As has been shown by Song et al. (2021), there exists an ordinary differential equation (ODE), called the probability flow ODE, that is able to reverse the forward process (4) (in the sense of yielding matching marginal distributions). Note that the precise form of the probability flow ODE depends on the parameterization of the forward process (e.g., (X_t) and $(\overline{X}_\tau)$ are associated with different, although related, ODEs). In particular, the probability flow ODE associated with process (4) takes the following form:

$$\mathrm{d}\frac{Y_\tau^{\mathsf{ode}}}{\sqrt{1-\tau}} = -\frac{1}{2(1-\tau)^{3/2}} s_\tau^\star(Y_\tau^{\mathsf{ode}})\mathrm{d}\tau; \tag{9}$$

see, e.g., Li et al. (2024a, Eqn. (20)). When setting $Y_{\tau_2}^{\mathsf{ode}} = \overline{X}_{\tau_2}$ and running (9) backward from τ_2 to τ_1, one has $Y_{\tau_1}^{\mathsf{ode}} \overset{\mathrm{d}}{=} \overline{X}_{\tau_1}$, allowing one to transport the distribution of $\overline{X}_{\tau_2}$ to that of $\overline{X}_{\tau_1}$.

Notation. For two probability measures P and Q, the total-variation distance between them is defined as $\mathsf{TV}(P, Q) \coloneqq \frac{1}{2}\int |\mathrm{d}P - \mathrm{d}Q|$. For any random object X, we let p_X denote its probability density function. For any matrix A, we denote by $\|A\|$ its spectral norm.

3 Algorithm

In this section, we present the main ideas and detailed procedure of the proposed algorithm. At a high level, our algorithm is an iterative procedure that attempts to approximately solve ODE (9) by approximating the true scores in the integrand using high-order polynomials.

Approximating an integral. Recall that $X_t = \overline{X}_{1-\overline{\alpha}_t}$ and $X_{t-1} = \overline{X}_{1-\overline{\alpha}_{t-1}}$, which makes apparent the importance of time points $1 - \overline{\alpha}_t$ and $1 - \overline{\alpha}_{t-1}$ when generating Y_{t-1}. Armed with the probability flow ODE (9), an ideal strategy for step $t-1$ would be to evaluate the following integral:

$$\frac{Y_{1-\overline{\alpha}_{t-1}}^{\mathsf{ode}}}{\sqrt{\overline{\alpha}_{t-1}}} = \frac{Y_{1-\overline{\alpha}_t}^{\mathsf{ode}}}{\sqrt{\overline{\alpha}_t}} - \int_{1-\overline{\alpha}_t}^{1-\overline{\alpha}_{t-1}} \frac{1}{2(1-\tau)^{3/2}} s_\tau^\star(Y_\tau^{\mathsf{ode}})\mathrm{d}\tau. \tag{10}$$

However, computing (10) involves evaluation an infinite number of score functions (i.e., those in the integral), making it computationally intensive. As a result, everything comes down to how to approximate the integral using a small number of score function evaluations.

Higher-order polynomials. In comparison to first-order solvers like DDIM that construct Y_{t-1} solely based on the score $s_{1-\overline{\alpha}_t}(\cdot)$, we propose to take advantage of the score functions at K time points — denoted by $\tau_{t,0}, \ldots, \tau_{t,K-1}$ — within the interval $[1 - \overline{\alpha}_{t-1}, 1 - \overline{\alpha}_t]$. To be precise, set

$$\tau_{t,0} = 1 - \overline{\alpha}_t \qquad \text{and} \qquad \tau_{t,K-1} = 1 - \overline{\alpha}_{t-1} \tag{11a}$$

to be the two endpoints of $[1-\overline{\alpha}_{t-1}, 1-\overline{\alpha}_t]$, and create equi-spaced points within this interval as

$$\tau_{t,i} = \tau_{t,0} - \frac{i}{K-1}(\tau_{t,0} - \tau_{t,K-1}), \qquad 0 < i < K, \tag{11b}$$

which clearly satisfy $\tau_{t,K-1} < \tau_{t,K-2} < \cdots < \tau_{t,0}$. It then boils down to how to exploit the score functions at these K points to approximately compute (10).

Our strategy is to attempt approximation by means of the Lagrange interpolation. To be precise, consider the renowned Lagrange basis polynomials as follows

$$\psi_i(\tau) := \frac{\prod_{i':i'\neq i}(\tau - \tau_{t,i'})}{\prod_{i':i'\neq i}(\tau_{t,i} - \tau_{t,i'})}, \qquad 0 \le i \le K. \tag{12}$$

Given a set of points $(\tau_{t,i}, (1-\tau_{t,i})^{-3/2} s_{\tau_{t,i}}(Y^{\mathsf{ode}}_{\tau_{t,i}}))$ for $0 \le i \le K-1$, we would like to approximate $\frac{1}{(1-\tau)^{3/2}} s^\star_\tau(x_\tau)$ via the Lagrange interpolating polynomial passing through the above K points, i.e.,

$$\frac{1}{(1-\tau)^{3/2}} s^\star_\tau(Y^{\mathsf{ode}}_\tau) \approx \sum\nolimits_{0\le i<K} \psi_i(\tau) \frac{s_{\tau_{t,i}}(Y^{\mathsf{ode}}_{\tau_{t,i}})}{(1-\tau_{t,i})^{3/2}}, \quad \forall \tau \in [\tau_{t,K-1}, \tau_{t,0}]. \tag{13}$$

Here, we have also replaced the true score $s^\star_t$ with the score estimate s_t on the right-hand side of (13). Note that the Lagrange interpolating polynomial on right-hand side of (13) forms a unique degree-$(K-1)$ polynomial passing through these K points.

Successive refinement. Noteworthily, the polynomial appproximation (13) still cannot be readily used, given that the points $\{Y^{\mathsf{ode}}_{\tau_{t,i}}\}$ are inaccessible in general. To remedy this issue, we propose a successive refinement scheme that alternates between estimating $\{Y^{\mathsf{ode}}_{\tau_{t,i}}\}$ and computing the the associated scores. More precisely, let us perform an iterative estimation procedure containing N rounds, where in each round n, we produce a sequence $\{x^{(n)}_{\tau_{t,i}}\}_{0\le i\le K-1}$ as an estimate of $\{Y^{\mathsf{ode}}_{\tau_{t,i}}\}$. When $\{x^{(n)}_{\tau_{t,i}}\}_{0\le i\le K-1}$ is available, we compute, for each $\tau' \in \{\tau_{t,0}, \ldots, \tau_{t,K-1}\}$,

$$\frac{x^{(n+1)}_{\tau'}}{\sqrt{1-\tau'}} = \frac{x_{\tau_{t,0}}}{\sqrt{1-\tau_{t,0}}} + \sum_{0\le j\le K-1} \gamma_{t,j}(\tau') \frac{1}{(1-\tau_{t,j})^{3/2}} s_{\tau_{t,j}}(x^{(n)}_{\tau_{t,j}}). \tag{14}$$

$$\text{where} \quad \gamma_{t,i}(\tau') := \int_{\tau'}^{\tau_{t,0}} \psi_i(\tau)\mathrm{d}\tau \quad \text{and} \quad x_{\tau_{t,0}} = Y_t. \tag{15}$$

In a nutshell, we employ $\{x^{(n)}_{\tau_{t,i}}\}_{0\le i\le K-1}$ to update the scores used in the polynomial approximation (13), which in turn helps us generate $\{x^{(n+1)}_{\tau_{t,i}}\}_{0\le i\le K-1}$ as improved estimates of $\{Y^{\mathsf{ode}}_{\tau_{t,i}}\}$.

Algorithm description. We are now ready to present our algorithm.

- **Initialization.** Let us initialize our algorithm with $x_{\tau_{T,0}} = Y_T \sim \mathcal{N}(0, I_d)$ and $x^{(0)}_{\tau_{T,i}} = x_{\tau_{T,0}}$ for each $0 < i < K$. To ease presentation, we also take $x^{(N)}_{\tau_{T+1,K-1}} = x_{\tau_{T,0}}$.
- **Iterative update rule.** Working backward from $t = T$ to $t = 1$, we generate the point Y_{t-1} as follows. Let us look at the time points $\{\tau_{t,i}\}_{0\le i<K}$, and initialize both the sequence $\{x^{(0)}_{\tau_{t,i}}\}_{0\le i\le K-1}$ and $x_{\tau_{t,0}}$ using the value $x^{(N)}_{\tau_{t+1,K-1}}$ obtained in the previous iteration. For each round $n = 0, \ldots, N-1$, we generate a sequence $\{x^{(n+1)}_{\tau_{t,i}}\}_{0\le i\le K-1}$ according to Eq. (14). After N rounds of successive refinement, we output $Y_{t-1} = x^{(N)}_{\tau_{t,K-1}}$.

The whole procedure of the proposed algorithm is summarized in Algorithm 1.

Several remarks are in order. To begin with, once the value of Y_T is given, the sequence $\{x^{(n)}_{\tau_{t,i}}\}$ is purely deterministic. Thus, when accurate score functions are available at the requested time points, the discrepancy between the distributions of Y_t and X_t mainly stems from the approximation error when interpolating a complicated function using polynomials of degree $K-1$. In addition, our algorithm requires computing the score functions at the sequence $\{x^{(n)}_{\tau_{t,i}}\}_{0\le i\le K-1, 0\le n\le N, 1\le t\le T}$, and hence the total number of score function evaluations is TKN. As we shall see momentarily in our theoretical development, K is taken to be a constant, whereas N is chosen to be logarithmically large. Consequently, the total number of score function evaluations is on the order of $\widetilde{O}(T)$.

4 Main theory

Encouragingly, the proposed algorithm comes with intriguing convergence guarantees, as we present in this section. Throughout this paper, we let q_t (resp. p_t) denote the distribution of X_t (resp. Y_t).

4.1 Assumptions

Let us gather the key assumptions needed to state our main theorem. First of all, the sequence $\{\beta_t\}$ that determines the variance of the injective noise (cf. (1)) is taken to be

$$\beta_1 = 1-\alpha_1 = T^{-c_0} \tag{16a}$$

$$\beta_t = 1-\alpha_t = \frac{c_1\log T}{T}\min\left\{\beta_1\left(1+\frac{c_1\log T}{T}\right)^t, 1\right\}, \qquad t\ge 2 \tag{16b}$$

for some large enough numerical constants $c_0, c_1 > 0$. In a nutshell, β_t undergoes two phases, growing exponentially fast at the beginning and then staying flat in the remaining steps. Such a two-phase schedule is consistent with what is used in the state-of-the-art diffusion theory work (e.g., Benton et al. (2024); Li et al. (2024c); Li and Yan (2024b); Huang et al. (2024b)).

Next, we turn to assumptions for the target data distribution. Our goal is to accommodate a very broad family of data distributions without imposing restrictions like smoothness or log-concavity. In fact, the only assumption we make about p_{data} is the following boundedness condition, where the radius is allowed to be polynomially large, with an arbitrarily large constant degree. A polynomially large support size can readily accommodate a very broad range of applications.

Assumption 1 (Target data distribution)**.** *The target distribution p_{data} satisfies*

$$\mathbb{P}\left(\|X_0\|_2 \le T^{c_R} \mid X_0 \sim p_{p_{\mathsf{data}}}\right) = 1$$

for some arbitrarily large constant $c_R > 0$.

In the presence of imperfect score estimation, it is natural to anticipate that the sampling quality is largely affected by the goodness of the score function estimates. In light of this, we introduce another set of assumptions to capture the quality of score function estimates. The first one below is concerned with the mean squared score estimation error, averaged over all time steps and iterations.

Assumption 2. *Suppose that the score estimates at hand satisfy*

$$\frac{1}{TNK}\sum_{t=1}^{T}\mathbb{E}_{Y_t\sim q_t}\left[\varepsilon^2_{\mathsf{score},t}(Y_t)\right] \le \varepsilon^2_{\mathsf{score}},$$

where $\varepsilon^2_{\mathsf{score},t}(x_{\tau_t,0}) = \sum_{i=0}^{K-1}\sum_{n=0}^{N-1}\left\|s_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}}) - s^\star_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}})\right\|_2^2$.

In stark contrast to the SDE-based samplers (e.g., DDPM) for which the ℓ_2 estimation error alone suffices for the stability analysis (e.g., Chen et al. (2022); Benton et al. (2024)), the ODE-based samplers require additional assumptions in order to preclude some unfavorable problem instances. See, e.g., Li et al. (2024c) for a lower bound that justifies why ℓ_2 score accuracy assumption alone is insufficient. In this work, we impose the following assumption concerning the goodness of the associated Jacobian errors.

Assumption 3. *Suppose that $s_t(\cdot)$ is continuously differentiable for each $1\le t\le T$, and satisfies*

$$\frac{1}{TNK}\sum_{t=1}^{T}\mathbb{E}_{Y_t\sim q_t}\left[\varepsilon^2_{\mathsf{Jacobi},t}(Y_t)\right] \le \varepsilon^2_{\mathsf{Jacobi}},$$

where $\varepsilon^2_{\mathsf{Jacobi},t}(x_{\tau_t,0}) = \sum_{i=0}^{K-1}\sum_{n=0}^{N-1}\left\|\frac{\partial s_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}})}{\partial x} - \frac{\partial s^\star_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}})}{\partial x}\right\|^2$.

Remark 1. When $N = K = 1$, these are equivalent to Li et al. (2024c, Assumptions 1 and 2).

4.2 Convergence guarantees

We are now in a position to present our main theoretical guarantees for the proposed algorithm.

Theorem 1. *Suppose that Assumptions 1, 2 and 3 hold. Let $K > 0$ be an arbitrary fixed integer. If $T \geq C_2 d^2 \log^3 T$ and $N = \lceil C_3 \log T \rceil$ for some large enough constants $C_2, C_3 > 0$, then Algorithm 1 achieves*

$$\mathsf{TV}\big(p_{X_1}, p_{Y_1}\big) \leq C_1 d^2 \left(\frac{d \log^2 T}{T}\right)^K \log^2 T + C_1 \sqrt{d \log^4 T}\, \varepsilon_{\mathsf{score}} + C_1 d (\log^{3/2} T) \varepsilon_{\mathsf{Jacobi}}. \tag{17}$$

When exact score functions are available (so that $\varepsilon_{\mathsf{score}} = \varepsilon_{\mathsf{Jacobi}} = 0$), Theorem 1 reveals that the number of iterations needed to approximate q_1 to within ε TV distance is no greater than

$$\text{(iteration complexity)} \qquad \widetilde{O}\left(\frac{d^{1+2/K}}{\varepsilon^{1/K}}\right). \tag{18}$$

Given that K is a fixed constant and $N \asymp \log T$, the total number of score function evaluations of the proposed algorithm is given by $O(NKT) = O(T \log T)$, which can be as small as $\widetilde{O}\big(\frac{d^{1+2/K}}{\varepsilon^{1/K}}\big)$ (cf. (18)). In particular, when K grows, one can approximately interpret this number of function evaluations as $\widetilde{O}\big(\frac{d^{1+o(1)}}{\varepsilon^{1/K}}\big)$, resulting in substantial speed-ups compared to prior theory in Li et al. (2024c,a); Li and Cai (2024); Wu et al. (2024). In the noisy case when we only have access to inexact scores, our result guarantees that the TV distance between our output distribution and the target distribution scales proportionally to the two error metrics $\varepsilon_{\mathsf{score}}$ and $\varepsilon_{\mathsf{Jacobi}}$. It is worth noting that: without assumptions on the smoothness of the score estimates, additional assumptions like the one concerning Jacobian errors cannot be avoided in general, as already illustrated by a counterexample in Li et al. (2023) for the probability flow ODE.

Notably, our approach bears some similarity with other higher-order ODE solvers recently developed in the literature, such as the DPM solver (Lu et al., 2022). While remarkable empirical success of such methods have been reported, their theoretical footings remain significantly underdeveloped. Recently, Huang et al. (2024a) analyzed the theoretical performance of the K-th order Runge-Kutta method for solving the probability flow ODE, for any constant order K. While their theory is comparable to ours, it requires higher-order smoothness on the score estimates (more precisely, they required the score estimate's first $(K+1)$-th derivatives to be bounded). In addition, their iteration complexity also scales linearly in the support size of p_{data}; in contrast, our theory is nearly independent from such a support size as long as it is polynomially bounded. Finally, while we were finalizing our paper, we became aware of a concurrent work Huang et al. (2025), which — under a different set of assumptions — also relaxes the high-order smoothness condition previously imposed in Huang et al. (2024a). Compared with our results, their analysis requires a stronger assumption on the support size, and their total-variation bound depends heavily upon the early-stopping parameter (denoted by τ therein).

5 Main analysis

This section outlines the proof steps for establishing Theorem 1. Here and throughout, we assume

$$\varepsilon_{\mathsf{score}} \leq (C_1 \sqrt{d} \log^2 T)^{-1}, \tag{19a}$$

$$\varepsilon_{\mathsf{Jacobi}} \leq (C_1 d \log^{3/2} T)^{-1}; \tag{19b}$$

otherwise, (17) is trivially satisfied given TV distance is always bounded by 1.

5.1 Additional notation

Before proceeding to the main proof, we find it helpful to introduce an additional set of notation. Fix $x_T \in \mathbb{R}^d$, we slightly abuse notation by letting $\{x^{(n)}_{\tau_{t,i}}\}$ denote the deterministic sequence defined by the update rule

(14) with $x_{\tau_{T,0}} = x_T$. We also define $x_t := x_{\tau_{t,0}}$ for all $1 \leq t \leq T-1$. We let $x_t^\star$ (resp. $Y_t^\star$) denote the solution of ODE (9) at $\tau = \tau_{t,0} = 1 - \overline{\alpha}_t$ with the initial condition $x^\star_{\tau_{t+1,0}} = x_{t+1}$ (resp. $x^\star_{\tau_{t+1,0}} = Y_{t+1}$). For any $\tau \in [\tau_{t,K-1}, \tau_{t,0}] = [1-\overline{\alpha}_{t-1}, 1-\overline{\alpha}_t]$, we denote by $x_\tau^\star$ the solution of ODE (9) at τ with the initial condition $x^\star_{\tau_{t,0}} = x_t$. For any $0 \leq i \leq K-1$, $1 \leq t \leq T$, $1 \leq n \leq N$, we define

$$\varepsilon^{(n)}_{\mathsf{score},t,i}(x^{(n)}_{\tau_{t,i}}) := \big\| s_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}}) - s^\star_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}}) \big\|_2; \quad \varepsilon^{(n)}_{\mathsf{Jacobi},t,i}(x^{(n)}_{\tau_{t,i}}) := \left\| \frac{\partial s_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}})}{\partial x} - \frac{\partial s^\star_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}})}{\partial x} \right\|.$$

It is also defined that

$$\xi_t(x_t) := \frac{\log T}{T}\left(d\sqrt{\sum_{i,n}\big(\varepsilon^{(n)}_{\mathsf{Jacobi},t,i}(x^{(n)}_{\tau_{t,i}})\big)^2} + \sqrt{d\log T}\sqrt{\sum_{i,n}\big(\varepsilon^{(n)}_{\mathsf{score},t,i}(x^{(n)}_{\tau_{t,i}})\big)^2} \right), \tag{20a}$$

$$S_t(x_T) := \sum_{k=2}^{t} \xi_t(x_t) \qquad \text{for } t \geq 2, \qquad \text{and } S_1(x_T) = 0. \tag{20b}$$

From the definitions of $\varepsilon_{\mathsf{score},t}(\cdot)$ and $\varepsilon_{\mathsf{Jacobi},t}(\cdot)$, it follows that

$$\frac{1}{T}\sum_t \mathbb{E}_{Y_t\sim q_t}\left[\varepsilon_{\mathsf{score},t}(Y_t)\right] \leq \sqrt{\frac{1}{T}\sum_t \mathbb{E}_{Y_t\sim q_t}\left[\varepsilon^2_{\mathsf{score},t}(Y_t)\right]} = \sqrt{NK}\varepsilon_{\mathsf{score}} \asymp \varepsilon_{\mathsf{score}}\sqrt{\log T}, \tag{21}$$

$$\frac{1}{T}\sum_t \mathbb{E}_{Y_t\sim q_t}\left[\varepsilon_{\mathsf{Jacobi},t}(Y_t)\right] \leq \sqrt{\frac{1}{T}\sum_t \mathbb{E}_{Y_t\sim q_t}\left[\varepsilon^2_{\mathsf{Jacobi},t}(Y_t)\right]} = \sqrt{NK}\varepsilon_{\mathsf{Jacobi}} \asymp \varepsilon_{\mathsf{Jacobi}}\sqrt{\log T}. \tag{22}$$

5.2 Preliminary facts

The following lemmas will be frequently used throughout the proof. The first lemma introduces some helpful properties concerning the learning rates.

Lemma 1. *For large enough T, one has*

$$\alpha_t \geq 1 - \frac{c_1 \log T}{T} \geq \frac{1}{2}, \qquad 1 \leq t \leq T \tag{23a}$$

$$\frac{1}{2}\frac{1-\alpha_t}{1-\overline{\alpha}_t} \leq \frac{1}{2}\frac{1-\alpha_t}{\alpha_t - \overline{\alpha}_t} \leq \frac{1-\alpha_t}{1-\overline{\alpha}_{t-1}} \leq \frac{4c_1\log T}{T}, \qquad 2 \leq t \leq T \tag{23b}$$

$$1 \leq \frac{1-\overline{\alpha}_t}{1-\overline{\alpha}_{t-1}} \leq 1 + \frac{4c_1 \log T}{T}, \qquad 2 \leq t \leq T \tag{23c}$$

$$\overline{\alpha}_T \leq \frac{1}{T^{c_2}}, \tag{23d}$$

$$\frac{\overline{\alpha}_{t+1}}{1-\overline{\alpha}_{t+1}} \leq \frac{\overline{\alpha}_t}{1-\overline{\alpha}_t} \leq \frac{4\overline{\alpha}_{t+1}}{1-\overline{\alpha}_{t+1}}, \qquad 1 \leq t < T \tag{23e}$$

$$\left|\frac{\tau_{t,i_1} - \tau_{t,i_2}}{\tau_{t,i_3}(1-\tau_{t,i_4})}\right| \leq 8c_1\frac{\log T}{T}, \qquad 2 \leq t \leq T, 0 \leq i_1, i_2, i_3, i_4 \leq K-1, \tag{23f}$$

$$|\gamma_{t,i}(\tau_{t,j})| \leq 2^K(\tau_{t,0} - \tau_{t,j}), \qquad \forall 0 \leq i,j \leq K-1 \tag{23g}$$

$$1 - \tau_{t,i} \asymp 1 - \tau_{t,j}, \qquad \tau_{t,i} \asymp \tau_{t,j}, \qquad \forall 0 \leq i,j \leq K-1, 2 \leq t \leq T. \tag{23h}$$

Here, c_2 is a sufficiently large constant.

For notational convenience, we define

$$\theta_\tau(x) := \max\left\{-\frac{\log p_{\overline{X}_\tau}(x)}{d\log T}, c_6\right\} \tag{24}$$

for some large enough constant $c_6 > 0$. The second lemma establishes the tail bounds of X_0 conditioned on the continuous-time forward process $\overline{X}_\tau$ (cf. (4)).

Lemma 2 (Lemma 1 in Li et al. (2024c)). *Suppose that Assumption 1 holds and $\tau \leq 1 - \frac{1}{T^{c_0}}$. Then for any $c_5 \geq 2$, conditioned on $\overline{X}_\tau = y$, we have*

$$\big\|\sqrt{1-\tau}X_0 - y\big\|_2 \leq 5c_5\sqrt{\theta_\tau(y)d\tau\log T} \tag{25}$$

with probability exceeding $1 - \exp(-c_5^2\theta_\tau(y)d\log T)$. Furthermore, one has

$$\mathbb{E}\left[\big\|\sqrt{1-\tau}X_0 - y\big\|_2 \mid \overline{X}_\tau = y\right] \leq 12\sqrt{\theta_\tau(y)d\tau\log T}, \tag{26a}$$

$$\mathbb{E}\left[\big\|\sqrt{1-\tau}X_0 - y\big\|_2^2 \mid \overline{X}_\tau = y\right] \leq 120\theta_\tau(y)d\tau\log T, \tag{26b}$$

$$\mathbb{E}\left[\big\|\sqrt{1-\tau}X_0 - y\big\|_2^3 \mid \overline{X}_\tau = y\right] \leq 1040\left(\theta_\tau(y)d\tau\log T\right)^{3/2}, \tag{26c}$$

$$\mathbb{E}\left[\big\|\sqrt{1-\tau}X_0 - y\big\|_2^4 \mid \overline{X}_\tau = y\right] \leq 10080\left(\theta_\tau(y)d\tau\log T\right)^{2},. \tag{26d}$$

The next lemma provides key properties of the score function $s_\tau^\star(\cdot)$.

Lemma 3. *One has*

$$\|s_\tau^\star(x)\|_2 \lesssim \sqrt{\frac{d\theta_\tau(x)\log T}{\tau}}, \tag{27a}$$

$$\left\|\frac{\partial s_\tau^\star(x)}{\partial x}\right\| \lesssim \frac{d\theta_\tau(x)\log T}{\tau}. \tag{27b}$$

Furthermore, if $-\log p_{\overline{X}_\tau}(\lambda x_1 + (1-\lambda)x_2) \lesssim d\log T$ for all $\lambda \in [0,1]$, then

$$\left\|\frac{\partial s_\tau^\star(x_1)}{\partial x} - \frac{\partial s_\tau^\star(x_2)}{\partial x}\right\| \lesssim \sqrt{\frac{d^3\log^3 T}{\tau^3}}\|x_1 - x_2\|_2. \tag{28}$$

The proofs of Lemma 1 and Lemma 3 are deferred to Sections B.1 and B.2, respectively.

The following lemma shows that if $-\log p_{\overline{X}_\tau}(x_\tau^\star)$ is not too large, then $-\log p_{\overline{X}_{\tau'}}(x_{\tau'}^\star)$ can be well-controlled for all τ' sufficiently close to τ.

Lemma 4. *Suppose that $-\log p_{\overline{X}_{\widetilde{\tau}}}(x_{\widetilde{\tau}}^\star) \leq \theta d\log T$ for some $\theta > 1$. Then for all $|\tau' - \widetilde{\tau}| \leq c_0\widetilde{\tau}(1-\widetilde{\tau})$ for some sufficiently small constant $c_0 > 0$, one has*

$$-\log p_{\overline{X}_{\tau'}}(x_{\tau'}^\star) \leq 2\theta d\log T. \tag{29}$$

The proof of Lemma 4 can be found in Section B.3. The last lemma reveals that the the distributions of X_T and Y_T are close.

Lemma 5 (Lemma 3 in Li et al. (2024c)). *Suppose that T is large enough. Then*

$$\big(\mathsf{TV}(p_{X_T} \parallel p_{Y_T})\big)^2 \leq \frac{1}{2}\mathsf{KL}(p_{X_T} \parallel p_{Y_T}) \lesssim \frac{1}{T^{200}}. \tag{30}$$

5.3 Main steps for proving Theorem 1

We now proceed to present the proof ideas, which comprises multiple steps as detailed below.

Step 1: controlling the density ratio. To establish the convergence rate (17), a pivotal step is to control the density ratio $\frac{p_{Y_{t-1}}(x_{t-1})}{p_{X_{t-1}}(x_{t-1})}\big/\frac{p_{Y_t}(x_t)}{p_{X_t}(x_t)}$. Our starting point is the following observation:

$$\frac{p_{Y_{t-1}}(x_{t-1})}{p_{X_{t-1}}(x_{t-1})} = \frac{p_{\sqrt{\alpha_t}Y_{t-1}}(\sqrt{\alpha_t}x_{t-1})}{p_{\sqrt{\alpha_t}X_{t-1}}(\sqrt{\alpha_t}x_{t-1})} = \frac{p_{\sqrt{\alpha_t}Y_{t-1}}(\sqrt{\alpha_t}x_{t-1})}{p_{Y_t}(x_t)}\left(\frac{p_{\sqrt{\alpha_t}X_{t-1}}(\sqrt{\alpha_t}x_{t-1})}{p_{X_t}(x_t)}\right)^{-1}\frac{p_{Y_t}(x_t)}{p_{X_t}(x_t)}$$

$$= \frac{p_{\sqrt{\alpha_t}Y_{t-1}}(\sqrt{\alpha_t}x_{t-1})}{p_{\sqrt{\alpha_t}Y_{t-1}^\star}(\sqrt{\alpha_t}x_{t-1}^\star)}\cdot\frac{p_{\sqrt{\alpha_t}Y_{t-1}^\star}(\sqrt{\alpha_t}x_{t-1}^\star)}{p_{Y_t}(x_t)}\left(\frac{p_{\sqrt{\alpha_t}X_{t-1}}(\sqrt{\alpha_t}x_{t-1})}{p_{\sqrt{\alpha_t}X_{t-1}}(\sqrt{\alpha_t}x_{t-1}^\star)}\cdot\frac{p_{\sqrt{\alpha_t}X_{t-1}}(\sqrt{\alpha_t}x_{t-1}^\star)}{p_{X_t}(x_t)}\right)^{-1}\frac{p_{Y_t}(x_t)}{p_{X_t}(x_t)}. \tag{31}$$

By virtue of the defition of $Y^\star_{t-1}$ and the fact that we can transport the distribution of X_t to the one of X_{t-1} via (9), one has

$$\frac{p_{\sqrt{\alpha_t}Y^\star_{t-1}}(\sqrt{\alpha_t}x^\star_{t-1})}{p_{Y_t}(x_t)} = \frac{p_{\sqrt{\alpha_t}X_{t-1}}(\sqrt{\alpha_t}x^\star_{t-1})}{p_{X_t}(x_t)}. \tag{32}$$

In addition, the following lemma provides upper bounds for $\frac{p_{\sqrt{\alpha_t}X_{t-1}}(\sqrt{\alpha_t}x_{t-1})}{p_{\sqrt{\alpha_t}X_{t-1}}(\sqrt{\alpha_t}x^\star_{t-1})}$ and $\frac{p_{\sqrt{\alpha_t}Y_{t-1}}(\sqrt{\alpha_t}x_{t-1})}{p_{\sqrt{\alpha_t}Y^\star_{t-1}}(\sqrt{\alpha_t}x^\star_{t-1})}$, respectively.

Lemma 6. *We have*

$$\frac{p_{\sqrt{\alpha_t}X_{t-1}}(\sqrt{\alpha_t}x_{t-1})}{p_{\sqrt{\alpha_t}X_{t-1}}(\sqrt{\alpha_t}x^\star_{t-1})} = \exp\left(O\left(\frac{\|x_{t-1}-x^\star_{t-1}\|_2^2}{1-\overline{\alpha}_{t-1}} + \sqrt{\frac{d\|x_{t-1}-x^\star_{t-1}\|_2^2\log T}{1-\overline{\alpha}_{t-1}}}\right)\right). \tag{33}$$

If, furthermore, $\left\|\frac{\partial x^{(N)}_{\tau_{t,K-1}}/\sqrt{1-\tau_{t,K-1}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\right\|^{-1} \lesssim 1$, *then*

$$\frac{p_{\sqrt{\alpha_t}Y_{t-1}}(\sqrt{\alpha_t}x_{t-1})}{p_{\sqrt{\alpha_t}Y^\star_{t-1}}(\sqrt{\alpha_t}x^\star_{t-1})} = \exp\left(O\left(d\left\|\frac{\partial x^\star_{\tau_{t,K-1}}/\sqrt{1-\tau_{t,K-1}}}{\partial x^\star_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}} - \frac{\partial x^{(N)}_{\tau_{t,K-1}}/\sqrt{1-\tau_{t,K-1}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\right\|\right)\right). \tag{34}$$

The proof of Lemma 6 is postponed to Section B.4. Combining (31), (32), (33) and (34), we know that if $\left\|\frac{\partial x^{(N)}_{\tau_{t,K-1}}/\sqrt{1-\tau_{t,K-1}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\right\|^{-1} \lesssim 1$, then we have

$$\frac{p_{Y_{t-1}}(x_{t-1})}{p_{X_{t-1}}(x_{t-1})} = \frac{p_{\sqrt{\alpha_t}Y_{t-1}}(\sqrt{\alpha_t}x_{t-1})}{p_{\sqrt{\alpha_t}X_{t-1}}(\sqrt{\alpha_t}x_{t-1})} = \exp(O(\zeta_t(x_t)))\frac{p_{Y_t}(x_t)}{p_{X_t}(x_t)}, \tag{35}$$

where

$$\zeta_t(x_t) := d\left\|\frac{\partial x^\star_{\tau_{t,K-1}}/\sqrt{1-\tau_{t,K-1}}}{\partial x^\star_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}} - \frac{\partial x^{(N)}_{\tau_{t,K-1}}/\sqrt{1-\tau_{t,K-1}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\right\| + \frac{\|x^\star_{\tau_{t,K-1}} - x^{(N)}_{\tau_{t,K-1}}\|_2^2}{1-\overline{\alpha}_{t-1}} + \sqrt{\frac{d\|x^\star_{\tau_{t,K-1}} - x^{(N)}_{\tau_{t,K-1}}\|_2^2\log T}{1-\overline{\alpha}_{t-1}}}. \tag{36}$$

The next lemma reveals that $\zeta_t(x_t)$ is small for "typical" points.

Lemma 7. *We define the events*

$$\mathcal{E}'_t := \left\{\varepsilon^{(n)}_{\mathsf{Jacob},t,i}(x^{(n)}_{\tau_{t,i}})\log T \le T,\ \textit{for all } i,n\right\}$$

and

$$\begin{aligned}\mathcal{E}_t := \mathcal{E}'_t &\cap \{-\log p_{X_{\tau_{t,i}}}\left(\lambda x^{(n)}_{\tau_{t,i}} + (1-\lambda)x^\star_{\tau_{t,i}}\right) \le C_4 d\log T,\ \textit{for all } i,n\} \\ &\cap \{-\log p_{X_\tau}(x^\star_\tau) \le C_4 d\log T \ \textit{for all } \tau_{t,K-1}\le\tau\le\tau_{t,0}\},\end{aligned} \tag{37}$$

where $C_4 > 0$ is a sufficiently large constant. Then on the event $\mathcal{E}_t$, for any $0 \le i \le K-1$, one has

$$\left\|\frac{\partial x^{(N)}_{\tau_{t,K-1}}/\sqrt{1-\tau_{t,K-1}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\right\|^{-1} \lesssim 1, \tag{38a}$$

$$\|x^{(N)}_{\tau_{t,i}} - x^\star_{\tau_{t,i}}\|_2^2 \le C\tau'^2\frac{\log^2 T}{T^2}\sum_{0\le n\le N}\sum_{0\le i<k}\left(\varepsilon^{(n)}_{\mathsf{score},t,i}(x^{(n)}_{\tau_{t,i}})\right)^2 + 3C\frac{d\tau'\log^3 T}{T^2}\left(\frac{d\log^2 T}{T}\right)^{2K}, \tag{38b}$$

$$\begin{aligned}&\left\|\frac{\partial x^{(N)}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}} - \frac{\partial x^\star_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}{\partial x^\star_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\right\|^2 \\ &\quad\lesssim \frac{d^3\log^7 T}{T^4}\sum_{i,n}\left(\varepsilon^{(n)}_{\mathsf{score},t,i}(x^{(n)}_{\tau_{t,i}})\right)^2 + \frac{\log^2 T}{T^2}\tau'^2\sum_{i,n}\left(\varepsilon^{(n)}_{\mathsf{Jacobi},t,i}(x^{(n)}_{\tau_{t,i}})\right)^2 + \left(\frac{d\log^2 T}{T}\right)^{2K+2}.\end{aligned} \tag{38c}$$

The proof of Lemma 7 is deferred to Section A. By virtue of Lemma 7 and (36), one can show that on the event $\mathcal{E}_t \cap \{\xi_t(x_t) \le c\}$, one has

$$\begin{aligned}
&\frac{p_{Y_{t-1}}(x_{t-1})}{p_{X_{t-1}}(x_{t-1})} \\
&\quad = \exp\left(O\left(\frac{d\log T}{T}\sqrt{\sum_{i,n}\big(\varepsilon^{(n)}_{\mathsf{Jacobi},t,i}(x^{(n)}_{\tau_{t,i}})\big)^2} + \frac{\sqrt{d\log^3 T}}{T}\sqrt{\sum_{i,n}\big(\varepsilon^{(n)}_{\mathsf{score},t,i}(x^{(n)}_{\tau_{t,i}})\big)^2} + d\Big(\frac{d\log^2 T}{T}\Big)^{K+1}\right)\right)\frac{p_{Y_t}(x_t)}{p_{X_t}(x_t)} \\
&\quad = \exp\left(O\left(\xi_t(x_t) + d\Big(\frac{d\log^2 T}{T}\Big)^{K+1}\right)\right)\frac{p_{Y_t}(x_t)}{p_{X_t}(x_t)}, \qquad (39)
\end{aligned}$$

provided that $T \gtrsim d^2\log^2 T$. Here, $\xi_t(\cdot)$ is defined in (20a). Furthermore, the following lemma shows that the log densities $\{-\log p_{X_{\tau_{t,i}}}\big(\lambda x^{(n+1)}_{\tau_{t,i}} + (1-\lambda)x^\star_{\tau_{t,i}}\big)\}_{i,n,\lambda\in[0,1]}$ can be well-controlled as long as $x_{\tau_{t,0}}$ is "typical" and the score estimation errors are not extremely large.

Lemma 8. *If* $C_{10}\frac{\theta_{\tau_{t,0}}(x_{\tau_{t,0}})d\log^2 T + \sqrt{\theta_{\tau_{t,0}}(x_{\tau_{t,0}})\sum_{i,n}\big(\varepsilon^{(n)}_{\mathsf{score},t,i}(x^{(n)}_{\tau_{t,i}})\big)^2 d\log^3 T}}{T} \le 1$ *for some large enough constant* $C_{10} > 0$*, then for all* $0 \le i \le K-1, 0 \le n \le N-1$*, one has*

$$-\log p_{X_{\tau_{t,i}}}\big(\lambda x^{(n+1)}_{\tau_{t,i}} + (1-\lambda)x^\star_{\tau_{t,i}}\big) \le 2.1 d\theta_t \log T, \qquad (40a)$$

$$\log\frac{p_{\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,i}}}X_{\tau_{t,i}}}\Big(\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,i}}}x^{(n+1)}_{\tau_{t,i}}\Big)}{p_{X_{\tau_{t,0}}}(x_{\tau_{t,0}})} \le \frac{4c_1 d\log T}{T} + C_{10}\left\{\frac{d^2\theta_t^2\log^4 T}{T^2} + \frac{\sqrt{d\theta_t\log^3 T}}{T}\sqrt{\sum_i\big(\varepsilon^{(n)}_{\mathsf{score},t,i}(x^{(n)}_{\tau_{t,i}})\big)^2}\right\}. \qquad (40b)$$

Step 2: decomposing the TV distance of interest. Next, let us define the sets

$$\mathcal{E} = \{x : q_1(x) > \max\{p_1(x), \exp(-c_6 d\log T)\}\}, \qquad (41)$$

and

$$\mathcal{I}_1 := \{x_T \mid S_T(x_T) \le c_3\} \qquad (42)$$

for some small enough constant $c_3 > 0$. Then we learn from Li et al. (2024c, Eqn. (51) and (53)) that

$$\begin{aligned}
\mathsf{TV}(q_1, p_1) &\le \mathbb{E}_{Y_1 \in p_1}\left[\Big(\frac{q_1(Y_1)}{p_1(Y_1)} - 1\Big)\mathbb{1}\{Y_1 \in \mathcal{E}\}\right] + \exp(-c_6 d\log T) \\
&= \underbrace{\mathbb{E}_{Y_T \in p_T}\left[\Big(\frac{q_1(Y_1)}{p_1(Y_1)} - 1\Big)\mathbb{1}\{Y_1 \in \mathcal{E}, Y_T \in \mathcal{I}_1\}\right]}_{=:\alpha_1} + \underbrace{\mathbb{E}_{Y_T \in p_T}\left[\Big(\frac{q_1(Y_1)}{p_1(Y_1)} - 1\Big)\mathbb{1}\{Y_1 \in \mathcal{E}, Y_T \notin \mathcal{I}_1\}\right]}_{=:\alpha_2} \\
&\quad + \exp(-c_6 d\log T). \qquad (43)
\end{aligned}$$

provided that $c_6 \ge 4(c_R + 2)$. This motivates us to bound α_1 and α_2 separately.

Step 3: bounding α_1. We let $\tau(x_T)$ denote the following quantity:

$$\tau(x_T) := \max\{2 \le t \le T+1 : S_{t-1}(x_T) \le c_3\}. \qquad (44)$$

Then for any $x_T \in \mathcal{I}_1$, one has

$$\tau(x_T) = T + 1. \qquad (45)$$

In addition, we have the following two important properties regarding $\tau(x_T)$:

Lemma 9. *If $-\log q_1(x_1) \le c_6 d \log T$, then for all $1 \le \ell < \tau(x_T)$, one has*

$$-\log q_\ell(x_\ell) \le 2c_6 d \log T \tag{46}$$

as long as $c_6 > 5c_1$.

Lemma 10. *For $\tau = \tau(x_T)$, We have*

$$\frac{q_1(x_1)}{p_1(x_1)} = \left(1 + O\left(\sum_{t<\tau} \xi_t(x_t) + d^2 \log^2 T \left(\frac{d \log^2 T}{T}\right)^K\right)\right) \frac{q_{\tau-1}(x_{\tau-1})}{p_{\tau-1}(x_{\tau-1})}, \tag{47}$$

$$\frac{q_\ell(x_\ell)}{2p_\ell(x_\ell)} \le \frac{q_1(x_1)}{p_1(x_1)} \le \frac{2q_\ell(x_\ell)}{p_\ell(x_\ell)}, \quad \forall \ell < \tau. \tag{48}$$

The proofs of Lemma 9 and Lemma 10 are postponed to Sections B.6 and B.7, respectively. By virtue of (45) and Lemma 10, one can show that

$$\begin{aligned}
&\mathbb{E}_{Y_T \in p_T}\left[\left(\frac{q_1(Y_1)}{p_1(Y_1)} - 1\right) \mathbb{1}\{Y_1 \in \mathcal{E}, Y_T \in \mathcal{I}_1\}\right] \\
&= \mathbb{E}_{Y_T \in p_T}\left[\left(\left(1 + \sum_t \xi_t(Y_t) + d^2 \log^2 T \left(\frac{d \log^2 T}{T}\right)^K\right) \frac{q_T(Y_T)}{p_T(Y_T)} - 1\right) \mathbb{1}\{Y_1 \in \mathcal{E}, Y_T \in \mathcal{I}_1\}\right] \\
&= \int \left\{\left(\left(1 + \sum_t \xi_t(x_t) + d^2 \log^2 T \left(\frac{d \log^2 T}{T}\right)^K\right) q_T(x_T) - p_T(x_T)\right) \mathbb{1}\{x_1 \in \mathcal{E}, x_T \in \mathcal{I}_1\}\right\} \mathrm{d}x_T \\
&\le \int |q_T(x_T) - p_T(x_T)| \mathrm{d}x_T + d^2 \log^2 T \left(\frac{d \log^2 T}{T}\right)^K + \int S_T(x_T) q_T(x_T) \mathbb{1}\{x_1 \in \mathcal{E}, x_T \in \mathcal{I}_1\} \mathrm{d}x_T.
\end{aligned} \tag{49}$$

Additionally, we make the observation that

$$\begin{aligned}
&\int S_T(x_T) q_T(x_T) \mathbb{1}\{x_1 \in \mathcal{E}, x_T \in \mathcal{I}_1\} \mathrm{d}x_T \\
&= \sum_{t=2}^T \int \frac{\log T}{T} \left(d \sqrt{\sum_{i,n} \left(\varepsilon^{(n)}_{\mathsf{Jacobi},t,i}(x^{(n)}_{\tau_{t,i}})\right)^2} + \sqrt{d \log T} \sqrt{\sum_{i,n} \left(\varepsilon^{(n)}_{\mathsf{score},t,i}(x^{(n)}_{\tau_{t,i}})\right)^2}\right) \mathbb{1}\{x_1 \in \mathcal{E}, x_T \in \mathcal{I}_1\} \mathrm{d}x_T \\
&= \sum_{t=2}^T \mathbb{E}_{Y_T \sim p_T}\left[\frac{\log T}{T} \left(d\varepsilon_{\mathsf{Jacobi},t}(Y_t) + \sqrt{d \log T} \varepsilon_{\mathsf{score},t}(Y_t)\right) \frac{q_t(Y_T)}{p_t(Y_T)} \mathbb{1}\{x_1 \in \mathcal{E}, x_T \in \mathcal{I}_1\}\right] \\
&\overset{(48)}{\le} 4 \sum_{t=2}^T \mathbb{E}_{Y_T \sim p_T}\left[\frac{\log T}{T} \left(d\varepsilon_{\mathsf{Jacobi},t}(Y_t) + \sqrt{d \log T} \varepsilon_{\mathsf{score},t}(Y_t)\right) \frac{q_t(Y_t)}{p_t(Y_t)} \mathbb{1}\{x_1 \in \mathcal{E}, x_T \in \mathcal{I}_1\}\right] \\
&\le 4 \sum_{t=2}^T \mathbb{E}_{Y_t \sim p_t}\left[\frac{\log T}{T} \left(d\varepsilon_{\mathsf{Jacobi},t}(Y_t) + \sqrt{d \log T} \varepsilon_{\mathsf{score},t}(Y_t)\right) \frac{q_t(Y_t)}{p_t(Y_t)}\right] \\
&= 4 \sum_{t=2}^T \mathbb{E}_{Y_t \sim q_t}\left[\frac{\log T}{T} \left(d\varepsilon_{\mathsf{Jacobi},t}(Y_t) + \sqrt{d \log T} \varepsilon_{\mathsf{score},t}(Y_t)\right)\right] \\
&\overset{(21) \text{ and } (22)}{\lesssim} d \log^{3/2} T \varepsilon_{\mathsf{Jacobi}} + \sqrt{d \log^4 T} \varepsilon_{\mathsf{score}}.
\end{aligned} \tag{50}$$

Putting the previous two inequalities together, one arrives at

$$\begin{aligned}
\alpha_1 &= \mathbb{E}_{Y_T \in p_T}\left[\left(\frac{q_1(Y_1)}{p_1(Y_1)} - 1\right) \mathbb{1}\{Y_1 \in \mathcal{E}, Y_T \in \mathcal{I}_1\}\right] \\
&\lesssim d^2 \log^2 T \left(\frac{d \log^2 T}{T}\right)^K + d \log^{3/2} T \varepsilon_{\mathsf{Jacobi}} + \sqrt{d \log^4 T} \varepsilon_{\mathsf{score}}.
\end{aligned} \tag{51}$$

Step 4: bounding α_2. Now, we turn to bounding α_2. We let $\tau = \tau(x_T)$ define the following sets:

$$\mathcal{I}_2 := \{x_T \mid c_3 \le S_\tau(x_T) \le 2c_3\}, \tag{52}$$

$$\mathcal{I}_3 := \left\{x_T \mid S_{\tau-1}(x_T) \le c_3, \xi_\tau(x_T) \ge c_3, \frac{q_{\tau-1}(x_{\tau-1})}{p_{\tau-1}(x_{\tau-1})} \le 8\frac{q_\tau(x_\tau)}{p_\tau(x_\tau)}\right\}, \tag{53}$$

$$\mathcal{I}_4 := \left\{x_T \mid S_{\tau-1}(x_T) \le c_3, \xi_\tau(x_T) \ge c_3, \frac{q_{\tau-1}(x_{\tau-1})}{p_{\tau-1}(x_{\tau-1})} > 8\frac{q_\tau(x_\tau)}{p_\tau(x_\tau)}\right\}, \tag{54}$$

It is straightforward to verify that $\mathcal{I}_1 \cup \mathcal{I}_2 \cup \mathcal{I}_3 \cup \mathcal{I}_4 = \mathbb{R}^d$. The following lemma indicates that α_2 can be well-controlled.

Lemma 11. *The following inequality holds:*

$$\mathbb{E}_{Y_T \sim p_T}\left[\Big(\frac{q_1(Y_1)}{p_1(Y_1)} - 1\Big)\mathbb{1}\{Y_1 \in \mathcal{E}, Y_T \in \mathcal{I}_2 \cup \mathcal{I}_3 \cup \mathcal{I}_4\}\right] \lesssim d^2 \log^2 T \left(\frac{d \log^2 T}{T}\right)^K + d \log^{3/2} T \varepsilon_{\mathsf{Jacobi}} + \sqrt{d \log^4 T}\varepsilon_{\mathsf{score}}. \tag{55}$$

An immediate consequence of Lemma 11 is

$$\begin{aligned}
\alpha_2 &= \mathbb{E}_{Y_T \in p_T}\left[\Big(\frac{q_1(Y_1)}{p_1(Y_1)} - 1\Big)\mathbb{1}\{Y_1 \in \mathcal{E}, Y_T \notin \mathcal{I}_1\}\right] \\
&\le \mathbb{E}_{Y_T \in p_T}\left[\Big(\frac{q_1(Y_1)}{p_1(Y_1)} - 1\Big)\mathbb{1}\{Y_1 \in \mathcal{E}, Y_T \in \mathcal{I}_2 \cup \mathcal{I}_3 \cup \mathcal{I}_4\}\right] \\
&\lesssim d^2 \log^2 T \left(\frac{d \log^2 T}{T}\right)^K + d \log^{3/2} T \varepsilon_{\mathsf{Jacobi}} + \sqrt{d \log^4 T}\varepsilon_{\mathsf{score}}.
\end{aligned} \tag{56}$$

Here, the second line comes from the facts $q_1(x) \ge p_1(x)$ for any $x \in \mathcal{E}$ and $\mathcal{I}_1^c \subseteq \mathcal{I}_2 \cup \mathcal{I}_3 \cup \mathcal{I}_4$.

Step 5: bounding $\mathsf{TV}(p_1, q_1)$ Combining (43), (51) and (56), one arrives at the advertised result:

$$\begin{aligned}
\mathsf{TV}(p_1, q_1) &\lesssim d^2 \log^2 T \left(\frac{d \log^2 T}{T}\right)^K + d \log^{3/2} T \varepsilon_{\mathsf{Jacobi}} + \sqrt{d \log^4 T}\varepsilon_{\mathsf{score}} + \exp(-c_6 d \log T) \\
&\asymp d^2 \log^2 T \left(\frac{d \log^2 T}{T}\right)^K + d \log^{3/2} T \varepsilon_{\mathsf{Jacobi}} + \sqrt{d \log^4 T}\varepsilon_{\mathsf{score}}.
\end{aligned} \tag{57}$$

6 Discussion and future directions

In this work, we have proposed an algorithm to speed up data generation in diffusion models. The proposed algorithm exploits K-th order ODE approximation combined with a successive refining scheme, which provably yields ε-precision (measured by the TV distance between the output distribution and the target distribution) within $\widetilde{O}\big(d^{1+2/K}/\varepsilon^{1/K}\big)$ iterations. Our theoretical guarantees hold under fairly mild assumptions on the target distribution, with no restrictions on smoothness and log-concavity, and are capable of accommodating imperfect score estimation. Our findings leave open several interesting questions for future investigation. For instance, while our theory allows the order K to be an arbitrary large constant, it is worth exploring whether K can be allowed to scale with T (e.g., whether K can reach the order of $\log T$), as accomplishing this might potentially lead to further acceleration. Additionally, our method is a deterministic sampler designed to speed up diffusion ODE, and it remains unclear whether a similar iteration complexity can be achieved with an accelerated stochastic sampler (i.e., the SDE counterpart). Furthermore, while this work focuses on the iteration complexity of the sampling phase, how to develop an end-to-end theory that also incorporates the score matching phase forms another direction for future study. Finally, if the target data distribution exhibits intrinsic low dimensionality, it would be interesting to examine whether the sampling process be further expedited.

References

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

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

Benton, J., Bortoli, V. D., Doucet, A., and Deligiannidis, G. (2024). Nearly d-linear convergence bounds for diffusion models via stochastic localization. In *The Twelfth International Conference on Learning Representations*.

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

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

Chen, S., Chewi, S., Lee, H., Li, Y., Lu, J., and Salim, A. (2023b). The probability flow ODE is provably fast. *arXiv preprint arXiv:2305.11798*.

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

Chen, S., Daras, G., and Dimakis, A. G. (2023c). Restoration-degradation beyond linear diffusions: A non-asymptotic analysis for DDIM-type samplers. *arXiv preprint arXiv:2303.03384*.

Cheng, X., Lu, J., Tan, Y., and Xie, Y. (2023). Convergence of flow-based generative models via proximal gradient descent in wasserstein space. *arXiv preprint arXiv:2310.17582*.

Croitoru, F.-A., Hondru, V., Ionescu, R. T., and Shah, M. (2023). Diffusion models in vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9):10850–10869.

Efron, B. (2011). Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614.

Gao, X. and Zhu, L. (2024). Convergence analysis for general probability flow odes of diffusion models in wasserstein distances. *arXiv preprint arXiv:2401.17958*.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11):139–144.

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

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

Huang, D. Z., Huang, J., and Lin, Z. (2024a). Convergence analysis of probability flow ode for score-based generative models. *arXiv preprint arXiv:2404.09730*.

Huang, D. Z., Huang, J., and Lin, Z. (2025). Fast convergence for high-order ode solvers in diffusion probabilistic models. *arXiv preprint arXiv:2506.13061*.

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

Ipsen, I. C. and Rehman, R. (2008). Perturbation bounds for determinants and characteristic polynomials. *SIAM Journal on Matrix Analysis and Applications*, 30(2):762–776.

Jiao, Y. and Li, G. (2024). Instance-dependent convergence theory for diffusion models. *arXiv e-prints*, pages arXiv–2410.

Jolicoeur-Martineau, A., Li, K., Piché-Taillefer, R., Kachman, T., and Mitliagkas, I. (2021). Gotta go fast when generating data with score-based models. *arXiv preprint arXiv:2105.14080*.

Kingma, D. P. and Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

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

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

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

Li, G., Huang, Y., Efimov, T., Wei, Y., Chi, Y., and Chen, Y. (2024a). Accelerating convergence of score-based diffusion models, provably. *arXiv preprint arXiv:2403.03852*.

Li, G., Huang, Z., and Wei, Y. (2024b). Towards a mathematical theory for consistency training in diffusion models. *arXiv preprint arXiv:2402.07802*.

Li, G. and Jiao, Y. (2024). Improved convergence rate for diffusion probabilistic models. In *The Thirteenth International Conference on Learning Representations*.

Li, G., Wei, Y., Chen, Y., and Chi, Y. (2023). Towards faster non-asymptotic convergence for diffusion-based generative models. *arXiv preprint arXiv:2306.09251*.

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

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

Li, G. and Yan, Y. (2024b). $O(d/T)$ convergence theory for diffusion probabilistic models under minimal assumptions. *arXiv preprint arXiv:2409.18959*.

Li, R., Di, Q., and Gu, Q. (2024d). Unified convergence analysis for score-based diffusion models with deterministic samplers. *arXiv preprint arXiv:2410.14237*.

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

Liang, Y., Ju, P., Liang, Y., and Shroff, N. (2024). Broadening target distributions for accelerated diffusion models via a novel analysis approach. *arXiv preprint arXiv:2402.13901*.

Liu, X., Wu, L., Ye, M., and Liu, Q. (2022). Let us build bridges: Understanding and extending diffusion generative models. *arXiv preprint arXiv:2208.14699*.

Lu, C., Zhou, Y., Bao, F., Chen, J., Li, C., and Zhu, J. (2022). Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps. *Advances in Neural Information Processing Systems*, 35:5775–5787.

Luhman, E. and Luhman, T. (2021). Knowledge distillation in iterative generative models for improved sampling speed. *arXiv preprint arXiv:2101.02388*.

Meng, C., Rombach, R., Gao, R., Kingma, D., Ermon, S., Ho, J., and Salimans, T. (2023). On distillation of guided diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14297–14306.

Nichol, A. Q. and Dhariwal, P. (2021). Improved denoising diffusion probabilistic models. In *International conference on machine learning*, pages 8162–8171. PMLR.

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

Salimans, T. and Ho, J. (2022). Progressive distillation for fast sampling of diffusion models. In *International Conference on Learning Representations*.

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

Song, J., Meng, C., and Ermon, S. (2020). Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*.

Song, Y., Dhariwal, P., Chen, M., and Sutskever, I. (2023). Consistency models.

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

Tang, J. and Yan, Y. (2025). Adaptivity and convergence of probability flow odes in diffusion generative models. *arXiv preprint arXiv:2501.18863*.

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

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

Xue, S., Yi, M., Luo, W., Zhang, S., Sun, J., Li, Z., and Ma, Z.-M. (2024). Sa-solver: Stochastic adams solver for fast sampling of diffusion models. *Advances in Neural Information Processing Systems*, 36.

Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., Zhang, W., Cui, B., and Yang, M.-H. (2023). Diffusion models: A comprehensive survey of methods and applications. *ACM Computing Surveys*, 56(4):1–39.

Zhang, Q. and Chen, Y. (2023). Fast sampling of diffusion models with exponential integrator. In *International Conference on Learning Representations*.

Zhao, W., Bai, L., Rao, Y., Zhou, J., and Lu, J. (2024). Unipc: A unified predictor-corrector framework for fast sampling of diffusion models. *Advances in Neural Information Processing Systems*, 36.

Zheng, K., Lu, C., Chen, J., and Zhu, J. (2023). Dpm-solver-v3: Improved diffusion ode solver with empirical model statistics. *Advances in Neural Information Processing Systems*, 36:55502–55542.

A Proof of Lemma 7

For $\tau_{t-1,0} \le \tau < \tau_{t,0}$, we let $x^\star_\tau$ denote the solution of (9) with the initial condition $x^\star_{\tau_{t,0}} = x_{\tau_{t,0}}$. We let

$$\begin{aligned}\xi^{(n)}_{\mathsf{score},\tau'} &:= \frac{x^{(n+1)}_{\tau'}}{\sqrt{1-\tau'}} - \frac{x_{\tau_{t,0}}}{\sqrt{1-\tau_{t,0}}} - \sum_{0\le i<K} \gamma_{t,i}(\tau')(1-\tau_{t,i})^{-3/2} s^\star_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}}) \\ &= \sum_{0\le i<K} \gamma_{t,i}(\tau')(1-\tau_{t,i})^{-3/2}\big[s_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}}) - s^\star_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}})\big].\end{aligned}$$

Then one can derive

$$\begin{aligned}&\frac{x^{(n+1)}_{\tau'}}{\sqrt{1-\tau'}} - \frac{x^\star_{\tau'}}{\sqrt{1-\tau'}} \\ &= \Big(\frac{x^{(n+1)}_{\tau'}}{\sqrt{1-\tau'}} - \frac{x_{\tau_{t,0}}}{\sqrt{1-\tau_{t,0}}}\Big) - \Big(\frac{x^\star_{\tau'}}{\sqrt{1-\tau'}} - \frac{x^\star_{\tau_{t,0}}}{\sqrt{1-\tau_{t,0}}}\Big) \\ &= \sum_{0\le i<K} \gamma_{t,i}(\tau')(1-\tau_{t,i})^{-3/2} s_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}}) + \int_{\tau_{t,0}}^{\tau'} \frac{1}{2(1-\tau)^{3/2}} s^\star_\tau \mathrm{d}\tau \\ &= \sum_{0\le i<K} \gamma_{t,i}(\tau')(1-\tau_{t,i})^{-3/2}\big[s_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}}) - s^\star_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}})\big] \\ &\quad + \left[\sum_{0\le i<K} \gamma_{t,i}(\tau')(1-\tau_{t,i})^{-3/2} s^\star_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}}) - \left(-\int_{\tau_{t,0}}^{\tau'} \frac{1}{2(1-\tau)^{3/2}} s^\star_\tau(x^\star_\tau)\mathrm{d}\tau\right)\right] \\ &= \xi^{(n)}_{\mathsf{score},\tau'} + \sum_{0\le i<K} \gamma_{t,i}(\tau')(1-\tau_{t,i})^{-3/2}\big(s^\star_{\tau_{t,i}}(x^\star_{\tau_{t,i}}) - s^\star_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}})\big) \\ &\quad + \left[\sum_{0\le i<K} \gamma_{t,i}(\tau')(1-\tau_{t,i})^{-3/2} s^\star_{\tau_{t,i}}(x^\star_{\tau_{t,i}}) - \left(-\int_{\tau_{t,0}}^{\tau'} \frac{1}{2(1-\tau)^{3/2}} s^\star_\tau(x^\star_\tau)\mathrm{d}\tau\right)\right]. \end{aligned} \tag{58}$$

Furthermore, we make the observation that

$$\begin{aligned}\frac{\partial x^{(n+1)}_{\tau'}/\sqrt{1-\tau'}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}} &= I + \sum_{0\le i<K} \gamma_{t,i}(\tau')(1-\tau_{t,i})^{-3/2} \frac{\partial s_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}})}{\partial x^{(n)}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}} \frac{\partial x^{(n)}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}} \\ &=: I + \sum_{0\le i<K} \gamma_{t,i}(\tau')(1-\tau_{t,i})^{-3/2} \frac{\partial s^\star_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}})}{\partial x^{(n)}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}} \frac{\partial x^{(n)}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}} + \xi^{(n)}_{\mathsf{Jacob},\tau'}. \end{aligned} \tag{59}$$

and

$$\frac{\partial x^\star_{\tau'}/\sqrt{1-\tau'}}{\partial x^\star_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}} = I - \int_{\tau_{t,0}}^{\tau'} \frac{1}{2(1-\tau)^{3/2}} \frac{\partial s^\star_\tau(x^\star_\tau)}{\partial x^\star_\tau/\sqrt{1-\tau}} \frac{\partial x^\star_\tau/\sqrt{1-\tau}}{\partial x^\star_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\mathrm{d}\tau.$$

Combining the previous two equations, we have

$$\begin{aligned}&\frac{\partial x^{(n+1)}_{\tau'}/\sqrt{1-\tau'}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}} - \frac{\partial x^\star_{\tau'}/\sqrt{1-\tau'}}{\partial x^\star_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}} \\ &= \xi^{(n)}_{\mathsf{Jacob},\tau'} + \sum_{0\le i<K} \gamma_{t,i}(\tau')(1-\tau_{t,i})^{-3/2} \frac{\partial s^\star_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}})}{\partial x^{(n)}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}} \frac{\partial x^{(n)}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}} \\ &\quad + \int_{\tau_{t,0}}^{\tau'} \frac{1}{2(1-\tau)^{3/2}} \frac{\partial s^\star_\tau(x^\star_\tau)}{\partial x^\star_\tau/\sqrt{1-\tau}} \frac{\partial x^\star_\tau/\sqrt{1-\tau}}{\partial x^\star_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\mathrm{d}\tau. \end{aligned} \tag{60}$$

Controlling estimation error caused by score estimation. Cauchy-Schwarz inequality tells us that

$$\begin{aligned}
\|\xi^{(n)}_{\mathsf{score},\tau'}\|_2^2 &\le \Big(\sum_{0\le i<K}\gamma_{t,i}(\tau')(1-\tau_{t,i})^{-3/2}\varepsilon^{(n)}_{\mathsf{score},t,i}(x^{(n)}_{\tau_{t,i}})\Big)^2 \\
&\le K\sum_{0\le i<K}\gamma_{t,i}^2(\tau')(1-\tau_{t,i})^{-3}\big\|s_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}})-s^\star_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}})\big\|_2^2 \\
&\le K\sum_{0\le i<K}2^{2K}(\tau'-\tau_{t,0})^2(1-\tau_{t,i})^{-3}\big\|s_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}})-s^\star_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}})\big\|_2^2 \\
&\asymp \sum_{0\le i<K}(\tau'-\tau_{t,0})^2(1-\tau_{t,i})^{-3}\big\|s_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}})-s^\star_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}})\big\|_2^2. \qquad (61)
\end{aligned}$$

here, the penultimate line makes use of (23g) and the last line holds since K is a constant.

We claim that on the event $\mathcal{E}_t$, for all $0\le n\le N$, we have

$$\left\|\frac{\partial x^{(n)}_{\tau'}/\sqrt{1-\tau'}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\right\|\asymp 1, \qquad (62a)$$

$$\|\xi^{(n)}_{\mathsf{Jacob},\tau'}\|\le\frac{1}{2}. \qquad (62b)$$

If (62a) and (62b) hold, then on the event $\mathcal{E}_t$, we have

$$\begin{aligned}
\|\xi^{(n)}_{\mathsf{Jacob},\tau'}\|^2 &= \left\|\sum_{0\le i<K}\gamma_{t,i}(\tau')(1-\tau_{t,i})^{-3/2}\left(\frac{\partial s_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}})}{\partial x^{(n)}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}-\frac{\partial s^\star_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}})}{\partial x^{(n)}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}\right)\frac{\partial x^{(n)}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\right\|^2 \\
&\le \left(\sum_{0\le i<K}\gamma_{t,i}(\tau')(1-\tau_{t,i})^{-3/2}\left\|\frac{\partial s_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}})}{\partial x^{(n)}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}-\frac{\partial s^\star_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}})}{\partial x^{(n)}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}\right\|\left\|\frac{\partial x^{(n)}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\right\|\right)^2 \\
&\lesssim \left(\sum_{0\le i<K}\gamma_{t,i}(\tau')(1-\tau_{t,i})^{-3/2}\left\|\frac{\partial s_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}})}{\partial x^{(n)}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}-\frac{\partial s^\star_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}})}{\partial x^{(n)}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}\right\|\right)^2 \\
&= \left(\sum_{0\le i<K}\gamma_{t,i}(\tau')(1-\tau_{t,i})^{-1}\varepsilon^{(n)}_{\mathsf{Jacobi},t,i}(x^{(n)}_{\tau_{t,i}})\right)^2 \\
&\overset{\text{Cauchy-Schwarz}}{\le} K\sum_{0\le i<K}\gamma^2_{t,i}(\tau')(1-\tau_{t,i})^{-2}\big(\varepsilon^{(n)}_{\mathsf{Jacobi},t,i}(x^{(n)}_{\tau_{t,i}})\big)^2 \\
&\overset{(23g)}{\lesssim} \sum_{0\le i<K}(\tau'-\tau_{t,0})^2(1-\tau_{t,i})^{-2}\big(\varepsilon^{(n)}_{\mathsf{Jacobi},t,i}(x^{(n)}_{\tau_{t,i}})\big)^2 \\
&\overset{\text{Lemma 1}}{\lesssim} \frac{\log^2 T}{T^2}\tau'^2\sum_{0\le i<K}\big(\varepsilon^{(n)}_{\mathsf{Jacobi},t,i}(x^{(n)}_{\tau_{t,i}})\big)^2, \qquad (63)
\end{aligned}$$

Controlling discretization error. We claim that if $-p_{\overline{X}_\tau}(x^\star_\tau)\lesssim d\theta\log T$ for $\theta\ge c_6$, then for any interger $k\ge 0$, there exists a constant $c_k>0$ depending only on k such that

$$\left\|\frac{\partial^k}{\partial\tau^k}\frac{s^\star_\tau(x^\star_\tau)}{(1-\tau)^{3/2}}\right\|_2\le c_k\sqrt{\frac{d\theta\log T}{\tau(1-\tau)^3}}\left(\frac{d\theta\log T}{\tau(1-\tau)}\right)^k, \qquad (64a)$$

and

$$\left\|\frac{\partial^k}{\partial\tau^k}\frac{1}{(1-\tau)^{3/2}}\frac{\partial s^\star_\tau(x^\star_\tau)}{\partial x^\star_\tau/\sqrt{1-\tau}}\frac{\partial x^\star_\tau/\sqrt{1-\tau}}{\partial x^\star_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\right\|\le c_k\left(\frac{d\theta\log T}{\tau(1-\tau)}\right)^{k+1}, \qquad (64b)$$

which we prove by induction at the end of this proof. Then Taylor's Theorem together with the fact that $\tau \asymp \tau_{t,0}$ and $1-\tau \asymp 1-\tau_{t,0}$ for all $\tau' \leq \tau \leq \tau_{t,0}$ tells us that

$$
\begin{aligned}
&\frac{x^\star_{\tau'}}{\sqrt{1-\tau'}} \\
&\quad= \frac{x^\star_{\tau_{t,0}}}{\sqrt{1-\tau_{t,0}}} - \int_{\tau_{t,0}}^{\tau'} \frac{1}{2(1-\tau)^{3/2}} s^\star_\tau \mathrm{d}\tau \\
&\quad= \frac{x^\star_{\tau_{t,0}}}{\sqrt{1-\tau_{t,0}}} + \sum_{0\leq i<K} \gamma_{t,i}(\tau')(1-\tau_{t,i})^{-3/2} s^\star_{\tau_{t,i}}(x^\star_{\tau_{t,i}}) + O\left(\sqrt{\frac{d(\tau'-\tau_{t,0})^2\log T}{\tau'(1-\tau')^3}}\left(\frac{d(\tau_{t,0}-\tau')\log T}{\tau'(1-\tau')}\right)^K\right),
\end{aligned} \tag{65}
$$

and

$$
\begin{aligned}
&\frac{\partial x^\star_{\tau'}/\sqrt{1-\tau'}}{\partial x^\star_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}} \\
&\quad= I - \int_{\tau_{t,0}}^{\tau'} \frac{1}{2(1-\tau)^{3/2}} \frac{\partial s^\star_\tau(x^\star_\tau)}{\partial x^\star_\tau/\sqrt{1-\tau}} \frac{\partial x^\star_\tau/\sqrt{1-\tau}}{\partial x^\star_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}} \mathrm{d}\tau \\
&\quad= I + \sum_{0\leq i<K} \gamma_{t,i}(\tau')(1-\tau_{t,i})^{-3/2} \frac{\partial s^\star_{\tau_{t,i}}(x^\star_{\tau_{t,i}})}{\partial x^\star_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}} \frac{\partial x^\star_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}{\partial x^\star_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}} + O\left(\left(\frac{d(\tau_{t,0}-\tau')\log T}{\tau'(1-\tau')}\right)^{K+1}\right).
\end{aligned} \tag{66}
$$

Putting everything together. Combing (58), (65) and Lemma 3, one has

$$
\begin{aligned}
&\left\|\frac{x^{(n+1)}_{\tau'}}{\sqrt{1-\tau'}} - \frac{x^\star_{\tau'}}{\sqrt{1-\tau'}}\right\|_2 \\
&\quad\lesssim \|\xi^{(n)}_{\mathsf{score},\tau'}\|_2 + \sum_{0\leq i<K} |\gamma_{t,i}(\tau')|(1-\tau_{t,i})^{-3/2}\big\|s^\star_{\tau_{t,i}}(x^\star_{\tau_{t,i}}) - s^\star_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}})\big\|_2 \\
&\qquad+ \left\|\sum_{0\leq i<K} \gamma_{t,i}(\tau')(1-\tau_{t,i})^{-3/2} s^\star_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}}) - \left(-\int_{\tau_{t,0}}^{\tau'} \frac{1}{2(1-\tau)^{3/2}} s^\star_\tau(x^\star_\tau)\mathrm{d}\tau\right)\right\|_2 \\
&\quad\lesssim \|\xi^{(n)}_{\mathsf{score},\tau'}\|_2 + \sum_{0\leq i<K} 2^K(\tau_{t,0}-\tau')(1-\tau_{t,i})^{-3/2}\big\|s^\star_{\tau_{t,i}}(x^\star_{\tau_{t,i}}) - s^\star_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}})\big\|_2 \\
&\qquad+ O\left(\sqrt{\frac{d(\tau'-\tau_{t,0})^2\log T}{\tau'(1-\tau')^3}}\left(\frac{d(\tau_{t,0}-\tau')\log T}{\tau'(1-\tau')}\right)^K\right) \\
&\quad\lesssim \sum_{0\leq i<K} \frac{d(\tau_{t,0}-\tau')\log T}{\tau_{t,i}(1-\tau_{t,i})}\left\|\frac{x^{(n)}_{\tau_{t,i}}}{\sqrt{1-\tau'}} - \frac{x^\star_{\tau_{t,i}}}{\sqrt{1-\tau'}}\right\|_2 \\
&\qquad+ \|\xi^{(n)}_{\mathsf{score},\tau'}\|_2 + O\left(\sqrt{\frac{d(\tau'-\tau_{t,0})^2\log T}{\tau'(1-\tau')^3}}\left(\frac{d(\tau_{t,0}-\tau')\log T}{\tau'(1-\tau')}\right)^K\right) \\
&\quad\lesssim \frac{d\log^2 T}{T}\sum_{0\leq i<K}\left\|\frac{x^{(n)}_{\tau_{t,i}}}{\sqrt{1-\tau'}} - \frac{x^\star_{\tau_{t,i}}}{\sqrt{1-\tau'}}\right\|_2 + \|\xi^{(n)}_{\mathsf{score},\tau'}\|_2 + O\left(\sqrt{\frac{d\tau'\log^3 T}{T^2(1-\tau')}}\left(\frac{d\log^2 T}{T}\right)^K\right),
\end{aligned}
$$

Here, the second inequality makes use of (23g) and Lemma 1; the third inequality is valid due to Lemma 3; the last line holds because of Lemma 1. Recalling that K is a constant, we know there exists constants $C>0$ such that

$$
\|x^{(n+1)}_{\tau'} - x^\star_{\tau'}\|_2^2 \leq C\left[\left(\frac{d\log^2 T}{T}\right)^2 \sum_{0\leq i<K}\|x^{(n)}_{\tau_{t,i}} - x^\star_{\tau_{t,i}}\|_2^2 + (1-\tau')\|\xi^{(n)}_{\mathsf{score},\tau'}\|_2^2 + \frac{d\tau'\log^3 T}{T^2}\left(\frac{d\log^2 T}{T}\right)^{2K}\right]
$$

$$\overset{(61)}{\leq} \frac{1}{3}\max_{0\leq i\leq K}\big\|x^{(n)}_{\tau_{t,i}} - x^\star_{\tau_{t,i}}\big\|_2^2 + C\tau'^2\frac{\log^2 T}{T^2}\sum_{0\leq i<K}\big(\varepsilon^{(n)}_{\mathsf{score},t,i}(x^{(n)}_{\tau_{t,i}})\big)^2 + C\frac{d\tau'\log^3 T}{T^2}\bigg(\frac{d\log^2 T}{T}\bigg)^{2K},$$

Then it follows by induction that, for all $N'\geq 1$,

$$\big\|x^{(N')}_{\tau'} - x^\star_{\tau'}\big\|_2^2 \leq \frac{1}{3^{N'}}\max_{0\leq i\leq K}\big\|x^{(0)}_{\tau_{t,i}} - x^\star_{\tau_{t,i}}\big\|_2^2 + C\tau'^2\frac{\log^2 T}{T^2}\sum_{0\leq n\leq N'}\sum_{0\leq i<K}\big(\varepsilon^{(n)}_{\mathsf{score},t,i}(x^{(n)}_{\tau_{t,i}})\big)^2$$
$$+ 2C\frac{d\tau'\log^3 T}{T^2}\bigg(\frac{d\log^2 T}{T}\bigg)^{2K}. \tag{67}$$

Recalling the definition of $x^\star_{\tau_{t,i}}$ and $x^{(0)}_{\tau_{t,i}}$, we invoke Lemma 3 to obtain

$$\big\|x^\star_{\tau_{t,i}} - x^{(0)}_{\tau_{t,i}}\big\|_2 = \big\|x^\star_{\tau_{t,i}} - x_{\tau_{t,0}}\big\|_2 = \bigg\|\int_{\tau_{t,0}}^{\tau_{t,i}} s^\star_\tau(x^\star_\tau)\mathrm{d}\tau\bigg\|_2 \leq (\tau_{t,0}-\tau_{t,i})\sup_{\tau_{t,i}\leq\tau\leq\tau_{t,0}}\|s^\star_\tau(x^\star_\tau)\|_2$$
$$\lesssim (\tau_{t,0}-\tau_{t,K-1})\sqrt{\frac{d\log T}{\tau_{t,K-1}}} = (\overline{\alpha}_{t-1}-\overline{\alpha}_t)\sqrt{\frac{d\log T}{1-\overline{\alpha}_{t-1}}} = \overline{\alpha}_{t-1}\sqrt{1-\alpha_t}\sqrt{\frac{1-\alpha_t}{1-\overline{\alpha}_{t-1}}}\sqrt{d\log T}$$
$$\overset{\text{Lemma 1}}{\leq} 2c_1\frac{\log T}{T}\sqrt{d\log T} \tag{68}$$

for all $0\leq i\leq k-1$, provided that $-\log p_\tau(x^\star_\tau)\lesssim d\log T$ for all $\tau_{t,k-1}\leq\tau\leq\tau_{t,0}$. As a consequence, as long as $N\geq C_0\log T$ for some large enough constant $C_0>0$, one has

$$\big\|x^{(N)}_{\tau'} - x^\star_{\tau'}\big\|_2^2 \leq C\tau'^2\frac{\log^2 T}{T^2}\sum_{0\leq n\leq N}\sum_{0\leq i<K}\big(\varepsilon^{(n)}_{\mathsf{score},t,i}(x^{(n)}_{\tau_{t,i}})\big)^2 + 3C\frac{d\tau'\log^3 T}{T^2}\bigg(\frac{d\log^2 T}{T}\bigg)^{2K}. \tag{69}$$

Similarly, (60), (62a), (66), Lemma 1 and Lemma 3 together imply that on $\mathcal{E}_t$,

$$\bigg\|\frac{\partial x^{(n+1)}_{\tau'}/\sqrt{1-\tau'}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}} - \frac{\partial x^\star_{\tau'}/\sqrt{1-\tau'}}{\partial x^\star_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\bigg\|$$
$$\lesssim \sum_{0\leq i<K}(\tau_{t,0}-\tau')(1-\tau_{t,i})^{-3/2}\bigg\|\frac{\partial s^\star_{\tau_{t,i}}(x^\star_{\tau_{t,i}})}{\partial x^\star_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}\frac{\partial x^\star_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}{\partial x^\star_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}} - \frac{\partial s^\star_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}})}{\partial x^{(n)}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}\frac{\partial x^{(n)}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\bigg\|$$
$$+ \|\xi^{(n)}_{\mathsf{Jacob},\tau'}\| + O\bigg(\bigg(\frac{d(\tau_{t,0}-\tau')\log T}{\tau'(1-\tau')}\bigg)^{K+1}\bigg)$$
$$\lesssim \sum_{0\leq i<K}(\tau_{t,0}-\tau')(1-\tau_{t,i})^{-3/2}\bigg\|\frac{\partial s^\star_{\tau_{t,i}}(x^\star_{\tau_{t,i}})}{\partial x^\star_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}\bigg\|\bigg\|\frac{\partial x^\star_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}{\partial x^\star_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}} - \frac{\partial x^{(n)}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\bigg\|$$
$$+ \sum_{0\leq i<K}(\tau_{t,0}-\tau')(1-\tau_{t,i})^{-3/2}\bigg\|\frac{\partial x^{(n)}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\bigg\|\bigg\|\frac{\partial s^\star_{\tau_{t,i}}(x^\star_{\tau_{t,i}})}{\partial x^\star_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}} - \frac{\partial s^\star_{\tau_{t,i}}(x^{(n)}_{\tau_{t,i}})}{\partial x^{(n)}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}\bigg\|$$
$$+ \|\xi^{(n)}_{\mathsf{Jacob},\tau'}\| + O\bigg(\bigg(\frac{d(\tau_{t,0}-\tau')\log T}{\tau'(1-\tau')}\bigg)^{K+1}\bigg)$$
$$\lesssim \sum_{0\leq i<K}(\tau_{t,0}-\tau')(1-\tau_{t,i})^{-3/2}\sqrt{1-\tau_{t,i}}\frac{d\log T}{\tau_{t,i}}\bigg\|\frac{\partial x^\star_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}{\partial x^\star_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}} - \frac{\partial x^{(n)}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\bigg\|$$
$$+ \sum_{0\leq i<K}(\tau_{t,0}-\tau')(1-\tau_{t,i})^{-3/2}\sqrt{1-\tau_{t,i}}\sqrt{\frac{d^3\log^3 T}{\tau_{t,i}^3}}\big\|x^{(n)}_{\tau_{t,i}} - x^\star_{\tau_{t,i}}\big\|_2$$
$$+ \|\xi^{(n)}_{\mathsf{Jacob},\tau'}\| + O\bigg(\bigg(\frac{d(\tau_{t,0}-\tau')\log T}{\tau'(1-\tau')}\bigg)^{K+1}\bigg)$$

$$\lesssim \frac{d\log^2 T}{T}\sum_{0\le i<K}\bigg\{\bigg\|\frac{\partial x^{(n)}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}-\frac{\partial x^{\star}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}{\partial x^{\star}_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\bigg\|+\sqrt{\frac{d\log T}{\tau_{t,i}}}\big\|x^{(n)}_{\tau_{t,i}}-x^{\star}_{\tau_{t,i}}\big\|_2\bigg\}$$
$$+\|\xi^{(n)}_{\mathsf{Jacob},\tau'}\|+O\bigg(\Big(\frac{d\log^2 T}{T}\Big)^{K+1}\bigg),$$

where the last inequality holds since $\frac{\tau_{t,0}-\tau_{t,i}}{\tau_{t,i}(1-\tau_{t,i})}\lesssim\frac{\log T}{T}$ and $\tau_{t,0}-\tau'\asymp\tau_{t,0}-\tau_{t,i}$ for all $1\le i<K$.

The previous inequality combined with (63), (67) and (68) yields that on the event $\mathcal{E}_t$,

$$\bigg\|\frac{\partial x^{(n+1)}_{\tau'}/\sqrt{1-\tau'}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}-\frac{\partial x^{\star}_{\tau'}/\sqrt{1-\tau'}}{\partial x^{\star}_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\bigg\|^2$$
$$\le C\frac{d^2\log^4 T}{T^2}\sum_{0\le i<K}\bigg\|\frac{\partial x^{(n)}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}-\frac{\partial x^{\star}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}{\partial x^{\star}_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\bigg\|^2$$
$$+C\frac{d^3\log^5 T}{T^2\tau'}\bigg(\frac{1}{3^{n+1}}\max_{0\le i\le K}\big\|x^{(0)}_{\tau_{t,i}}-x^{\star}_{\tau_{t,i}}\big\|_2^2+C\tau'^2\frac{\log^2 T}{T^2}\sum_{0\le j\le n+1}\sum_{0\le i<K}\big(\varepsilon^{(j)}_{\mathsf{score},t,i}(x^{(j)}_{\tau_{t,i}})\big)^2$$
$$+2C\frac{d\tau'\log^3 T}{T^2}\Big(\frac{d\log^2 T}{T}\Big)^{2K}\bigg)$$
$$+C\frac{\log^2 T}{T^2}\tau'^2\sum_{0\le i<K}\big(\varepsilon^{(n)}_{\mathsf{Jacobi},t,i}(x^{(n)}_{\tau_{t,i}})\big)^2+C\Big(\frac{d\log^2 T}{T}\Big)^{2K+2}$$
$$\le C\frac{d^2\log^4 T}{T^2}\sum_{0\le i<K}\bigg\|\frac{\partial x^{(n)}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}-\frac{\partial x^{\star}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}{\partial x^{\star}_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\bigg\|^2$$
$$+C\frac{d^3\log^5 T}{T^2\tau'}\bigg(\frac{1}{3^{n+1}}4c_1^2\frac{d\log^3 T}{T^2}+C\tau'^2\frac{\log^2 T}{T^2}\sum_{0\le j\le n+1}\sum_{0\le i<K}\big(\varepsilon^{(j)}_{\mathsf{score},t,i}(x^{(j)}_{\tau_{t,i}})\big)^2$$
$$+2C\frac{d\tau'\log^3 T}{T^2}\Big(\frac{d\log^2 T}{T}\Big)^{2K}\bigg)$$
$$+C\frac{\log^2 T}{T^2}\tau'^2\sum_{0\le i<K}\big(\varepsilon^{(n)}_{\mathsf{Jacobi},t,i}(x^{(n)}_{\tau_{t,i}})\big)^2+C\Big(\frac{d\log^2 T}{T}\Big)^{2K+2}. \tag{70}$$

Repeating similar arguments as in (69) yields

$$\bigg\|\frac{\partial x^{(N)}_{\tau'}/\sqrt{1-\tau'}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}-\frac{\partial x^{\star}_{\tau'}/\sqrt{1-\tau'}}{\partial x^{\star}_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\bigg\|^2$$
$$\lesssim\frac{d^3\log^7 T}{T^4}\sum_{i,n}\big(\varepsilon^{(n)}_{\mathsf{score},t,i}(x^{(n)}_{\tau_{t,i}})\big)^2+\frac{\log^2 T}{T^2}\tau'^2\sum_{i,n}\big(\varepsilon^{(n)}_{\mathsf{Jacobi},t,i}(x^{(n)}_{\tau_{t,i}})\big)^2+\Big(\frac{d\log^2 T}{T}\Big)^{2K+2}. \tag{71}$$

In addition, recognizing that (62a) implies

$$\bigg\|\frac{\partial x^{(N)}_{\tau_{t,K-1}}/\sqrt{1-\tau_{t,K-1}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\bigg\|^{-1}\lesssim 1,$$

we have finished the proof of Lemma 7.

Proofs of (62a) and (62b). Assuming that the event $\mathcal{E}_t$ occurs, we will prove (62a) and (62b) by induction. For the base case $n=0$, since $x^{(0)}_{\tau'}=x^{(0)}_{\tau_{t,i}}=x_{\tau_{t,0}}$, we have

$$1\lesssim\sqrt{\frac{1-\overline{\alpha}_t}{1-\overline{\alpha}_{t+1}}}\le\bigg\|\frac{\partial x^{(0)}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\bigg\|=\bigg\|\frac{\sqrt{1-\tau_{t,0}}}{\sqrt{1-\tau_{t,i}}}I\bigg\|=\frac{\sqrt{1-\tau_{t,0}}}{\sqrt{1-\tau_{t,i}}}\le 1.$$

Moreover, it can be shown that on the event $\mathcal{E}_t$,

$$\begin{aligned}
\|\xi^{(0)}_{\mathsf{Jacob},\tau'}\| &= \Big\| \sum_{0\le i<K} \gamma_{t,i}(\tau')(1-\tau_{t,i})^{-3/2}\Big(\frac{\partial s_{\tau_{t,i}}(x^{(0)}_{\tau_{t,i}})}{\partial x^{(0)}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}} - \frac{\partial s^\star_{\tau_{t,i}}(x^{(0)}_{\tau_{t,i}})}{\partial x^{(0)}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}\Big)\frac{\partial x^{(0)}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\Big\| \\
&\lesssim \sum_{0\le i<K} 2^K(\tau_{t,0}-\tau')(1-\tau_{t,i})^{-1}\Big\|\frac{\partial s_{\tau_{t,i}}(x^{(0)}_{\tau_{t,i}})}{\partial x^{(0)}_{\tau_{t,i}}} - \frac{\partial s^\star_{\tau_{t,i}}(x^{(0)}_{\tau_{t,i}})}{\partial x^{(0)}_{\tau_{t,i}}}\Big\| \\
&\overset{(23f)}{\ll} K\cdot 2^K\cdot \tau_{t,i}\frac{\log T}{T}\cdot\frac{T}{\log T} \\
&\lesssim 1, \qquad (72)
\end{aligned}$$

Therefore, (62a) and (62b) hold for $n=0$. Supposing that (62a) and (62b) hold for n, we would like to show that these two claims still hold for $n+1$. In view of (59), one has

$$\Big\|\frac{\partial x^{(n+1)}_{\tau'}/\sqrt{1-\tau'}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\Big\| \le 1 + 2^K \sum_{0\le i<K}(\tau_{t,0}-\tau')\frac{d\log T}{\tau_{t,i}(1-\tau_{t,i})}\Big\|\frac{\partial x^{(n)}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\Big\| + \|\xi^{(n)}_{\mathsf{Jacob},\tau'}\| \lesssim 1,$$

provided that $-p_{\overline{X}_{\tau_{t,i}}}(x^{(n)}_{\tau_{t,i}}) \lesssim d\log T$. Here, the first inequality makes use of Lemma 3 and the second inequality holds due to the induction hypotheses (62a) and (62b) for n. Similarly, one has

$$\Big\|\frac{\partial x^{(n+1)}_{\tau'}/\sqrt{1-\tau'}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\Big\| \ge 1 - 2^K \sum_{0\le i<K}(\tau_{t,0}-\tau')\frac{d\log T}{\tau_{t,i}(1-\tau_{t,i})}\Big\|\frac{\partial x^{(n)}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\Big\| - \|\xi^{(n)}_{\mathsf{Jacob},\tau'}\| \gtrsim 1$$

In addition, repeating a similar argument as in (72) yields

$$\|\xi^{(n+1)}_{\mathsf{Jacob},\tau'}\| \le \frac{1}{2}.$$

Therefore, (62a) and (62b) hold for $n+1$, and consequently we have finished the proof.

Proof of Claims (64a) and (64b). It remains to verify (64a) and (64b). We let

$$u_k := \frac{\partial^k}{\partial\tau^k}\frac{s^\star_\tau(x^\star_\tau)}{(1-\tau)^{3/2}}.$$

The score function can be expressed as

$$\begin{aligned}
\frac{s^\star_\tau(\sqrt{1-\tau}x)}{(1-\tau)^{3/2}} &= -\frac{1}{\tau(1-\tau)}\int_{x_0} p_{X_0\,|\,\overline{X}_\tau}(x_0\,|\,\sqrt{1-\tau}x)(x-x_0)\mathrm{d}x_0 \\
&= -\frac{1}{\tau(1-\tau)}\cdot\frac{\int_{x_0} p_{X_0}(x_0)p_{\overline{X}_\tau\,|\,X_0}(\sqrt{1-\tau}x\,|\,x_0)(x-x_0)\mathrm{d}x_0}{\int_{x_0} p_{X_0}(x_0)p_{\overline{X}_\tau\,|\,X_0}(\sqrt{1-\tau}x\,|\,x_0)\mathrm{d}x_0}, \qquad (73)
\end{aligned}$$

where

$$(2\pi\tau)^{d/2}p_{\overline{X}_\tau\,|\,X_0}(\sqrt{1-\tau}x\,|\,x_0) = \exp\Big(-\frac{(1-\tau)\|x-x_0\|_2^2}{2\tau}\Big).$$

Then for sufficiently small δ, we have

$$\begin{aligned}
&\frac{s^\star_{\tau+\delta}(x^\star_{\tau+\delta})}{(1-\tau-\delta)^{3/2}} \\
&\quad= -\frac{1}{(\tau+\delta)(1-\tau-\delta)}\int_{x_0} p_{X_0\,|\,\overline{X}_{\tau+\delta}}(x_0\,|\,x^\star_{\tau+\delta})\Big(\frac{x^\star_{\tau+\delta}}{\sqrt{1-\tau-\delta}} - x_0\Big)\mathrm{d}x_0
\end{aligned}$$

$$= -\frac{1}{(\tau+\delta)(1-\tau-\delta)}\bigg(\frac{x^\star_{\tau+\delta}}{\sqrt{1-\tau-\delta}} - \frac{x^\star_\tau}{\sqrt{1-\tau}} + \int_{x_0} p_{X_0 \,|\, \overline{X}_{\tau+\delta}}(x_0 \,|\, x^\star_{\tau+\delta})\bigg(\frac{x^\star_\tau}{\sqrt{1-\tau}} - x_0\bigg)\mathrm{d}x_0\bigg)$$

$$= -\frac{1}{(\tau+\delta)(1-\tau-\delta)}\left(\frac{x^\star_{\tau+\delta}}{\sqrt{1-\tau-\delta}} - \frac{x^\star_\tau}{\sqrt{1-\tau}} + \frac{\int_{x_0} p_{X_0}(x_0) p_{\overline{X}_{\tau+\delta} \,|\, X_0}(x^\star_{\tau+\delta} \,|\, x_0)\left(\frac{x^\star_\tau}{\sqrt{1-\tau}} - x_0\right)\mathrm{d}x_0}{\int_{x_0} p_{X_0}(x_0) p_{\overline{X}_{\tau+\delta} \,|\, X_0}(x^\star_{\tau+\delta} \,|\, x_0)\mathrm{d}x_0}\right)$$

$$= -\frac{1}{(\tau+\delta)(1-\tau-\delta)}\left(\frac{x^\star_{\tau+\delta}}{\sqrt{1-\tau-\delta}} - \frac{x^\star_\tau}{\sqrt{1-\tau}} + \frac{\int_{x_0} p_{X_0}(x_0) p_{\overline{X}_{\tau} \,|\, X_0}(x^\star_{\tau} \,|\, x_0)\exp(\Delta)\left(\frac{x^\star_\tau}{\sqrt{1-\tau}} - x_0\right)\mathrm{d}x_0}{\int_{x_0} p_{X_0}(x_0) p_{\overline{X}_{\tau} \,|\, X_0}(x^\star_{\tau} \,|\, x_0)\exp(\Delta)\mathrm{d}x_0}\right)$$

$$= -\frac{1}{(\tau+\delta)(1-\tau-\delta)}\left(\frac{x^\star_{\tau+\delta}}{\sqrt{1-\tau-\delta}} - \frac{x^\star_\tau}{\sqrt{1-\tau}} + \frac{\int_{x_0} p_{X_0 \,|\, \overline{X}_{\tau}}(x_0 \,|\, x^\star_{\tau})\exp(\Delta)\big(\frac{x^\star_\tau}{\sqrt{1-\tau}} - x_0\big)\mathrm{d}x_0}{\int_{x_0} p_{X_0 \,|\, \overline{X}_{\tau}}(x_0 \,|\, x^\star_{\tau})\exp(\Delta)\mathrm{d}x_0}\right), \tag{74}$$

where

$$\Delta := \frac{(1-\tau)\|x^\star_\tau/\sqrt{1-\tau} - x_0\|_2^2}{\tau} - \frac{(1-\tau-\delta)\|x^\star_{\tau+\delta}/\sqrt{1-\tau-\delta} - x_0\|_2^2}{\tau+\delta} =: \sum_{k=1}^{\infty}\frac{\delta^k}{k!}v_k.$$

Here, the last equation of (74) makes use of the Bayes rule. We prove by induction that for any quantity $\overline{C} \geq 2$,

$$\|u_k\|_2 \leq c_k\sqrt{\frac{d\theta\log T}{\tau(1-\tau)^3}}\bigg(\frac{d\theta\log T}{\tau(1-\tau)}\bigg)^k, \quad \text{for } k = 0,1,\dots,K, \tag{75a}$$

$$|v_k| \leq C_k\overline{C}^2\bigg(\frac{d\theta\log T}{\tau(1-\tau)}\bigg)^k, \quad \text{for } k = 1,\dots,K, \quad \text{provided that } \bigg\|\frac{x^\star_\tau}{\sqrt{1-\tau}} - x_0\bigg\|_2 \leq 5\overline{C}\sqrt{\frac{d\theta\tau\log T}{1-\tau}}. \tag{75b}$$

For $k=0$, we have

$$\begin{aligned}
\|u_0\|_2 &= \bigg\|\frac{s^\star_\tau(x^\star_\tau)}{(1-\tau)^{3/2}}\bigg\|_2 \\
&= \frac{1}{\tau(1-\tau)}\bigg\|\int_{x_0} p_{X_0 \,|\, \overline{X}_\tau}(x_0 \,|\, x^\star_\tau)\bigg(\frac{x^\star_\tau}{\sqrt{1-\tau}} - x_0\bigg)\mathrm{d}x_0\bigg\|_2 \\
&= \frac{1}{\tau(1-\tau)}\int_{x_0} p_{X_0 \,|\, \overline{X}_\tau}(x_0 \,|\, x^\star_\tau)\bigg\|\frac{x^\star_\tau}{\sqrt{1-\tau}} - x_0\bigg\|\mathrm{d}x_0 \\
&= \frac{1}{\tau(1-\tau)}\mathbb{E}\bigg[\bigg\|\frac{x^\star_\tau}{\sqrt{1-\tau}} - X_0\bigg\|_2 \,\bigg|\, \overline{X}_\tau = x^\star_\tau\bigg] \\
&\overset{\text{Lemma 2}}{\lesssim} \frac{1}{\tau(1-\tau)}\frac{1}{\sqrt{1-\tau}}\sqrt{d\theta\tau\log T} \\
&\asymp \sqrt{\frac{d\theta\log T}{\tau(1-\tau)^3}}.
\end{aligned}$$

Suppose that (75a) and (75b) hold for $k \leq k_0$. We would like to prove (75a) and (75b) for $k = k_0 + 1$. Recalling that

$$\frac{\partial}{\partial\tau}\frac{x^\star_\tau}{\sqrt{1-\tau}} = -\frac{1}{2(1-\tau)^{3/2}}s^\star_\tau(x^\star_\tau),$$

we have

$$\frac{\partial^k}{\partial\tau^k}\frac{x^\star_\tau}{\sqrt{1-\tau}} = -\frac{\partial^{k-1}}{\partial\tau^{k-1}}\frac{1}{2(1-\tau)^{3/2}}s^\star_\tau(x^\star_\tau).$$

Then we know from Taylor expansion that

$$\frac{x^\star_{\tau+\delta}}{\sqrt{1-\tau-\delta}} - \frac{x^\star_\tau}{\sqrt{1-\tau}} = -\frac{1}{2}\sum_{k=1}^{\infty}\frac{\delta^k}{k!}u_{k-1}.$$

Furthermore, it is straightforward to verify that

$$\frac{1}{(\tau+\delta)(1-\tau-\delta)} - \frac{1}{\tau(1-\tau)} = \sum_{k=1}^{\infty} \delta^k \big[(1-\tau)^{-k-1} - (-\tau)^{-k-1}\big], \tag{76}$$

and

$$\begin{aligned}
&\frac{(1-\tau-\delta)\|x_{\tau+\delta}^{\star}/\sqrt{1-\tau-\delta} - x_0\|_2^2}{\tau+\delta} \\
&\quad = \Big(\frac{1-\tau}{\tau} - \sum_{k=1}^{\infty} \delta^k (-\tau)^{-k-1}\Big) \Big\| \frac{x_\tau^\star}{\sqrt{1-\tau}} - x_0 + \frac{x_{\tau+\delta}^\star}{\sqrt{1-\tau-\delta}} - \frac{x_\tau^\star}{\sqrt{1-\tau}} \Big\|_2^2 \\
&\quad = \Big(\frac{1-\tau}{\tau} - \sum_{k=1}^{\infty} \delta^k (-\tau)^{-k-1}\Big)\Big(\Big\| \frac{x_\tau^\star}{\sqrt{1-\tau}} - x_0 \Big\|_2^2 - \Big(\frac{x_\tau^\star}{\sqrt{1-\tau}} - x_0\Big)^\top \sum_{k=1}^{\infty} \frac{\delta^k}{k!} u_{k-1} + \frac{1}{4} \Big\| \sum_{k=1}^{\infty} \frac{\delta^k}{k!} u_{k-1} \Big\|_2^2 \Big).
\end{aligned} \tag{77}$$

Then we immediately have

$$\begin{aligned}
\sum_{k=1}^{\infty} \frac{\delta^k}{k!} v_k = \sum_{k=1}^{\infty} \Big[&- (-\tau)^{-k-1} \Big\| \frac{x_\tau^\star}{\sqrt{1-\tau}} - x_0 \Big\|_2^2 - \frac{1}{k!} \frac{1-\tau}{\tau} \Big(\frac{x_\tau^\star}{\sqrt{1-\tau}} - x_0\Big)^\top u_{k-1} \\
&+ \Big(\frac{x_\tau^\star}{\sqrt{1-\tau}} - x_0\Big)^\top \sum_{\ell=1}^{k-1} (-\tau)^{-\ell-1} \frac{1}{(k-\ell)!} u_{k-\ell-1} + \frac{1}{4} \frac{1-\tau}{\tau} \sum_{\ell=1}^{k-1} \frac{1}{\ell!(k-\ell)!} u_{\ell-1} u_{k-\ell-1} \\
&- \sum_{\ell=1}^{k-2} \sum_{j=1}^{k-1-\ell} \frac{1}{j!(k-\ell-j)!} (-\tau)^{-\ell-1} u_{j-1} u_{k-\ell-j-1} \Big] \delta^k
\end{aligned}$$

for sufficiently small $\delta > 0$. Comparing the coefficients of δ^{k_0+1} on both sides and making use of the induction assumption, one has

$$\begin{aligned}
|v_{k_0+1}| \le \;& \tau^{-k_0} \Big\| \frac{x_\tau^\star}{\sqrt{1-\tau}} - x_0 \Big\|_2^2 + \frac{1}{(k_0+1)!} \frac{1-\tau}{\tau} \Big\| \frac{x_\tau^\star}{\sqrt{1-\tau}} - x_0 \Big\|_2 \|u_{k_0}\|_2 \\
&+ \Big\| \frac{x_\tau^\star}{\sqrt{1-\tau}} - x_0 \Big\|_2 \sum_{\ell=1}^{k_0} \tau^{-\ell-1} \frac{1}{(k_0+1-\ell)!} u_{k_0-\ell} \\
&+ \frac{1}{4} \sum_{\ell=1}^{k_0} \frac{1}{\ell!(k_0+1-\ell)!} u_{\ell-1} u_{k_0-\ell} + \sum_{\ell=1}^{k_0-1} \sum_{j=1}^{k_0-\ell} \frac{1}{j!(k_0+1-\ell-j)!} \tau^{-\ell-1} u_{j-1} u_{k_0-\ell-j} \\
\lesssim \;& \tau^{k_0} \overline{C}^2 \frac{d\theta\tau\log T}{1-\tau} + \frac{1-\tau}{\tau} \overline{C} \sqrt{\frac{d\theta\tau\log T}{1-\tau}} \sqrt{\frac{d\theta\log T}{\tau(1-\tau)^3}} \Big(\frac{d\theta\log T}{\tau(1-\tau)}\Big)^{k_0} \\
&+ \overline{C} \sqrt{\frac{d\theta\tau\log T}{1-\tau}} \sum_{\ell=1}^{k_0} \tau^{-\ell-1} \sqrt{\frac{d\theta\log T}{\tau(1-\tau)^3}} \Big(\frac{d\theta\log T}{\tau(1-\tau)}\Big)^{k_0-\ell} \\
&+ \frac{1-\tau}{\tau} \sum_{\ell=1}^{k_0} \sqrt{\frac{d\theta\log T}{\tau(1-\tau)^3}} \Big(\frac{d\theta\log T}{\tau(1-\tau)}\Big)^{\ell-1} \sqrt{\frac{d\theta\log T}{\tau(1-\tau)^3}} \Big(\frac{d\theta\log T}{\tau(1-\tau)}\Big)^{k_0-\ell} \\
&+ \sum_{\ell=1}^{k_0-1} \sum_{j=1}^{k_0-\ell} \tau^{-\ell-1} \sqrt{\frac{d\theta\log T}{\tau(1-\tau)^3}} \Big(\frac{d\theta\log T}{\tau(1-\tau)}\Big)^{j-1} \sqrt{\frac{d\theta\log T}{\tau(1-\tau)^3}} \Big(\frac{d\theta\log T}{\tau(1-\tau)}\Big)^{k_0-\ell-j} \\
\lesssim \;& \overline{C}^2 \Big(\frac{d\theta\log T}{\tau(1-\tau)}\Big)^{k_0+1},
\end{aligned}$$

provided that $\|\frac{x_\tau^\star}{\sqrt{1-\tau}} - x_0\|_2 \le 5\overline{C}\sqrt{\frac{d\theta\tau\log T}{1-\tau}}$. Therefore, we have verified (75b) for $k = k_0+1$.

Then, we prove that (75a) holds for $k = k_0 + 1$. We let $e^{\Delta} =: \sum_{k=0}^{\infty} \frac{\delta^k}{k!} w_k = \sum_{k=0}^{\infty} \frac{\delta^k}{k!} w_k(x_0)$ denote the Taylor expansion of e^{Δ}. Then one can show that $w_0 = 1$ and for all $1 \le k \le k_0 + 1$,

$$|w_k| = \left| \sum_{\ell=1}^{k} \sum_{j_1+\cdots+j_\ell=k} \frac{k!}{j_1!\cdots j_\ell!} v_{j_1} \cdots v_{j_\ell} \right| \lesssim \overline{C}^{2k} \left(\frac{d\theta \log T}{\tau(1-\tau)} \right)^k, \tag{78}$$

provided that $\|\frac{x_\tau^\star}{\sqrt{1-\tau}} - x_0\|_2 \le 5\overline{C}\sqrt{\frac{d\theta\tau\log T}{1-\tau}}$.

We define

$$\mathcal{C}_0 = \left\{ x_0 \in \mathbb{R}^d : \Big\| \frac{x_\tau^\star}{\sqrt{1-\tau}} - x_0 \Big\|_2 \le 5c_7 \sqrt{\frac{d\theta\tau\log T}{1-\tau}} \right\},$$
$$\mathcal{C}_\ell = \left\{ x_0 \in \mathbb{R}^d : 5\cdot 2^{\ell-1} c_7 \sqrt{\frac{d\tau\log T}{1-\tau}} < \Big\| \frac{x_\tau^\star}{\sqrt{1-\tau}} - x_0 \Big\|_2 \le 5\cdot 2^{\ell} c_7 \sqrt{\frac{d\theta\tau\log T}{1-\tau}} \right\}, \quad k \ge 1.$$

where c_7 is a sufficiently large constant. Let a_k (resp. b_k) denote the k-th order derivative $\frac{\partial^k}{\partial\tau^k}\int_{x_0} p_{X_0\,|\,\overline{X}_\tau}(x_0\,|\,x_\tau^\star)\exp(\Delta)\big(\frac{x_\tau^\star}{\sqrt{1-\tau}} - x_0\big)\mathrm{d}x_0$ (resp. $\frac{\partial^k}{\partial\tau^k}\int_{x_0} p_{X_0\,|\,\overline{X}_\tau}(x_0\,|\,x_\tau^\star)\exp(\Delta)\mathrm{d}x_0$). Then for $0 \le k \le k_0 + 1$,

$$\begin{aligned}
\|a_k\|_2 &= \Big\| \int_{x_0} p_{X_0\,|\,\overline{X}_\tau}(x_0\,|\,x_\tau^\star)\Big(\frac{x_\tau^\star}{\sqrt{1-\tau}} - x_0\Big) w_k \mathrm{d}x_0 \Big\|_2 \\
&= \Big\| \sum_{\ell=0}^{\infty} \int_{x_0\in\mathcal{C}_\ell} p_{X_0\,|\,\overline{X}_\tau}(x_0\,|\,x_\tau^\star)\Big(\frac{x_\tau^\star}{\sqrt{1-\tau}} - x_0\Big) w_k \mathrm{d}x_0 \Big\|_2 \\
&\le \Big\| \int_{x_0\in\mathcal{C}_0} p_{X_0\,|\,\overline{X}_\tau}(x_0\,|\,x_\tau^\star)\Big(\frac{x_\tau^\star}{\sqrt{1-\tau}} - x_0\Big) w_k \mathrm{d}x_0 \Big\|_2 \\
&\quad + \sum_{\ell=1}^{\infty} \Big\| \int_{x_0\in\mathcal{C}_\ell} p_{X_0\,|\,\overline{X}_\tau}(x_0\,|\,x_\tau^\star)\Big(\frac{x_\tau^\star}{\sqrt{1-\tau}} - x_0\Big) w_k \mathrm{d}x_0 \Big\|_2 \\
&\le \sup_{x_0\in\mathcal{C}_0} \Big\| \frac{x_\tau^\star}{\sqrt{1-\tau}} - x_0 \Big\|_2 \|w_k\|_2 \\
&\quad + \sum_{\ell=1}^{\infty} \mathbb{P}\left(\Big\| \frac{x_\tau^\star}{\sqrt{1-\tau}} - X_0 \Big\|_2 \le 5\cdot 2^\ell c_7 \sqrt{\frac{d\theta\tau\log T}{1-\tau}} \;\Big|\; \overline{X}_\tau = x_\tau^\star \right) \sup_{x_0\in\mathcal{C}_\ell} \Big\| \frac{x_\tau^\star}{\sqrt{1-\tau}} - x_0 \Big\|_2 \|w_k\|_2 \\
&\overset{(*)}{\lesssim} 5c_7 \sqrt{\frac{d\theta\tau\log T}{1-\tau}} \cdot c_7^{2k} \left(\frac{d\theta\log T}{\tau(1-\tau)} \right)^k \\
&\quad + \sum_{\ell=1}^{\infty} \exp\left(-2^{2\ell} c_7^2 d\theta\log T\right) \cdot 2^\ell c_7 \sqrt{\frac{d\theta\tau\log T}{1-\tau}} \cdot (5\cdot 2^\ell c_7)^{2k} \left(\frac{d\theta\log T}{\tau(1-\tau)} \right)^k \\
&\lesssim \sqrt{\frac{d\theta\tau\log T}{1-\tau}} \left(\frac{d\theta\log T}{\tau(1-\tau)} \right)^k + \sum_{\ell=1}^{\infty} \exp\left(-2^{\ell} c_7^2 d\theta\log T\right) \sqrt{\frac{d\theta\tau\log T}{1-\tau}} \left(\frac{d\theta\log T}{\tau(1-\tau)} \right)^k \\
&\asymp \sqrt{\frac{d\theta\tau\log T}{1-\tau}} \left(\frac{d\theta\log T}{\tau(1-\tau)} \right)^k,
\end{aligned}$$

and more specifically,

$$a_0 = \int_{x_0} p_{X_0\,|\,\overline{X}_\tau}(x_0\,|\,x_\tau^\star)\Big(\frac{x_\tau^\star}{\sqrt{1-\tau}} - x_0\Big) w_0 \mathrm{d}x_0 = \int_{x_0} p_{X_0\,|\,\overline{X}_\tau}(x_0\,|\,x_\tau^\star)\Big(\frac{x_\tau^\star}{\sqrt{1-\tau}} - x_0\Big) \mathrm{d}x_0.$$

Here, $(*)$ is valid due to Lemma 2 and (78). Similarly, for all $1 \le k \le k_0 + 1$, one can show that

$$b_0 = 1, \qquad \text{and} \qquad |b_k| \lesssim \left(\frac{d\theta\log T}{\tau(1-\tau)} \right)^k, \quad \forall 1 \le k \le k_0 + 1.$$

We denote by

$$d_k := \frac{\partial^k}{\partial \tau^k} \frac{\int_{x_0} p_{X_0 \mid \overline{X}_\tau}(x_0 \mid x_\tau^\star) \exp(\Delta)\big(\frac{x_\tau^\star}{\sqrt{1-\tau}} - x_0\big)\mathrm{d}x_0}{\int_{x_0} p_{X_0 \mid \overline{X}_\tau}(x_0 \mid x_\tau^\star) \exp(\Delta)\mathrm{d}x_0}$$

the k-th derivative of $\frac{\int_{x_0} p_{X_0 \mid \overline{X}_\tau}(x_0 \mid x_\tau^\star) \exp(\Delta)\big(\frac{x_\tau^\star}{\sqrt{1-\tau}} - x_0\big)\mathrm{d}x_0}{\int_{x_0} p_{X_0 \mid \overline{X}_\tau}(x_0 \mid x_\tau^\star) \exp(\Delta)\mathrm{d}x_0}$. Then one can easily verify that

$$d_0 = \int_{x_0} p_{X_0 \mid \overline{X}_\tau}(x_0 \mid x_\tau^\star)\Big(\frac{x_\tau^\star}{\sqrt{1-\tau}} - x_0\Big)\mathrm{d}x_0, \tag{79a}$$

$$|d_k| = |a_k - \sum_{\ell=0}^{k-1} d_\ell b_{k-\ell}| \lesssim \sqrt{\frac{d\theta\tau\log T}{1-\tau}}\Big(\frac{d\theta\log T}{\tau(1-\tau)}\Big)^k, \quad \text{for all } 0 \le k \le k_0+1. \tag{79b}$$

Now, we are ready to prove (75b) for $t = t_0 + 1$. In view of (74) and (73), one has

$$\begin{aligned}
\sum_{k=1}^{\infty} \frac{\delta^k}{k!} u_k &= \frac{s_{\tau+\delta}^\star(x_{\tau+\delta}^\star)}{(1-\tau-\delta)^{3/2}} - \frac{s_\tau^\star(x_\tau^\star)}{(1-\tau)^{3/2}} \\
&= -\frac{1}{(\tau+\delta)(1-\tau-\delta)}\left(\frac{x_{\tau+\delta}^\star}{\sqrt{1-\tau-\delta}} - \frac{x_\tau^\star}{\sqrt{1-\tau}} + \frac{\int_{x_0} p_{X_0 \mid \overline{X}_\tau}(x_0 \mid x_\tau^\star) \exp(\Delta)\big(\frac{x_\tau^\star}{\sqrt{1-\tau}} - x_0\big)\mathrm{d}x_0}{\int_{x_0} p_{X_0 \mid \overline{X}_\tau}(x_0 \mid x_\tau^\star) \exp(\Delta)\mathrm{d}x_0}\right) \\
&\quad + \frac{1}{\tau(1-\tau)}\int_{x_0} p_{X_0 \mid \overline{X}_\tau}(x_0 \mid \sqrt{1-\tau}x)(x - x_0)\mathrm{d}x_0 \\
&= -\left(\frac{1}{\tau(1-\tau)} + \sum_{k=1}^{\infty}\delta^k\big[(1-\tau)^{-k-1} - (-\tau)^{-k-1}\big]\right)\left(-\frac{1}{2}\sum_{k=1}^{\infty}\frac{\delta^k}{k!}u_{k-1} + \sum_{k=0}^{\infty}\frac{\delta^k}{k!}d_k\right) + \frac{1}{\tau(1-\tau)}d_0 \\
&= \sum_{k=1}^{\infty}\left(\frac{1}{\tau(1-\tau)}\frac{u_{k-1} - 2d_k}{2k!} + \sum_{\ell=1}^{k}\big[(-\tau)^{-\ell-1} - (1-\tau)^{-\ell-1}\big]\frac{2d_{k-\ell} - u_{k-1-\ell}}{2(k-\ell)!}\right)\delta^k.
\end{aligned} \tag{80}$$

By virtue of the induction hypothesis (75a) for $1 \le k \le k_0$ and (79b), we have

$$\begin{aligned}
|u_{k_0+1}| &= \left|\frac{1}{\tau(1-\tau)}\Big(\frac{u_{k_0}}{2} - d_{k_0+1}\Big) + \sum_{\ell=1}^{k_0+1}\big[(-\tau)^{-\ell-1} - (1-\tau)^{-\ell-1}\big]\frac{(k_0+1)!}{2(k_0+1-\ell)!}(2d_{k_0+1-\ell} - u_{k_0-\ell})\right| \\
&\lesssim \frac{1}{\tau(1-\tau)}\left(\sqrt{\frac{d\theta\log T}{\tau(1-\tau)^3}}\Big(\frac{d\theta\log T}{\tau(1-\tau)}\Big)^{k_0} + \sqrt{\frac{d\theta\tau\log T}{1-\tau}}\Big(\frac{d\theta\log T}{\tau(1-\tau)}\Big)^{k_0+1}\right) \\
&\quad + \sum_{\ell=1}^{k_0+1}\frac{1}{[\tau(1-\tau)]^{\ell+1}}\left[\sqrt{\frac{d\theta\tau\log T}{1-\tau}}\Big(\frac{d\theta\log T}{\tau(1-\tau)}\Big)^{k_0+1-\ell} + \sqrt{\frac{d\theta\log T}{\tau(1-\tau)^3}}\Big(\frac{d\theta\log T}{\tau(1-\tau)}\Big)^{k_0-\ell}\right] \\
&\asymp \sqrt{\frac{d\theta\log T}{\tau(1-\tau)^3}}\Big(\frac{d\theta\log T}{\tau(1-\tau)}\Big)^{k_0+1},
\end{aligned}$$

which has finished the proof of the induction step. Therefore, we have verified Claim (64a).

Claim (64b) can be proved by using similar arguments. We omit the details here for the sake of brevity.

B Proofs of technical lemmas

B.1 Proof of Lemma 1

Equations (23a) - (23e) follow directly from Li et al. (2024c, (26a) - (26e)).

Proof of (23f). For any $2 \le t \le T, 0 \le i_1, i_2, i_3, i_4 \le K-1$, one has

$$\Big|\frac{\tau_{t,i_1} - \tau_{t,i_2}}{\tau_{t,i_3}(1-\tau_{t,i_4})}\Big| \le \frac{\tau_{t,0} - \tau_{t,k-1}}{\tau_{t,k-1}(1-\tau_{t,0})} = \frac{\overline{\alpha}_{t-1} - \overline{\alpha}_t}{\overline{\alpha}_t(1-\overline{\alpha}_{t-1})} = \frac{1}{\alpha_t}\frac{1-\alpha_t}{1-\overline{\alpha}_{t-1}} \overset{(23a) \text{ and } (23b)}{\le} 2 \cdot \frac{4c_1\log T}{T} = \frac{8c_1\log T}{T}.$$

Proof of (23g). By virtue of (15), one has

$$|\gamma_{t,i}(\tau_{t,j})| \le (\tau_{t,0} - \tau_{t,j}) \sup_{\tau \in [\tau_{t,K-1}, \tau_{t,0}]} |\psi_i(\tau)| .$$

We observe that

$$\sup_{\tau \in [\tau_{t,K-1}, \tau_{t,0}]} |\psi_i(\tau)| = \sup_{\eta \in [0, K-1]} \left| \frac{\prod_{i' \neq i}(K-1-i'-\eta)}{\prod_{i' \neq i}(i-i')} \right| \le \frac{(K-1)!}{(\lfloor \frac{K-1}{2} \rfloor)!\,(\lfloor \frac{K-1}{2} \rfloor)!} = \binom{K-1}{\lfloor \frac{K-1}{2} \rfloor} \le 2^K.$$

Putting the previous two inequalities together, we have completed the proof of (23g).

Proof of (23h). For all $0 \le i \le K-1, 2 \le t \le T$, one has

$$\overline{\alpha}_t = 1 - \tau_{t,0} \le 1 - \tau_{t,i} \le 1 - \tau_{t,K-1} = \overline{\alpha}_{t-1} \qquad \text{and} \qquad 1 - \overline{\alpha}_{t-1} = \tau_{t,K-1} \le \tau_{t,i} \le \tau_{t,0} = 1 - \overline{\alpha}_t.$$

Therefore, it suffices to show that

$$\overline{\alpha}_t \asymp \overline{\alpha}_{t-1} \qquad \text{and} \qquad 1 - \overline{\alpha}_{t-1} \asymp 1 - \overline{\alpha}_t,$$

which are direct consequences of (23a) and (23c), respectively.

B.2 Proof of Lemma 3

Elementary calculations reveal that

$$\begin{aligned}
\|s_\tau^\star(x)\|_2 &= \frac{\sqrt{1-\tau}}{\tau} \left\| \int_{x_0} p_{X_0 \,|\, \overline{X}_\tau}(x_0 \,|\, x) \left(\frac{x}{\sqrt{1-\tau}} - x_0 \right) \mathrm{d}x_0 \right\|_2 \\
&= \frac{\sqrt{1-\tau}}{\tau} \int_{x_0} p_{X_0 \,|\, \overline{X}_\tau}(x_0 \,|\, x) \left\| \frac{x}{\sqrt{1-\tau}} - x_0 \right\| \mathrm{d}x_0 \\
&= \frac{\sqrt{1-\tau}}{\tau} \mathbb{E}\left[\left\| \frac{x}{\sqrt{1-\tau}} - X_0 \right\|_2 \;\middle|\; \overline{X}_\tau = x \right] \\
&\overset{\text{Lemma 2}}{\lesssim} \frac{\sqrt{1-\tau}}{\tau} \frac{1}{\sqrt{1-\tau}} \sqrt{d\theta_\tau(x)\tau \log T} \\
&\asymp \sqrt{\frac{d\theta_\tau(x) \log T}{\tau}},
\end{aligned}$$

which has finished the proof of (27a).

Turning to (27b), by virtue of (24) and (25b) in Li et al. (2024c), we have

$$\frac{\partial s_\tau^\star(x_\tau)}{\partial x} = -\frac{1}{\tau} I_d + \frac{1}{\tau^2} \mathsf{Cov}\left(\overline{X}_\tau - \sqrt{1-\tau} X_0 \mid \overline{X}_\tau = x \right). \tag{81}$$

We make the observation that

$$\begin{aligned}
\left\| \mathsf{Cov}\left(\overline{X}_\tau - \sqrt{1-\tau} X_0 \mid \overline{X}_\tau = x \right) \right\| &\le \left\| \mathbb{E}\left[\left(x - \sqrt{1-\tau} X_0 \right) \left(x - \sqrt{1-\tau} X_0 \right)^\top \mid \overline{X}_\tau = x \right] \right\| \\
&\le \mathbb{E}\left[\left\| x - \sqrt{1-\tau} X_0 \right\|_2^2 \mid \overline{X}_\tau = x \right] \\
&\lesssim d\theta_\tau(x) \tau \log T.
\end{aligned}$$

Here, the last line holds due to Lemma 2. Then we immediately have

$$\left\| \frac{\partial s_\tau^\star(x_\tau)}{\partial x} \right\| \lesssim \frac{1}{\tau} + \frac{1}{\tau^2} d\theta_\tau(x) \tau \log T \asymp \frac{d\theta_\tau(x) \log T}{\tau}.$$

Eqn. (28) follows by an argument similar to that of Li and Cai (2024, Claim (89)); we omit the details for brevity.

B.3 Proof of Lemma 4

We let J_τ denote the Jacobian matrix

$$J_\tau(\sqrt{1-\tau}x) := \frac{\partial}{\partial x}\frac{s_\tau^\star(\sqrt{1-\tau}x)}{(1-\tau)^{3/2}}$$
$$= -\frac{1}{\tau(1-\tau)}I_d + \frac{1}{\tau^2}\Big(\mathbb{E}\big[(x-X_0)(x-X_0)^\top \mid \overline{X}_\tau = \sqrt{1-\tau}x\big]$$
$$- \mathbb{E}\big[x-X_0 \mid \overline{X}_\tau = -\sqrt{1-\tau}x\big]\mathbb{E}\big[x-X_0 \mid \overline{X}_\tau = \sqrt{1-\tau}x\big]^\top\Big). \quad (82)$$

We define

$$\theta' = \sup_{\tau':|\tau'-\widetilde{\tau}|\le c_0\widetilde{\tau}(1-\widetilde{\tau})} \frac{-\log p_{\overline{X}_{\tau'}}(x_{\tau'}^\star)}{d\log T} < \infty.$$

By virtue of Lemma 2, we have

$$\mathsf{tr}\left(J_\tau(x_\tau^\star)\right) \le -\frac{d}{\tau(1-\tau)} + \frac{1}{\tau^2(1-\tau)}\mathsf{tr}\left(\mathbb{E}\big[(x_\tau^\star - \sqrt{1-\tau}X_0)(x_\tau^\star - \sqrt{1-\tau}X_0)^\top \mid \overline{X}_\tau = x_\tau^\star\big]\right)$$
$$= -\frac{d}{\tau(1-\tau)} + \frac{1}{\tau^2(1-\tau)}\mathbb{E}\left[\left\|x_\tau^\star - \sqrt{1-\tau}X_0\right\|_2^2 \mid \overline{X}_\tau = x_\tau^\star\right]$$
$$\le -\frac{d}{\tau(1-\tau)} + \frac{120}{\tau^2(1-\tau)}\theta' d\tau\log T \le \frac{120}{\tau(1-\tau)}\theta' d\log T. \quad (83)$$

In addition, (9)tells us that

$$\frac{\partial}{\partial\tau}\frac{\partial x_\tau^\star/\sqrt{1-\tau}}{\partial x_{\tau'}^\star/\sqrt{1-\tau'}} = \frac{\partial}{\partial x_{\tau'}^\star/\sqrt{1-\tau'}}\frac{\partial x_\tau^\star/\sqrt{1-\tau}}{\partial\tau}$$
$$= \frac{\partial}{\partial x_\tau^\star/\sqrt{1-\tau}}\left(-\frac{1}{2(1-\tau)^{3/2}}s_\tau^\star(x_\tau^\star)\right)$$
$$= \frac{\partial}{\partial x_{\tau'}^\star/\sqrt{1-\tau'}}\left(-\frac{1}{2(1-\tau)^{3/2}}s_\tau^\star(x_\tau^\star)\right)\frac{\partial x_\tau^\star/\sqrt{1-\tau}}{\partial x_{\tau'}^\star/\sqrt{1-\tau'}}$$
$$= -\frac{1}{2}J_\tau\left(x_\tau^\star\right)\frac{\partial x_\tau^\star/\sqrt{1-\tau}}{\partial x_{\tau'}^\star/\sqrt{1-\tau'}}. \quad (84)$$

Applying Jacobi's formula

$$\frac{\mathrm{d}}{\mathrm{d}t}\det A(t) = (\det A(t))\cdot\mathrm{tr}\left(A(t)^{-1}\cdot\frac{\mathrm{d}A(t)}{\mathrm{d}t}\right)$$

yields

$$\frac{\partial}{\partial\tau}\det\left(\frac{\partial x_\tau^\star/\sqrt{1-\tau}}{\partial x_{\tau'}^\star/\sqrt{1-\tau'}}\right) = \det\left(\frac{\partial x_\tau^\star/\sqrt{1-\tau}}{\partial x_{\tau'}^\star/\sqrt{1-\tau'}}\right)\cdot\mathrm{tr}\left(\left(\frac{\partial x_\tau^\star/\sqrt{1-\tau}}{\partial x_{\tau'}^\star/\sqrt{1-\tau'}}\right)^{-1}\cdot\frac{\partial}{\partial\tau}\frac{\partial x_\tau^\star/\sqrt{1-\tau}}{\partial x_{\tau'}^\star/\sqrt{1-\tau'}}\right)$$
$$= -\frac{1}{2}\mathrm{tr}\left(J_\tau\left(x_\tau^\star\right)\right)\det\left(\frac{\partial x_\tau^\star/\sqrt{1-\tau}}{\partial x_{\tau'}^\star/\sqrt{1-\tau'}}\right). \quad (85)$$

By solving this equation, we know that

$$\det\left(\frac{\partial x_{\tau''}^\star/\sqrt{1-\tau''}}{\partial x_{\tau'}^\star/\sqrt{1-\tau'}}\right) = \exp\left(-\frac{1}{2}\int_{\tau'}^{\tau''}\mathrm{tr}\left(J_\tau\left(x_\tau^\star\right)\right)\mathrm{d}\tau\right) \quad (86)$$

for all τ', τ''. In view of (83) and (85), one has

$$\frac{\mathrm{d}}{\mathrm{d}\tau'}\left(\log\frac{p_{\overline{X}_{\tau'}/\sqrt{1-\tau'}}\left(x_{\tau'}^\star/\sqrt{1-\tau'}\right)}{p_{\overline{X}_{\widetilde{\tau}}/\sqrt{1-\widetilde{\tau}}}\left(x_{\widetilde{\tau}}^\star/\sqrt{1-\widetilde{\tau}}\right)}\right) = -\frac{\mathrm{d}}{\mathrm{d}\tau'}\log\det\left(\frac{\partial x_{\tau'}^\star/\sqrt{1-\tau'}}{\partial x_{\widetilde{\tau}}^\star/\sqrt{1-\widetilde{\tau}}}\right) = \frac{1}{2}\mathrm{tr}\left(J_{\tau'}\left(x_{\tau'}^\star\right)\right). \quad (87)$$

Moreover, recognizing that for all $|\tau' - \widetilde{\tau}| \le c_0\widetilde{\tau}(1-\widetilde{\tau})$,

$$\left|\frac{1}{\tau'(1-\tau')} - \frac{1}{\widetilde{\tau}(1-\widetilde{\tau})}\right| = \frac{|(\tau'-\widetilde{\tau})(1-\tau'-\widetilde{\tau})|}{\tau'(1-\tau')\widetilde{\tau}(1-\widetilde{\tau})} \le \frac{c_0\widetilde{\tau}(1-\widetilde{\tau})}{\tau'(1-\tau')\widetilde{\tau}(1-\widetilde{\tau})} = \frac{c_0}{\tau'(1-\tau')},$$

we have

$$\frac{1}{\tau'(1-\tau')} \le \frac{1}{1-c_0}\frac{1}{\widetilde{\tau}(1-\widetilde{\tau})} \le \frac{2}{\widetilde{\tau}(1-\widetilde{\tau})}.$$

Combining (83), (87) and the previous inequality gives us

$$-\log p_{\overline{X}_{\tau'}/\sqrt{1-\tau'}}\big(x^\star_{\tau'}/\sqrt{1-\tau'}\big) \le -\log p_{\overline{X}_{\widetilde{\tau}}/\sqrt{1-\widetilde{\tau}}}\big(x^\star_{\widetilde{\tau}}/\sqrt{1-\widetilde{\tau}}\big) + |\tau'-\widetilde{\tau}|\frac{120}{\tau(1-\tau)}\theta' d\log T.$$

Furthermore, by noting that $p_{\overline{X}_{\tau}/\sqrt{1-\tau}}(x^\star_{\tau}/\sqrt{1-\tau}) = (1-\tau)^{d/2}p_{\overline{X}_{\tau}}(x^\star_{\tau})$, we have

$$\begin{aligned}
-\log p_{\overline{X}_{\tau'}}(x^\star_{\tau'}) &\le -\log p_{\overline{X}_{\widetilde{\tau}}}(x^\star_{\widetilde{\tau}}) + d\log\left(\frac{1-\tau'}{1-\widetilde{\tau}}\right) + |\tau'-\widetilde{\tau}|\frac{120}{\tau(1-\tau)}\theta' d\log T \\
&\le -\log p_{\overline{X}_{\widetilde{\tau}}}(x^\star_{\widetilde{\tau}}) + d\frac{|\tau'-\widetilde{\tau}|}{1-\widetilde{\tau}} + 60c_0\theta' d\log T \\
&\le \theta d\log T + c_0 d\tau + 60c_0\theta' d\log T \\
&\le \theta d\log T + \frac{1}{2}\theta' d\log T,
\end{aligned}$$

provided that $|\tau'-\widetilde{\tau}| \le c_0\widetilde{\tau}(1-\widetilde{\tau})$. Here, the second inequality makes use of $\log(1+x) \le x$. Recalling the definition of θ', we know that

$$\theta' d\log T = \sup_{\tau':|\tau'-\widetilde{\tau}|\le c_0\widetilde{\tau}(1-\widetilde{\tau})} -\log p_{\overline{X}_{\tau'}}(x^\star_{\tau'}) \le \theta d\log T + \frac{1}{2}\theta' d\log T,$$

which implies

$$\theta' \le 2\theta.$$

Therefore, for all $|\tau'-\widetilde{\tau}| \le c_0\widetilde{\tau}(1-\widetilde{\tau})$, we have

$$-\log p_{\overline{X}_{\tau'}}(x^\star_{\tau'}) \le 2\theta d\log T,$$

which has finished the proof of Lemma 4.

B.4 Proof of Lemma 6

In addition, elementary calculations show that

$$\frac{p_{\sqrt{\alpha_t}X_{t-1}}(\sqrt{\alpha_t}x_{t-1})}{p_{\sqrt{\alpha_t}X_{t-1}}(\sqrt{\alpha_t}x^\star_{t-1})} = \int_{x_0} p_{X_0\,|\,\sqrt{\alpha_t}X_{t-1}}(x_0|\sqrt{\alpha_t}x^\star_{t-1})\exp\Big(-\frac{\|x_{t-1}-\sqrt{\overline{\alpha}_{t-1}}x_0\|_2^2 - \|x^\star_{t-1}-\sqrt{\overline{\alpha}_{t-1}}x_0\|_2^2}{2(1-\overline{\alpha}_{t-1})}\Big)\mathrm{d}x_0.$$

We define $u := x_{t-1} - x^\star_{t-1}$. Then we have

$$\begin{aligned}
\frac{p_{\sqrt{\alpha_t}X_{t-1}}(\sqrt{\alpha_t}x_{t-1})}{p_{\sqrt{\alpha_t}X_{t-1}}(\sqrt{\alpha_t}x^\star_{t-1})} &= \int_{x_0} p_{X_0\,|\,\sqrt{\alpha_t}X_{t-1}}(x_0|\sqrt{\alpha_t}x^\star_{t-1})\exp\Big(-\frac{\|u\|_2^2 + 2u^\top(x^\star_{t-1}-\sqrt{\overline{\alpha}_{t-1}}x_0)}{2(1-\overline{\alpha}_{t-1})}\Big)\mathrm{d}x_0 \\
&= \exp\Big(-\frac{\|u\|_2^2}{2(1-\overline{\alpha}_{t-1})}\Big)\int_{x_0} p_{X_0\,|\,\sqrt{\alpha_t}X_{t-1}}(x_0|\sqrt{\alpha_t}x^\star_{t-1})\exp\Big(-\frac{u^\top(x^\star_{t-1}-\sqrt{\overline{\alpha}_{t-1}}x_0)}{1-\overline{\alpha}_{t-1}}\Big)\mathrm{d}x_0 \\
&\le \exp\Big(-\frac{\|u\|_2^2}{2(1-\overline{\alpha}_{t-1})}\Big)\mathbb{E}\left[\exp\Big(-\frac{u^\top(x^\star_{t-1}-\sqrt{\overline{\alpha}_{t-1}}X_0)}{1-\overline{\alpha}_{t-1}}\Big)\;\Big|\; X_{t-1} = x^\star_{t-1}\right]
\end{aligned}$$

$$\leq \exp\Big(-\frac{\|u\|_2^2}{2(1-\overline{\alpha}_{t-1})}\Big)\mathbb{E}\left[\exp\Big(\frac{\|u\|_2\|x_{t-1}^\star-\sqrt{\overline{\alpha}_{t-1}}X_0\|_2}{1-\overline{\alpha}_{t-1}}\Big)\;\Big|\;X_{t-1}=x_{t-1}^\star\right]. \tag{88}$$

By virtue of Lemma 2, one has

$$\begin{aligned}
&\mathbb{E}\left[\exp\Big(\frac{\|u\|_2\|x_{t-1}^\star-\sqrt{\overline{\alpha}_{t-1}}X_0\|_2}{1-\overline{\alpha}_{t-1}}\Big)\;\Big|\;X_{t-1}=x_{t-1}^\star\right]\\
&\quad=\int_0^\infty \mathbb{P}\left(\exp\Big(\frac{\|u\|_2\|x_{t-1}^\star-\sqrt{\overline{\alpha}_{t-1}}X_0\|_2}{1-\overline{\alpha}_{t-1}}\Big)>w\;\Big|\;X_{t-1}=x_{t-1}^\star\right)\mathrm{d}w\\
&\quad=1+\int_1^\infty \mathbb{P}\left(\exp\Big(\frac{\|u\|_2\|x_{t-1}^\star-\sqrt{\overline{\alpha}_{t-1}}X_0\|_2}{1-\overline{\alpha}_{t-1}}\Big)>w\;\Big|\;X_{t-1}=x_{t-1}^\star\right)\mathrm{d}w\\
&\quad\overset{v=\frac{\sqrt{1-\overline{\alpha}_{t-1}}\log w}{\|u\|_2}}{=}1+\int_0^\infty e^{\frac{\|u\|_2 v}{\sqrt{1-\overline{\alpha}_{t-1}}}}\frac{\|u\|_2}{\sqrt{1-\overline{\alpha}_{t-1}}}\mathbb{P}\left(\|x_{t-1}^\star-\sqrt{\overline{\alpha}_{t-1}}X_0\|_2>\sqrt{1-\overline{\alpha}_{t-1}}v\;\Big|\;X_{t-1}=x_{t-1}^\star\right)\mathrm{d}v\\
&\quad\leq 1+5c_5\sqrt{c_6 d\log T}\cdot\frac{\|u\|_2}{\sqrt{1-\overline{\alpha}_{t-1}}}\exp\left(\frac{\|u\|_2\cdot 5c_5\sqrt{c_6 d\log T}}{\sqrt{1-\overline{\alpha}_{t-1}}}\right)+\int_{5c_5\sqrt{c_6 d\log T}}^\infty e^{\frac{\|u\|_2 v}{\sqrt{1-\overline{\alpha}_{t-1}}}}\frac{\|u\|_2}{\sqrt{1-\overline{\alpha}_{t-1}}}e^{-\frac{v^2}{25}}\mathrm{d}v\\
&\quad\leq 1+5c_5\sqrt{c_6 d\log T}\cdot\frac{\|u\|_2}{\sqrt{1-\overline{\alpha}_{t-1}}}\exp\left(\frac{\|u\|_2\cdot 5c_5\sqrt{c_6 d\log T}}{\sqrt{1-\overline{\alpha}_{t-1}}}\right)\\
&\qquad+\frac{\|u\|_2}{\sqrt{1-\overline{\alpha}_{t-1}}}\exp\left(\frac{25\|u\|_2^2}{4(1-\overline{\alpha}_{t-1})}\right)\int_{-\infty}^\infty\exp\left(-\frac{(v-\frac{25\|u\|_2}{2\sqrt{1-\overline{\alpha}_{t-1}}})^2}{25}\right)\mathrm{d}v\\
&\quad=1+5c_5\sqrt{c_6 d\log T}\cdot\frac{\|u\|_2}{\sqrt{1-\overline{\alpha}_{t-1}}}\exp\left(\frac{\|u\|_2\cdot 5c_5\sqrt{c_6 d\log T}}{\sqrt{1-\overline{\alpha}_{t-1}}}\right)+5\sqrt{2\pi}\frac{\|u\|_2}{\sqrt{1-\overline{\alpha}_{t-1}}}\exp\left(\frac{25\|u\|_2^2}{4(1-\overline{\alpha}_{t-1})}\right)\\
&\quad\leq\exp\left(O\Big(\frac{\|x_{t-1}-x_{t-1}^\star\|_2^2}{1-\overline{\alpha}_{t-1}}+\sqrt{\frac{d\|x_{t-1}-x_{t-1}^\star\|_2^2\log T}{1-\overline{\alpha}_{t-1}}}\Big)\right).
\end{aligned}\tag{89}$$

where the last line holds since $1+a\exp(b)+c\exp(d)\leq(1+a+c)\exp(\max\{b,d\})\leq\exp(a+b+c+d)$ for all $a,b,c,d\geq 0$. Combining (88) and the previous inequality, we have finished the proof of (33).

In addition, we make the reservation that $p_{\phi(Y)}(\phi(x))=\mathsf{det}\Big(\frac{\partial\phi(x)}{\partial x}\Big)^{-1}p_Y(x)$. Then we can write

$$\begin{aligned}
\frac{p_{\sqrt{\alpha_t}Y_{t-1}}(\sqrt{\alpha_t}x_{t-1})}{p_{\sqrt{\alpha_t}Y_{t-1}^\star}(\sqrt{\alpha_t}x_{t-1}^\star)}&=\frac{p_{\frac{1}{\sqrt{\overline{\alpha}_{t-1}}}Y_{t-1}}\big(\frac{1}{\sqrt{\overline{\alpha}_{t-1}}}x_{t-1}\big)}{p_{\frac{1}{\sqrt{\overline{\alpha}_{t-1}}}Y_{t-1}^\star}\big(\frac{1}{\sqrt{\overline{\alpha}_{t-1}}}x_{t-1}^\star\big)}\\
&\overset{x_{\tau_{t,0}}=x_{\tau_{t,0}}^\star}{=}\frac{p_{\frac{1}{\sqrt{1-\tau_{t,K-1}}}Y_{t-1}}\big(\frac{1}{\sqrt{1-\tau_{t,K-1}}}x_{\tau_{t,K-1}^{(N)}}\big)/p_{\frac{1}{\sqrt{1-\tau_{t,0}}}Y_t}\big(\frac{1}{1-\tau_{t,0}}x_{\tau_{t,0}}\big)}{p_{\frac{1}{\sqrt{\overline{\alpha}_{t-1}}}Y_{t-1}^\star}\big(\frac{1}{\sqrt{\overline{\alpha}_{t-1}}}x_{t-1}^\star\big)/p_{\frac{1}{\sqrt{1-\tau_{t,0}}}Y_t}\big(\frac{1}{1-\tau_{t,0}}x_{\tau_{t,0}}^\star\big)}\\
&=\mathsf{det}\left(\frac{\partial x_{\tau_{t,K-1}}^\star/\sqrt{1-\tau_{t,K-1}}}{\partial x_{\tau_{t,0}}^\star/\sqrt{1-\tau_{t,0}}}\right)/\mathsf{det}\left(\frac{\partial x_{\tau_{t,K-1}}^{(N)}/\sqrt{1-\tau_{t,K-1}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\right)\\
&=1+\frac{\mathsf{det}\left(\frac{\partial x_{\tau_{t,K-1}}^\star/\sqrt{1-\tau_{t,K-1}}}{\partial x_{\tau_{t,0}}^\star/\sqrt{1-\tau_{t,0}}}\right)-\mathsf{det}\left(\frac{\partial x_{\tau_{t,K-1}}^{(N)}/\sqrt{1-\tau_{t,K-1}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\right)}{\mathsf{det}\left(\frac{\partial x_{\tau_{t,K-1}}^{(N)}/\sqrt{1-\tau_{t,K-1}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\right)}.
\end{aligned}$$

We know from Ipsen and Rehman (2008, Corollary 2.14) that

$$\begin{aligned}
&\left|\frac{\mathsf{det}\left(\frac{\partial x_{\tau_{t,K-1}}^\star/\sqrt{1-\tau_{t,K-1}}}{\partial x_{\tau_{t,0}}^\star/\sqrt{1-\tau_{t,0}}}\right)-\mathsf{det}\left(\frac{\partial x_{\tau_{t,K-1}}^{(N)}/\sqrt{1-\tau_{t,K-1}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\right)}{\mathsf{det}\left(\frac{\partial x_{\tau_{t,K-1}}^{(N)}/\sqrt{1-\tau_{t,K-1}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\right)}\right|\\
&\quad\leq\left(\left\|\frac{\partial x_{\tau_{t,K-1}}^{(N)}/\sqrt{1-\tau_{t,K-1}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\right\|^{-1}\left\|\frac{\partial x_{\tau_{t,K-1}}^\star/\sqrt{1-\tau_{t,K-1}}}{\partial x_{\tau_{t,0}}^\star/\sqrt{1-\tau_{t,0}}}-\frac{\partial x_{\tau_{t,K-1}}^{(N)}/\sqrt{1-\tau_{t,K-1}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\right\|+1\right)^d-1
\end{aligned}$$

$$\leq \exp\left(d\left\|\frac{\partial x^{(N)}_{\tau_{t,K-1}}/\sqrt{1-\tau_{t,K-1}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\right\|^{-1}\left\|\frac{\partial x^{\star}_{\tau_{t,K-1}}/\sqrt{1-\tau_{t,K-1}}}{\partial x^{\star}_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}-\frac{\partial x^{(N)}_{\tau_{t,K-1}}/\sqrt{1-\tau_{t,K-1}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\right\|\right)-1.$$

The previous two inequalities together tell us that (34) holds if $\left\|\frac{\partial x^{(N)}_{\tau_{t,K-1}}/\sqrt{1-\tau_{t,K-1}}}{\partial x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}}\right\|^{-1} \lesssim 1.$

B.5 Proof of Lemma 8

We prove Lemma 8 by induction. For notational convenience, we let $\theta_t := \theta_{\tau_{t,0}}(x^{(0)}_{\tau_{t,i}})$. Recalling that $x^{\star}_{\tau}$ is the solution of ODE (9) at τ with the initial condition $x^{\star}_{\tau_{t,0}} = x_{\tau_{t,0}}$, we know from Lemma 4 that $-\log p_{\overline{X}_{\tau}}(x^{\star}_{\tau}) \leq 2\theta_t d\log T$ for all $\tau_{t,K-1} \leq \tau \leq \tau_{t,0}$. For all $\lambda \in [0,1]$, one can derive

$$\begin{aligned}\frac{p_{\overline{X}_{\tau_{t,i}}}\big(\lambda x^{(0)}_{\tau_{t,i}}+(1-\lambda)x^{\star}_{\tau_{t,i}}\big)}{p_{\overline{X}_{\tau_{t,i}}}(x^{\star}_{\tau_{t,i}})} &= \exp\Big(-\frac{\|u\|_2^2}{2\tau_{t,i}}\Big)\int_{x_0} p_{X_0\,|\,\overline{X}_{\tau_{t,i}}}\big(x_0\,|\,x^{\star}_{\tau_{t,i}}\big)\exp\Big(-\frac{u^{\top}(x^{\star}_{\tau_{t,i}}-\sqrt{1-\tau_{t,i}}x_0)}{\tau_{t,i}}\Big)\mathrm{d}x_0\\ &\geq \exp\Big(-\frac{\|u\|_2^2}{2\tau_{t,i}}\Big)\mathbb{E}\left[\exp\Big(-\frac{\|u\|_2\|x^{\star}_{\tau_{t,i}}-\sqrt{1-\tau_{t,i}}X_0\|_2}{\tau_{t,i}}\Big)\;\Big|\;\overline{X}_{\tau_{t,i}}=x^{\star}_{\tau_{t,i}}\right],\end{aligned}$$

where $u=\lambda(x^{(0)}_{\tau_{t,i}}-x^{\star}_{\tau_{t,i}})$. In view of (89) and Jensen's inequality, we know that

$$\begin{aligned}&\mathbb{E}\left[\exp\Big(-\frac{\|u\|_2\|x^{\star}_{\tau_{t,i}}-\sqrt{1-\tau_{t,i}}X_0\|_2}{\tau_{t,i}}\Big)\;\Big|\;\overline{X}_{\tau_{t,i}}=x^{\star}_{\tau_{t,i}}\right]\\ &\quad\geq\left(\mathbb{E}\left[\exp\Big(\frac{\|u\|_2\|x^{\star}_{\tau_{t,i}}-\sqrt{1-\tau_{t,i}}X_0\|_2}{\tau_{t,i}}\Big)\;\Big|\;\overline{X}_{\tau_{t,i}}=x^{\star}_{\tau_{t,i}}\right]\right)^{-1}\\ &\quad\geq\exp\left(-O\left(\frac{\big\|x^{(0)}_{\tau_{t,i}}-x^{\star}_{\tau_{t,i}}\big\|_2^2}{\tau_{t,i}}+\sqrt{\frac{d\big\|x^{(0)}_{\tau_{t,i}}-x^{\star}_{\tau_{t,i}}\big\|_2^2\log T}{\tau_{t,i}}}\right)\right)\\ &\quad\geq\exp\left(-O\left(\frac{\big\|x^{(0)}_{\tau_{t,i}}-x^{\star}_{\tau_{t,i}}\big\|_2^2}{1-\overline{\alpha}_{t-1}}+\sqrt{\frac{d\big\|x^{(0)}_{\tau_{t,i}}-x^{\star}_{\tau_{t,i}}\big\|_2^2\log T}{1-\overline{\alpha}_{t-1}}}\right)\right).\end{aligned}$$

where the last line makes use of $1-\overline{\alpha}_{t-1}\asymp\tau_{t,i}$. By virtue of (68), one has

$$\big\|x^{(0)}_{\tau_{t,i}}-x^{\star}_{\tau_{t,i}}\big\|_2\lesssim\overline{\alpha}_{t-1}(1-\alpha_t)\frac{1}{\sqrt{1-\overline{\alpha}_{t-1}}}\sqrt{\theta_t d\log T}. \tag{90}$$

The previous two inequalities together with Lemma 1 and (33) imply that

$$\left|\log\frac{p_{\overline{X}_{\tau_{t,i}}}\big(\lambda x^{(0)}_{\tau_{t,i}}+(1-\lambda)x^{\star}_{\tau_{t,i}}\big)}{p_{\overline{X}_{\tau_{t,i}}}(x^{\star}_{\tau_{t,i}})}\right|=O\left(\frac{\theta_t d\log^3 T}{T^2}+\frac{\sqrt{\theta_t}d\log^2 T}{T}\right)=O\left(\frac{\sqrt{\theta_t}d\log^2 T}{T}\right).$$

Therefore, for any i and $\lambda\in[0,1]$, we have

$$-\log p_{\overline{X}_{\tau_{t,i}}}\big(\lambda x^{(0)}_{\tau_{t,i}}+(1-\lambda)x^{\star}_{\tau_{t,i}}\big)\leq 2.1d\theta_t\log T. \tag{91}$$

Then Lemma 3 tells us that

$$\big\|s^{\star}_{\tau_{t,i}}\big(\lambda x^{(0)}_{\tau_{t,i}}+(1-\lambda)x^{\star}_{\tau_{t,i}}\big)\big\|\lesssim\sqrt{\frac{d\theta_t\log T}{\tau_{t,i}}},\qquad\forall 0\leq i\leq K-1. \tag{92}$$

We define

$$u_{t,i}^{(n)} := \sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,i}}} x_{\tau_{t,i}}^{(n)} - x_{\tau_{t,0}}$$
$$= \sqrt{1-\tau_{t,0}} \sum_{0\le j<K} \gamma_{t,j}(\tau_{t,i})(1-\tau_{t,j})^{-3/2} s_{\tau_{t,j}}(x_{\tau_{t,j}}^{(n-1)}).$$

For $n = 1$, recognizing that

$$\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,i}}}\overline{X}_{\tau_{t,i}} \overset{\text{d.}}{=} \sqrt{1-\tau_{t,0}}X_0 + \sqrt{\frac{(1-\tau_{t,0})\tau_{t,i}}{1-\tau_{t,i}}}Z, \qquad \text{where } Z \sim N(0, I_d),$$

one can apply the Bayesian rule to derive

$$p_{\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,i}}}\overline{X}_{\tau_{t,i}}}\left(\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,i}}}x_{\tau_{t,i}}^{(1)}\right)$$
$$= \int_{x_0} p_{X_0}(x_0) p_{\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,i}}}\overline{X}_{\tau_{t,i}} \mid X_0}\left(\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,i}}}x_{\tau_{t,i}}^{(1)} \;\Big|\; x_0\right) \mathrm{d}x_0$$
$$= \int_{x_0} p_{X_0}(x_0) \frac{1}{\left(2\pi\frac{(1-\tau_{t,0})\tau_{t,i}}{1-\tau_{t,i}}\right)^{d/2}} \exp\left(-\frac{\left\|\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,i}}}x_{\tau_{t,i}}^{(1)} - \sqrt{1-\tau_{t,0}}x_0\right\|_2^2}{2\frac{(1-\tau_{t,0})\tau_{t,i}}{1-\tau_{t,i}}}\right) \mathrm{d}x_0$$
$$= \int_{x_0} p_{X_0}(x_0) \frac{1}{\left(2\pi\frac{(1-\tau_{t,0})\tau_{t,i}}{1-\tau_{t,i}}\right)^{d/2}} \exp\left(-\frac{\left\|\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,i}}}x_{\tau_{t,i}}^{(1)} - \sqrt{1-\tau_{t,0}}x_0\right\|_2^2}{2\frac{(1-\tau_{t,0})\tau_{t,i}}{1-\tau_{t,i}}}\right) \mathrm{d}x_0$$
$$= \int_{x_0} p_{X_0}(x_0) \frac{1}{\left(2\pi\frac{(1-\tau_{t,0})\tau_{t,i}}{1-\tau_{t,i}}\right)^{d/2}} \exp\left(-\frac{\left\|x_{\tau_{t,0}} - \sqrt{1-\tau_{t,0}}x_0\right\|_2^2}{2\tau_{t,0}}\right)$$
$$\cdot \exp\left(-\frac{(\tau_{t,0}-\tau_{t,i})\left\|x_{\tau_{t,0}} - \sqrt{1-\tau_{t,0}}x_0\right\|_2^2}{2(1-\tau_{t,0})\tau_{t,0}\tau_{t,i}} - \frac{\left\|u_{t,i}^{(1)}\right\|_2^2 + 2\left\langle u_{t,i}^{(1)}, x_{\tau_{t,0}} - \sqrt{1-\tau_{t,0}}x_0\right\rangle}{2\frac{(1-\tau_{t,0})\tau_{t,i}}{1-\tau_{t,i}}}\right) \mathrm{d}x_0.$$

Therefore, we can rewrite

$$\frac{p_{\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,i}}}\overline{X}_{\tau_{t,i}}}\left(\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,i}}}x_{\tau_{t,i}}^{(1)}\right)}{p_{\overline{X}_{\tau_{t,0}}}(x_{\tau_{t,0}})}$$
$$= \frac{1}{p_{\overline{X}_{\tau_{t,0}}}(x_{\tau_{t,0}})} \int_{x_0} p_{X_0}(x_0) p_{\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,i}}}\overline{X}_{\tau_{t,i}} \mid X_0}\left(\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,i}}}x_{\tau_{t,i}}^{(1)} \;\Big|\; x_0\right) \mathrm{d}x_0$$
$$= \left(\frac{(1-\tau_{t,i})\tau_{t,0}}{(1-\tau_{t,0})\tau_{t,i}}\right)^{d/2} \int_{x_0} p_{X_0 \mid \overline{X}_{\tau_{t,0}}}(x_0 \mid x_{\tau_{t,0}})$$
$$\cdot \exp\left(-\frac{(\tau_{t,i}-\tau_{t,0})\left\|x_{\tau_{t,0}} - \sqrt{1-\tau_{t,0}}x_0\right\|_2^2}{2(1-\tau_{t,0})\tau_{t,0}\tau_{t,i}} - \frac{\left\|u_{t,i}^{(1)}\right\|_2^2 + 2\left\langle u_{t,i}^{(1)}, x_{\tau_{t,0}} - \sqrt{1-\tau_{t,0}}x_0\right\rangle}{2\frac{(1-\tau_{t,0})\tau_{t,i}}{1-\tau_{t,i}}}\right) \mathrm{d}x_0$$
$$= \left(1 + \frac{d}{2}\frac{\tau_{t,0}-\tau_{t,i}}{(1-\tau_{t,0})\tau_{t,i}} + O\left(d^2\left(\frac{\tau_{t,0}-\tau_{t,i}}{(1-\tau_{t,0})\tau_{t,i}}\right)^2\right)\right)$$
$$\cdot \int_{x_0} p_{X_0 \mid \overline{X}_{\tau_{t,0}}}(x_0 \mid x_{\tau_{t,0}})$$

$$\cdot \exp\left(-\frac{(\tau_{t,0}-\tau_{t,i})\left\|x_{\tau_{t,0}}-\sqrt{1-\tau_{t,0}}x_0\right\|_2^2}{2(1-\tau_{t,0})\tau_{t,0}\tau_{t,i}}-\frac{\left\|u_{t,i}^{(1)}\right\|_2^2+2\langle u_{t,i}^{(1)},x_{\tau_{t,0}}-\sqrt{1-\tau_{t,0}}x_0\rangle}{2\frac{(1-\tau_{t,0})\tau_{t,i}}{1-\tau_{t,i}}}\right)\mathrm{d}x_0. \tag{93}$$

By virtue of (92), we can derive

$$\begin{aligned}
&\left\|u_{t,i}^{(1)}-\left(-\sqrt{1-\tau_{t,0}}\int_{\tau_{t,0}}^{\tau_{t,i}}\frac{1}{2(1-\tau)^{3/2}}s_\tau^\star(x_\tau^\star)\mathrm{d}\tau\right)\right\|_2\\
&=\sqrt{1-\tau_{t,0}}\Bigg\|\sum_{0\le j<K}\gamma_{t,j}(\tau_{t,i})(1-\tau_{t,j})^{-3/2}\big(s_{\tau_{t,j}}(x_{\tau_{t,j}}^{(0)})-s_{\tau_{t,j}}^\star(x_{\tau_{t,j}}^{(0)})\big)\\
&\qquad\qquad+\left[\sum_{0\le j<k}\gamma_{t,j}(\tau_{t,i})(1-\tau_{t,j})^{-3/2}s_{\tau_{t,j}}^\star(x_{\tau_{t,j}}^\star)-\left(-\int_{\tau_{t,0}}^{\tau'}\frac{1}{2(1-\tau)^{3/2}}s_\tau^\star(x_\tau^\star)\mathrm{d}\tau\right)\right]\Bigg\|_2\\
&\le\sqrt{1-\tau_{t,0}}\sum_{0\le j<K}\left|\gamma_{t,j}(\tau_{t,i})\right|(1-\tau_{t,j})^{-3/2}\left\|s_{\tau_{t,j}}(x_{\tau_{t,j}}^{(0)})-s_{\tau_{t,j}}^\star(x_{\tau_{t,j}}^{(0)})\right\|\\
&\quad+\sqrt{1-\tau_{t,0}}\left\|\sum_{0\le j<K}\gamma_{t,j}(\tau_{t,i})(1-\tau_{t,j})^{-3/2}s_{\tau_{t,j}}^\star(x_{\tau_{t,j}}^\star)-\left(-\int_{\tau_{t,0}}^{\tau'}\frac{1}{2(1-\tau)^{3/2}}s_\tau^\star(x_\tau^\star)\mathrm{d}\tau\right)\right\|\\
&\asymp\sum_{0\le j<K}(\tau_{t,0}-\tau_{t,i})(1-\tau_{t,0})^{-1}\left\|s_{\tau_{t,j}}(x_{\tau_{t,j}}^{(0)})-s_{\tau_{t,j}}^\star(x_{\tau_{t,j}}^{(0)})\right\|_2\\
&\quad+\sqrt{1-\tau_{t,0}}\sqrt{\frac{d\theta_t(\tau_{t,i}-\tau_{t,0})^2\log T}{\tau_{t,0}(1-\tau_{t,0})^3}}\left(\frac{d\theta_t(\tau_{t,0}-\tau_{t,i})\log T}{\tau_{t,0}(1-\tau_{t,0})}\right)^K\\
&\lesssim\frac{\tau_{t,0}\log T}{T}\sum_{0\le j<K}\left\|s_{\tau_{t,j}}(x_{\tau_{t,j}}^{(0)})-s_{\tau_{t,j}}^\star(x_{\tau_{t,j}}^{(0)})\right\|_2+\sqrt{\frac{d\theta_t\tau_{t,0}\log^3 T}{T^2}}\left(\frac{d\theta_t\log^2 T}{T}\right)^K\\
&\lesssim\frac{\tau_{t,0}\log T}{T}\sqrt{K\sum_i\big(\varepsilon_{\mathsf{score},t,i}^{(n)}(x_{\tau_{t,i}}^{(0)})\big)^2}+\sqrt{\frac{d\theta_t\tau_{t,0}\log^3 T}{T^2}}\left(\frac{d\theta_t\log^2 T}{T}\right)^K\\
&\asymp\frac{\tau_{t,0}\log T}{T}\sqrt{\sum_i\big(\varepsilon_{\mathsf{score},t,i}^{(0)}(x_{\tau_{t,i}}^{(0)})\big)^2}+\sqrt{\frac{d\theta_t\tau_{t,0}\log^3 T}{T^2}}\left(\frac{d\theta_t\log^2 T}{T}\right)^K.
\end{aligned}\tag{94}$$

Here, the fourth line makes use of (23g), (23h), (64a) and Taylor's Theorem; the fifth line holds due to (23f); the fifth line comes from Cauchy-Schwarz inequality and the last line is valid since K is a constant. By virtue of Lemmas 1, 3 and 4, one has

$$\begin{aligned}
\left\|-\sqrt{1-\tau_{t,0}}\int_{\tau_{t,0}}^{\tau_{t,i}}\frac{1}{2(1-\tau)^{3/2}}s_\tau^\star(x_\tau^\star)\mathrm{d}\tau\right\|_2&\lesssim\sqrt{1-\tau_{t,0}}\int_{\tau_{t,i}}^{\tau_{t,0}}\frac{1}{2(1-\tau)^{3/2}}\sqrt{\frac{d\theta_t\log T}{\tau}}d\tau\\
&\lesssim\sqrt{1-\tau_{t,0}}\frac{\tau_{t,0}-\tau_{t,i}}{2(1-\tau_{t,0})^{3/2}}\sqrt{\frac{d\theta_t\log T}{\tau_{t,0}}}\\
&=\frac{\tau_{t,0}-\tau_{t,i}}{2(1-\tau_{t,0})\tau_{t,0}}\sqrt{\tau_{t,0}d\theta_t\log T} \qquad (95)\\
&\lesssim\frac{\sqrt{\tau_{t,0}d\theta_t\log^3 T}}{T}. \qquad (96)
\end{aligned}$$

Combining the previous two inequalities, we obtain

$$\left\|u_{t,i}^{(1)}\right\|_2\lesssim\frac{\sqrt{\tau_{t,0}d\theta_t\log^3 T}}{T}+\frac{\tau_{t,0}\log T}{T}\sqrt{\sum_i\big(\varepsilon_{\mathsf{score},t,i}^{(0)}(x_{\tau_{t,i}}^{(0)})\big)^2}. \tag{97}$$

We denote

$$\mathcal{E}_\ell^{\mathsf{typical}} := \{x_0 : \|x_{\tau_{t,0}} - \sqrt{1-\tau_{t,0}}x_0\|_2 \le 5\ell\sqrt{d\theta_t\tau_{t,0}\log T}\}, \qquad \ell = 1, 2, \dots$$

Then for any $x_0 \in \mathcal{E}_\ell^{\text{typical}}$, we have

$$\frac{(\tau_{t,0}-\tau_{t,i})\left\|x_{\tau_{t,0}} - \sqrt{1-\tau_{t,0}}x_0\right\|_2^2}{2(1-\tau_{t,0})\tau_{t,0}\tau_{t,i}} \le \frac{25\ell^2}{2}\frac{(\tau_{t,0}-\tau_{t,i})d\theta_t\log T}{(1-\tau_{t,0})\tau_{t,i}} \overset{\text{Lemma 1}}{\lesssim} \ell^2\frac{d\theta_t\log^2 T}{T}, \tag{98}$$

$$\begin{aligned}
&\left|\frac{\left\|u_{t,i}^{(1)}\right\|_2^2 + 2\left\langle u_{t,i}^{(1)}, x_{\tau_{t,0}} - \sqrt{1-\tau_{t,0}}x_0\right\rangle}{2\frac{(1-\tau_{t,0})\tau_{t,i}}{1-\tau_{t,i}}}\right| \\
&\quad\le \frac{\left\|u_{t,i}^{(1)}\right\|_2^2 + 2\left\|u_{t,i}^{(1)}\right\|_2\|x_{\tau_{t,0}} - \sqrt{1-\tau_{t,0}}x_0\|_2}{2\frac{(1-\tau_{t,0})\tau_{t,i}}{1-\tau_{t,i}}} \\
&\quad\overset{\text{(97) and Lemma 1}}{\lesssim} \frac{d\theta_t\log^3 T}{T^2} + \frac{\tau_{t,0}\log^2 T}{T^2}\sum_i\left(\varepsilon_{\mathsf{score},t,i}^{(0)}(x_{\tau_{t,i}}^{(0)})\right)^2 + \ell\frac{d\theta_t\log^2 T}{T} \\
&\qquad\qquad + \frac{\ell\sqrt{d\theta_t\tau_{t,0}\log^3 T}}{T}\sqrt{\sum_i\left(\varepsilon_{\mathsf{score},t,i}^{(0)}(x_{\tau_{t,i}}^{(0)})\right)^2} \\
&\quad\lesssim \ell\frac{d\theta_t\log^2 T}{T} + \frac{\ell\sqrt{d\theta_t\tau_{t,0}\log^3 T}}{T}\sqrt{\sum_i\left(\varepsilon_{\mathsf{score},t,i}^{(0)}(x_{\tau_{t,i}}^{(0)})\right)^2}.
\end{aligned} \tag{99}$$

Combining the previous two inequalities and the fact $\exp(-z) = 1 - z + O(z^2)$, we know that for any $x_0 \in \mathcal{E}_2^{\mathsf{typical}}$,

$$\begin{aligned}
&\exp\left(-\frac{(\tau_{t,0}-\tau_{t,i})\left\|x_{\tau_{t,0}} - \sqrt{1-\tau_{t,0}}x_0\right\|_2^2}{2(1-\tau_{t,0})\tau_{t,0}\tau_{t,i}} - \frac{\left\|u_{t,i}^{(1)}\right\|_2^2 + 2\left\langle u_{t,i}^{(1)}, x_{\tau_{t,0}} - \sqrt{1-\tau_{t,0}}x_0\right\rangle}{2\frac{(1-\tau_{t,0})\tau_{t,i}}{1-\tau_{t,i}}}\right) \\
&= 1 - \frac{(\tau_{t,0}-\tau_{t,i})\left\|x_{\tau_{t,0}} - \sqrt{1-\tau_{t,0}}x_0\right\|_2^2}{2(1-\tau_{t,0})\tau_{t,0}\tau_{t,i}} - \frac{\left\|u_{t,i}^{(1)}\right\|_2^2 + 2\left\langle u_{t,i}^{(1)}, x_{\tau_{t,0}} - \sqrt{1-\tau_{t,0}}x_0\right\rangle}{2\frac{(1-\tau_{t,0})\tau_{t,i}}{1-\tau_{t,i}}} \\
&\quad + O\left(\frac{d^2\theta_t^2\log^4 T}{T^2} + \frac{d\theta_t\tau_{t,0}\log^3 T}{T^2}\sum_i\left(\varepsilon_{\mathsf{score},t,i}^{(0)}(x_{\tau_{t,i}}^{(0)})\right)^2\right) \\
&= 1 - \frac{(\tau_{t,0}-\tau_{t,i})\left\|x_{\tau_{t,0}} - \sqrt{1-\tau_{t,0}}x_0\right\|_2^2}{2(1-\tau_{t,0})\tau_{t,0}\tau_{t,i}} - \frac{2\left\langle u_{t,i}^{(1)}, x_{\tau_{t,0}} - \sqrt{1-\tau_{t,0}}x_0\right\rangle}{2\frac{(1-\tau_{t,0})\tau_{t,i}}{1-\tau_{t,i}}} \\
&\quad + O\left(\frac{d^2\theta_t^2\log^4 T}{T^2} + \frac{d\theta_t\tau_{t,0}\log^3 T}{T^2}\sum_i\left(\varepsilon_{\mathsf{score},t,i}^{(0)}(x_{\tau_{t,i}}^{(0)})\right)^2\right),
\end{aligned} \tag{100}$$

and for any $x_0 \in \mathcal{E}_\ell^{\mathsf{typical}}$,

$$-\frac{(\tau_{t,0}-\tau_{t,i})\left\|x_{\tau_{t,0}} - \sqrt{1-\tau_{t,0}}x_0\right\|_2^2}{2(1-\tau_{t,0})\tau_{t,0}\tau_{t,i}} - \frac{\left\|u_{t,i}^{(1)}\right\|_2^2 + 2\left\langle u_{t,i}^{(1)}, x_{\tau_{t,0}} - \sqrt{1-\tau_{t,0}}x_0\right\rangle}{2\frac{(1-\tau_{t,0})\tau_{t,i}}{1-\tau_{t,i}}} \le \ell d\theta_t. \tag{101}$$

We make the observation that

$$\int_{x_0} p_{X_0 \mid \overline{X}_{\tau_{t,0}}}(x_0 \mid x_{\tau_{t,0}})$$

$$\cdot \exp\left(-\frac{(\tau_{t,0}-\tau_{t,i})\left\|x_{\tau_{t,0}}-\sqrt{1-\tau_{t,0}}x_0\right\|_2^2}{2(1-\tau_{t,0})\tau_{t,0}\tau_{t,i}}-\frac{\left\|u_{t,i}^{(1)}\right\|_2^2+2\big\langle u_{t,i}^{(1)},x_{\tau_{t,0}}-\sqrt{1-\tau_{t,0}}x_0\big\rangle}{2\frac{(1-\tau_{t,0})\tau_{t,i}}{1-\tau_{t,i}}}\right)\mathrm{d}x_0$$

$$= \int_{x_0\in\mathcal{E}_2^{\mathsf{typical}}} p_{X_0\,|\,\overline{X}_{\tau_{t,0}}}(x_0\,|\,x_{\tau_{t,0}})$$

$$\cdot \exp\left(-\frac{(\tau_{t,0}-\tau_{t,i})\left\|x_{\tau_{t,0}}-\sqrt{1-\tau_{t,0}}x_0\right\|_2^2}{2(1-\tau_{t,0})\tau_{t,0}\tau_{t,i}}-\frac{\left\|u_{t,i}^{(1)}\right\|_2^2+2\big\langle u_{t,i}^{(1)},x_{\tau_{t,0}}-\sqrt{1-\tau_{t,0}}x_0\big\rangle}{2\frac{(1-\tau_{t,0})\tau_{t,i}}{1-\tau_{t,i}}}\right)\mathrm{d}x_0$$

$$+\sum_{\ell=3}^{\infty}\int_{x_0\in\mathcal{E}_\ell^{\mathsf{typical}}\setminus\mathcal{E}_{\ell-1}^{\mathsf{typical}}} p_{X_0\,|\,\overline{X}_{\tau_{t,0}}}(x_0\,|\,x_{\tau_{t,0}})$$

$$\cdot \exp\left(-\frac{(\tau_{t,0}-\tau_{t,i})\left\|x_{\tau_{t,0}}-\sqrt{1-\tau_{t,0}}x_0\right\|_2^2}{2(1-\tau_{t,0})\tau_{t,0}\tau_{t,i}}-\frac{\left\|u_{t,i}^{(1)}\right\|_2^2+2\big\langle u_{t,i}^{(1)},x_{\tau_{t,0}}-\sqrt{1-\tau_{t,0}}x_0\big\rangle}{2\frac{(1-\tau_{t,0})\tau_{t,i}}{1-\tau_{t,i}}}\right)\mathrm{d}x_0$$

$$\overset{(100)\text{ and }(101)}{=} \int_{x_0\in\mathcal{E}_2^{\mathsf{typical}}} p_{X_0\,|\,\overline{X}_{\tau_{t,0}}}(x_0\,|\,x_{\tau_{t,0}})$$

$$\cdot\left(1-\frac{(\tau_{t,0}-\tau_{t,i})\left\|x_{\tau_{t,0}}-\sqrt{1-\tau_{t,0}}x_0\right\|_2^2}{2(1-\tau_{t,0})\tau_{t,0}\tau_{t,i}}-\frac{2\big\langle u_{t,i}^{(1)},x_{\tau_{t,0}}-\sqrt{1-\tau_{t,0}}x_0\big\rangle}{2\frac{(1-\tau_{t,0})\tau_{t,i}}{1-\tau_{t,i}}}\right)\mathrm{d}x_0$$

$$+O\left(\frac{d^2\theta_t^2\log^4 T}{T^2}+\frac{d\theta_t\tau_{t,0}\log^3 T}{T^2}\sum_i\big(\varepsilon_{\mathsf{score},t,i}^{(0)}(x_{\tau_{t,i}}^{(0)})\big)^2\right)$$

$$+O\left(\sum_{\ell=3}^{\infty}\int_{x_0\in\mathcal{E}_\ell^{\mathsf{typical}}\setminus\mathcal{E}_{\ell-1}^{\mathsf{typical}}} p_{X_0\,|\,\overline{X}_{\tau_{t,0}}}(x_0\,|\,x_{\tau_{t,0}})\mathrm{d}x_0\cdot\exp\left(\ell d\theta_t\right)\right) \tag{102}$$

$$\overset{(94)\text{ and Lemma 2}}{=} \int_{x_0\in\mathcal{E}_2^{\mathsf{typical}}} p_{X_0\,|\,\overline{X}_{\tau_{t,0}}}(x_0\,|\,x_{\tau_{t,0}})\left(1-\frac{(\tau_{t,0}-\tau_{t,i})\left\|x_{\tau_{t,0}}-\sqrt{1-\tau_{t,0}}x_0\right\|_2^2}{2(1-\tau_{t,0})\tau_{t,0}\tau_{t,i}}\right)\mathrm{d}x_0$$

$$-\int_{x_0\in\mathcal{E}_2^{\mathsf{typical}}} p_{X_0\,|\,\overline{X}_{\tau_{t,0}}}(x_0\,|\,x_{\tau_{t,0}})\frac{2\big\langle -\sqrt{1-\tau_{t,0}}\int_{\tau_{t,0}}^{\tau_{t,i}}\frac{1}{2(1-\tau)^{3/2}}s_\tau^\star(x_\tau^\star)\mathrm{d}\tau,x_{\tau_{t,0}}-\sqrt{1-\tau_{t,0}}x_0\big\rangle}{2\frac{(1-\tau_{t,0})\tau_{t,i}}{1-\tau_{t,i}}}\mathrm{d}x_0$$

$$+O\left(\left(\frac{\tau_{t,0}\log T}{T}\sqrt{\sum_i\big(\varepsilon_{\mathsf{score},t,i}^{(0)}(x_{\tau_{t,i}}^{(0)})\big)^2}+\sqrt{\frac{d\theta_t\tau_{t,0}\log^3 T}{T^2}}\left(\frac{d\theta_t\log^2 T}{T}\right)^K\right)\frac{\sqrt{d\theta_t\tau_{t,0}\log T}}{\frac{(1-\tau_{t,0})\tau_{t,i}}{1-\tau_{t,i}}}\right)$$

$$+O\left(\frac{d^2\theta_t^2\log^4 T}{T^2}+\frac{d\theta_t\tau_{t,0}\log^3 T}{T^2}\sum_i\big(\varepsilon_{\mathsf{score},t,i}^{(0)}(x_{\tau_{t,i}}^{(0)})\big)^2\right)+O\left(\sum_{\ell=3}^{\infty}\exp\left(-(\ell-1)^2d\theta\log T\right)\cdot\exp\left(\ell d\theta_t\right)\right). \tag{103}$$

Furthermore, (96), (98), (99) and Lemma 2 together imply that

RHS of (103)

$$= 1-\frac{(\tau_{t,0}-\tau_{t,i})\int_{x_0}p_{X_0\,|\,\overline{X}_{\tau_{t,0}}}(x_0\,|\,x_{\tau_{t,0}})\left\|x_{\tau_{t,0}}-\sqrt{1-\tau_{t,0}}x_0\right\|_2^2\mathrm{d}x_0}{2(1-\tau_{t,0})\tau_{t,0}\tau_{t,i}}$$

$$-\frac{2\big\langle -\sqrt{1-\tau_{t,0}}\int_{\tau_{t,0}}^{\tau_{t,i}}\frac{1}{2(1-\tau)^{3/2}}s_\tau^\star(x_\tau^\star)\mathrm{d}\tau,\int_{x_0}p_{X_0\,|\,\overline{X}_{\tau_{t,0}}}(x_0\,|\,x_{\tau_{t,0}})(x_{\tau_{t,0}}-\sqrt{1-\tau_{t,0}}x_0)\mathrm{d}x_0\big\rangle}{2\frac{(1-\tau_{t,0})\tau_{t,i}}{1-\tau_{t,i}}}$$

$$-\int_{x_0\notin\mathcal{E}_2^{\mathsf{typical}}} p_{X_0\,|\,\overline{X}_{\tau_{t,0}}}(x_0\,|\,x_{\tau_{t,0}})\mathrm{d}x_0$$

$$+\sum_{\ell=3}^{\infty}\int_{x_0\in\mathcal{E}_\ell^{\mathsf{typical}}\setminus\mathcal{E}_{\ell-1}^{\mathsf{typical}}} p_{X_0\,|\,\overline{X}_{\tau_{t,0}}}(x_0\,|\,x_{\tau_{t,0}})\frac{(\tau_{t,0}-\tau_{t,i})\left\|x_{\tau_{t,0}}-\sqrt{1-\tau_{t,0}}x_0\right\|_2^2}{2(1-\tau_{t,0})\tau_{t,0}\tau_{t,i}}\mathrm{d}x_0$$

$$
\begin{aligned}
&+ O\Bigg(\Big\| - \sqrt{1-\tau_{t,0}} \int_{\tau_{t,0}}^{\tau_{t,i}} \frac{1}{2(1-\tau)^{3/2}} s^{\star}_{\tau}(x^{\star}_{\tau}) \mathrm{d}\tau \Big\|_2 \\
&\qquad \sum_{\ell=3}^{\infty} \int_{x_0 \in \mathcal{E}^{\mathsf{typical}}_{\ell} \setminus \mathcal{E}^{\mathsf{typical}}_{\ell-1}} p_{X_0 \mid \overline{X}_{\tau_{t,0}}}(x_0 \mid x_{\tau_{t,0}}) \left| \frac{\big\|u^{(1)}_{t,i}\big\|_2^2 + 2\big\langle u^{(1)}_{t,i}, x_{\tau_{t,0}} - \sqrt{1-\tau_{t,0}} x_0 \big\rangle}{2\frac{(1-\tau_{t,0})\tau_{t,i}}{1-\tau_{t,i}}} \right| \mathrm{d}x_0 \Bigg) \\
&+ O\left(\frac{d^2 \theta_t^2 \log^4 T}{T^2} + \frac{\sqrt{d\theta_t \tau_{t,0} \log^3 T}}{T} \sqrt{\sum_i \big(\varepsilon^{(0)}_{\mathsf{score},t,i}(x^{(0)}_{\tau_{t,i}})\big)^2} \right) \\
&= 1 - \frac{(\tau_{t,0} - \tau_{t,i}) \int_{x_0} p_{X_0 \mid \overline{X}_{\tau_{t,0}}}(x_0 \mid x_{\tau_{t,0}}) \big\| x_{\tau_{t,0}} - \sqrt{1-\tau_{t,0}} x_0 \big\|_2^2 \mathrm{d}x_0}{2(1-\tau_{t,0})\tau_{t,0}\tau_{t,i}} \\
&\quad - \frac{2\big\langle - \sqrt{1-\tau_{t,0}} \int_{\tau_{t,0}}^{\tau_{t,i}} \frac{1}{2(1-\tau)^{3/2}} s^{\star}_{\tau}(x^{\star}_{\tau}) \mathrm{d}\tau, \int_{x_0} p_{X_0 \mid \overline{X}_{\tau_{t,0}}}(x_0 \mid x_{\tau_{t,0}})(x_{\tau_{t,0}} - \sqrt{1-\tau_{t,0}} x_0) \mathrm{d}x_0 \big\rangle}{2\frac{(1-\tau_{t,0})\tau_{t,i}}{1-\tau_{t,i}}} \\
&+ O\left(\frac{d^2 \theta_t^2 \log^4 T}{T^2} + \frac{\sqrt{d\theta_t \tau_{t,0} \log^3 T}}{T} \sqrt{\sum_i \big(\varepsilon^{(0)}_{\mathsf{score},t,i}(x^{(0)}_{\tau_{t,i}})\big)^2} \right). \qquad (104)
\end{aligned}
$$

In addition, one can show that

$$
\begin{aligned}
&\left\| 2\sqrt{1-\tau_{t,0}} \int_{\tau_{t,i}}^{\tau_{t,0}} \frac{1}{2(1-\tau)^{3/2}} s^{\star}_{\tau}(x^{\star}_{\tau}) \mathrm{d}\tau - \frac{\tau_{t,0} - \tau_{t,i}}{1-\tau_{t,i}} s^{\star}_{\tau_{t,0}}(x_{\tau_{t,0}}) \right\|_2 \\
&\quad \le \left\| 2\sqrt{1-\tau_{t,0}} \int_{\tau_{t,i}}^{\tau_{t,0}} \left(\frac{1}{2(1-\tau)^{3/2}} s^{\star}_{\tau}(x^{\star}_{\tau}) - \frac{1}{2(1-\tau_{t,0})^{3/2}} s^{\star}_{\tau_{t,0}}(x^{\star}_{\tau_{t,0}}) \right) \mathrm{d}\tau \right\|_2 \\
&\qquad + \left(\frac{\tau_{t,0} - \tau_{t,i}}{1-\tau_{t,i}} - \frac{\tau_{t,0} - \tau_{t,i}}{1-\tau_{t,0}} \right) \big\| s^{\star}_{\tau_{t,0}}(x_{\tau_{t,0}}) \big\|_2 \\
&\quad \lesssim 2\sqrt{1-\tau_{t,0}} \int_{\tau_{t,i}}^{\tau_{t,0}} (\tau_{t,0} - \tau) \sup_{\tau' \in [\tau, \tau_{t,0}]} \sqrt{\frac{d\theta_t \log T}{\tau'(1-\tau')^3}} \frac{d\theta_t \log T}{\tau'(1-\tau')} \mathrm{d}\tau + \frac{(\tau_{t,0} - \tau_{t,i})^2}{(1-\tau_{t,i})(1-\tau_{t,0})} \sqrt{\frac{d\theta_t \log T}{\tau_{t,0}}} \\
&\quad \lesssim \sqrt{1-\tau_{t,0}} (\tau_{t,0} - \tau_{t,i})^2 \frac{d^{3/2} \theta_t^{3/2} \log^{3/2} T}{\tau_{t,0}^{3/2} (1-\tau_{t,0})^{5/2}} + \frac{\tau_{t,0} - \tau_{t,i}}{1-\tau_{t,i}} \frac{\tau_{t,0} - \tau_{t,i}}{1-\tau_{t,0}} \sqrt{\frac{d\theta_t \log T}{\tau_{t,0}}} \\
&\quad \lesssim \tau_{t,0}^{1/2} \frac{d^{3/2} \theta_t^{3/2} \log^{7/2} T}{T^2} + \tau_{t,0}^{3/2} \frac{d^{1/2} \theta_t^{1/2} \log^{5/2} T}{T^2} \asymp \tau_{t,0}^{1/2} \frac{d^{3/2} \theta_t^{3/2} \log^{7/2} T}{T^2}.
\end{aligned}
$$

Here, the third line is valid due to (64a) and Lemma 3; the penultimate line and the last line result from Lemma 1. Then we have

$$
\begin{aligned}
&\frac{\Big| \big\langle - 2\sqrt{1-\tau_{t,0}} \int_{\tau_{t,0}}^{\tau_{t,i}} \frac{1}{2(1-\tau)^{3/2}} s^{\star}_{\tau}(x^{\star}_{\tau}) \mathrm{d}\tau - \frac{\tau_{t,0} - \tau_{t,i}}{1-\tau_{t,i}} s^{\star}_{\tau_{t,0}}(x_{\tau_{t,0}}), \int_{x_0} p_{X_0 \mid \overline{X}_{\tau_{t,0}}}(x_0 \mid x_{\tau_{t,0}})(x_{\tau_{t,0}} - \sqrt{1-\tau_{t,0}} x_0) \mathrm{d}x_0 \big\rangle \Big|}{2\frac{(1-\tau_{t,0})\tau_{t,i}}{1-\tau_{t,i}}} \\
&\overset{\text{Lemma 1}}{\lesssim} \frac{\big\| 2\sqrt{1-\tau_{t,0}} \int_{\tau_{t,i}}^{\tau_{t,0}} \frac{1}{2(1-\tau)^{3/2}} s^{\star}_{\tau}(x^{\star}_{\tau}) \mathrm{d}\tau - \frac{\tau_{t,0} - \tau_{t,i}}{1-\tau_{t,i}} s^{\star}_{\tau_{t,0}}(x_{\tau_{t,0}}) \big\|_2}{\tau_{t,0}} \\
&\qquad \cdot \Big\| \int_{x_0} p_{X_0 \mid \overline{X}_{\tau_{t,0}}}(x_0 \mid x_{\tau_{t,0}})(x_{\tau_{t,0}} - \sqrt{1-\tau_{t,0}} x_0) \mathrm{d}x_0 \Big\|_2 \\
&\overset{\text{Lemma 2}}{\lesssim} \frac{\tau_{t,0}^{1/2} \frac{d^{3/2} \theta_t^{3/2} \log^{7/2} T}{T^2} \cdot \sqrt{\theta_t d \tau_{t,0} \log T}}{\tau_{t,0}} \\
&= \frac{d^2 \theta_t^2 \log^4 T}{T^2}. \qquad (105)
\end{aligned}
$$

Combining (103), (104) and (105), one has

$$\begin{aligned}
&\int_{x_0} p_{X_0 \,|\, \overline{X}_{\tau_{t,0}}}(x_0 \,|\, x_{\tau_{t,0}}) \\
&\quad \cdot \exp\left(-\frac{(\tau_{t,0}-\tau_{t,i})\left\|x_{\tau_{t,0}}-\sqrt{1-\tau_{t,0}}x_0\right\|_2^2}{2(1-\tau_{t,0})\tau_{t,0}\tau_{t,i}} - \frac{\left\|u_{t,i}^{(1)}\right\|_2^2 + 2\left\langle u_{t,i}^{(1)}, x_{\tau_{t,0}} - \sqrt{1-\tau_{t,0}}x_0\right\rangle}{2\frac{(1-\tau_{t,0})\tau_{t,i}}{1-\tau_{t,i}}}\right)\mathrm{d}x_0 \\
&= 1 - \frac{(\tau_{t,0}-\tau_{t,i})\int_{x_0} p_{X_0 \,|\, \overline{X}_{\tau_{t,0}}}(x_0 \,|\, x_{\tau_{t,0}})\left\|x_{\tau_{t,0}}-\sqrt{1-\tau_{t,0}}x_0\right\|_2^2\mathrm{d}x_0}{2(1-\tau_{t,0})\tau_{t,0}\tau_{t,i}} \\
&\quad - \frac{\left\langle \frac{\tau_{t,0}-\tau_{t,i}}{1-\tau_{t,i}} s^{\star}_{\tau_{t,0}}(x_{\tau_{t,0}}), \int_{x_0} p_{X_0 \,|\, \overline{X}_{\tau_{t,0}}}(x_0 \,|\, x_{\tau_{t,0}})(x_{\tau_{t,0}}-\sqrt{1-\tau_{t,0}}x_0)\mathrm{d}x_0\right\rangle}{2\frac{(1-\tau_{t,0})\tau_{t,i}}{1-\tau_{t,i}}} \\
&\quad + O\left(\frac{d^2\theta_t^2\log^4 T}{T^2} + \frac{\sqrt{d\theta_t\tau_{t,0}\log^3 T}}{T}\sqrt{\sum_i \left(\varepsilon_{\mathsf{score},t,i}^{(0)}(x_{\tau_{t,i}}^{(0)})\right)^2}\right) \\
&\overset{(8)}{=} 1 - \frac{\tau_{t,0}-\tau_{t,i}}{2(1-\tau_{t,0})\tau_{t,0}\tau_{t,i}}\Bigg[\int_{x_0} p_{X_0 \,|\, \overline{X}_{\tau_{t,0}}}(x_0 \,|\, x_{\tau_{t,0}})\left\|x_{\tau_{t,0}}-\sqrt{1-\tau_{t,0}}x_0\right\|_2^2\mathrm{d}x_0 \\
&\qquad\qquad - \left\|\int_{x_0} p_{X_0 \,|\, \overline{X}_{\tau_{t,0}}}(x_0 \,|\, x_{\tau_{t,0}})(x_{\tau_{t,0}}-\sqrt{1-\tau_{t,0}}x_0)\mathrm{d}x_0\right\|_2^2\Bigg] \\
&\quad + O\left(\frac{d^2\theta_t^2\log^4 T}{T^2} + \frac{\sqrt{d\theta_t\tau_{t,0}\log^3 T}}{T}\sqrt{\sum_i \left(\varepsilon_{\mathsf{score},t,i}^{(0)}(x_{\tau_{t,i}}^{(0)})\right)^2}\right) \\
&\overset{\text{Jensen's inequality}}{\le} 1 + O\left(\frac{d^2\theta_t^2\log^4 T}{T^2} + \frac{\sqrt{d\theta_t\tau_{t,0}\log^3 T}}{T}\sqrt{\sum_i \left(\varepsilon_{\mathsf{score},t,i}^{(0)}(x_{\tau_{t,i}}^{(0)})\right)^2}\right).
\end{aligned}$$

Equation (93) together with the previous inequality and Lemma 1 implies that

$$\log\frac{p_{\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,i}}}\overline{X}_{\tau_{t,i}}}\left(\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,i}}}x_{\tau_{t,i}}^{(1)}\right)}{p_{\overline{X}_{\tau_{t,0}}}(x_{\tau_{t,0}})} \le \frac{4c_1 d\log T}{T} + C_{10}\left\{\frac{d^2\theta_t^2\log^4 T}{T^2} + \frac{\sqrt{d\theta_t\log^3 T}}{T}\sqrt{\sum_i \left(\varepsilon_{\mathsf{score},t,i}^{(0)}(x_{\tau_{t,i}}^{(0)})\right)^2}\right\}. \tag{106}$$

Similarly, for any $\lambda \in [0,1]$, one has

$$\begin{aligned}
&\left|\log\frac{p_{\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,i}}}\overline{X}_{\tau_{t,i}}}\left(\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,i}}}\left(\lambda x^{\star}_{\tau_{t,i}} + (1-\lambda)x_{\tau_{t,i}}^{(1)}\right)\right)}{p_{\overline{X}_{\tau_{t,0}}}(x_{\tau_{t,0}})}\right| \\
&\le C_{10}\left\{\frac{d\theta_t\log^2 T}{T} + \frac{\ell\sqrt{d\theta_t\log^3 T}}{T}\sqrt{\sum_i \left(\varepsilon_{\mathsf{score},t,i}^{(0)}(x_{\tau_{t,i}}^{(0)})\right)^2}\right\}
\end{aligned} \tag{107}$$

Repeating similar arguments and using induction yields that for all $0 \le n \le N-1$, $0 \le i \le K-1$,

$$-\log p_{\overline{X}_{\tau_{t,i}}}\left(\lambda x_{\tau_{t,i}}^{(n+1)} + (1-\lambda)x^{\star}_{\tau_{t,i}}\right) \le 2.1 d\theta_t\log T, \tag{108a}$$

$$\log\frac{p_{\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,i}}}\overline{X}_{\tau_{t,i}}}\left(\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,i}}}x_{\tau_{t,i}}^{(n+1)}\right)}{p_{\overline{X}_{\tau_{t,0}}}(x_{\tau_{t,0}})} \le \frac{4c_1 d\log T}{T} + C_{10}\left\{\frac{d^2\theta_t^2\log^4 T}{T^2} + \frac{\sqrt{d\theta_t\log^3 T}}{T}\sqrt{\sum_i \left(\varepsilon_{\mathsf{score},t,i}^{(n)}(x_{\tau_{t,i}}^{(n)})\right)^2}\right\}. \tag{108b}$$

B.6 Proof of Lemma 9

We prove Lemma 9 by contradiction. Suppose that there exists $\ell \in [1, \tau(x_T))$ such that $-\log q_\ell(x_\ell) > 2c_6 d \log T$. We let $1 < t \leq \ell$ denote the smallest time step satisfying

$$\theta_t(x_t) = \max\left\{-\frac{\log q_t(x_t)}{d \log T}, c_6\right\} > 2c_6. \tag{109}$$

Since $-\log q_t(x_t) \leq c_6 d \log T$, we have

$$\theta_t(x_t) \geq 2\theta_1(x_1). \tag{110}$$

Morevoer, by repeating similar arguments as in Li et al. (2024c, Eqn. (129)), one can derive

$$\theta_1(x_1), \ldots, \theta_t(x_t) \in [c_6, 4c_6]. \tag{111}$$

By virtue of Lemma 8, for any $2 \leq j \leq t$, one has

$$\begin{aligned}
&\log q_{j-1}(x_{j-1}) - \log q_j(x_j) \\
&\quad = \log \frac{p_{\overline{X}_{\tau_{j,K-1}}}\big(x^{(n)}_{\tau_{j,K-1}}\big)}{p_{\overline{X}_{\tau_{j,0}}}\big(x_{\tau_{j,0}}\big)} \\
&\quad = \log \frac{p_{\sqrt{\frac{1-\tau_{j,0}}{1-\tau_{j,K-1}}}\overline{X}_{\tau_{j,K-1}}}\Big(\sqrt{\frac{1-\tau_{j,0}}{1-\tau_{j,K-1}}}x^{(n)}_{\tau_{j,K-1}}\Big)}{p_{\overline{X}_{\tau_{j,0}}}\big(x_{\tau_{j,0}}\big)} + \log\left[\left(\frac{1-\tau_{j,0}}{1-\tau_{j,K-1}}\right)^{d/2}\right] \\
&\quad \leq \frac{4c_1 d \log T}{T} + C_{10}\left\{\frac{d^2\theta_t^2 \log^4 T}{T^2} + \frac{\sqrt{d\theta_t \log^3 T}}{T}\sqrt{\sum_i \big(\varepsilon^{(n)}_{\mathsf{score},t,i}(x^{(n)}_{\tau_{t,i}})\big)^2}\right\} + \frac{d}{2}\log \alpha_j \\
&\quad \leq \frac{4c_1 d \log T}{T} + C_{10}\left\{\frac{d^2\theta_t^2 \log^4 T}{T^2} + \frac{\sqrt{d\theta_t \log^3 T}}{T}\sqrt{\sum_i \big(\varepsilon^{(n)}_{\mathsf{score},t,i}(x^{(n)}_{\tau_{t,i}})\big)^2}\right\},
\end{aligned} \tag{112}$$

where we makes use of the fact $p_{\overline{X}_{\tau_{j,K-1}}}\big(x^{(n)}_{\tau_{j,K-1}}\big) = \Big(\frac{1-\tau_{j,0}}{1-\tau_{j,K-1}}\Big)^{d/2} p_{\sqrt{\frac{1-\tau_{j,0}}{1-\tau_{j,K-1}}}\overline{X}_{\tau_{j,K-1}}}\Big(\sqrt{\frac{1-\tau_{j,0}}{1-\tau_{j,K-1}}}x^{(n)}_{\tau_{j,K-1}}\Big)$ in the second line. Putting (109), (110), (111) and (112) together leads to

$$\begin{aligned}
c_6 = \theta_1(x_1) \leq \theta_t(x_t) - \theta_1(x_1) &= -\frac{\log q_t(x_t)}{d \log T} - \theta_1(x_1) \leq \frac{-\log q_t(x_t) + \log q_1(x_1)}{d \log T} \\
&= \frac{1}{d \log T}\sum_{j=2}^{t}\big(\log q_{j-1}(x_{j-1}) - \log q_j(x_j)\big) \\
&\leq 4c_1 + C_{11}\left(\frac{d \log^3 T}{T} + \frac{S_{\tau(x_T)-1}(x_T)}{d \log T}\right) \\
&\leq 5c_1.
\end{aligned} \tag{113}$$

Here, the last inequality holds due to (44). This contradicts our assumption $c_6 > 5c_1$. Therefore, we know that (46) holds for all $1 \leq \ell < \tau(x_T)$.

B.7 Proof of Lemma 10

First, we prove that the event $\{S_t(x_T) \leq c_3\}$ implies $\mathcal{E}_t \cap \{\xi_t(x_t) \leq c_3\}$ for all $t < \tau$. We know from Lemma 9 that

$$-\log q_\ell(x_\ell) \leq 2c_6 d \log T, \qquad \forall 1 \leq \ell \leq t. \tag{114}$$

Then Lemma 4 and Lemma 1 (more precisely, (23f)) together imply that for any $\tau \in [\tau_{t,K-1}, \tau_{t,0}]$, one has

$$-\log p_{\overline{X}_\tau}(x^\star_\tau) \le 4c_6 d \log T. \tag{115}$$

Furthermore, it is straightforward to verify that

$$\begin{aligned}
&C_{10}\frac{\theta_{\tau_{t,0}}(x_{\tau_{t,0}})d\log^2 T + \sqrt{\theta_{\tau_{t,0}}(x_{\tau_{t,0}})\sum_{i,n}\big(\varepsilon^{(n)}_{\mathsf{score},t,i}(x^{(n)}_{\tau_{t,i}})\big)^2 d\log^3 T}}{T}\\
&\asymp \frac{d\log^2 T + \sqrt{\sum_{i,n}\big(\varepsilon^{(n)}_{\mathsf{score},t,i}(x^{(n)}_{\tau_{t,i}})\big)^2 d\log^3 T}}{T}\\
&\le \frac{d\log^2 T}{T} + \xi_t(x_t)\\
&\le \frac{d\log^2 T}{T} + S_t(x_T)\\
&\ll 1.
\end{aligned} \tag{116}$$

By virtue of Lemma 8, we know that for all $0 \le i \le K-1, 0 \le n \le N$,

$$-\log p_{\overline{X}_{\tau_{t,i}}}\left(\lambda x^{(n)}_{\tau_{t,i}} + (1-\lambda)x^\star_{\tau_{t,i}}\right) \le 2.1 d\theta_t \log T. \tag{117}$$

In addition, for all $0 \le i \le K-1, 0 \le n \le N$, we have

$$\frac{\varepsilon^{(n)}_{\mathsf{Jacob},t,i}(x^{(n)}_{\tau_{t,i}})\log T}{T} \le \xi_t(x_t) \le S_t(x_T) \le c_3 \ll 1. \tag{118}$$

Combining (115), (117) and (118) together yields

$$\{S_t(x_T) \le c_3\} \subseteq \mathcal{E}_t \cap \{\xi_t(x_t) \le c_3\}. \tag{119}$$

Observing that $S_t(x_T) \le c_3$ for all $t < \tau$, (39) tells us that

$$\frac{p_{Y_{t-1}}(x_{t-1})}{p_{X_{t-1}}(x_{t-1})} = \exp\left(O\left(\xi_t(x_t) + d\left(\frac{d\log^2 T}{T}\right)^{K+1}\right)\right)\frac{p_{Y_t}(x_t)}{p_{X_t}(x_t)}, \qquad \forall 2 \le t < \tau(x_T). \tag{120}$$

As a result, one has

$$\begin{aligned}
\frac{q_1(x_1)}{p_1(x_1)} &= \prod_{t=2}^{\tau-1}\exp\left(O\left(\xi_t(x_t) + d\left(\frac{d\log^2 T}{T}\right)^{K+1}\right)\right)\cdot\frac{q_{\tau-1}(x_{\tau-1})}{p_{\tau-1}(x_{\tau-1})}\\
&= \exp\left(O\left(\sum_{t<\tau}\xi_t(x_t) + d^2\log^2 T\left(\frac{d\log^2 T}{T}\right)^K\right)\right)\frac{q_{\tau-1}(x_{\tau-1})}{p_{\tau-1}(x_{\tau-1})}\\
&= \left(1 + O\left(\sum_{t<\tau}\xi_t(x_t) + d^2\log^2 T\left(\frac{d\log^2 T}{T}\right)^K\right)\right)\frac{q_{\tau-1}(x_{\tau-1})}{p_{\tau-1}(x_{\tau-1})},
\end{aligned} \tag{121}$$

which has finished the proof of (47). Here, the last equation holds since $\sum_{t<\tau}\xi_t(x_t) + d^2\log^2 T\left(\frac{d\log^2 T}{T}\right)^K \le c_3 + d^2\log^2 T\left(\frac{d\log^2 T}{T}\right)^K \ll 1$ and $\exp(z) = 1 + O(z)$ for all $|z| < 1$. Similarly, we can also prove (48).

B.8 Proof of Lemma 11

We can prove Lemma 11 by using same arguments in the proof of Li et al. (2024c, Lemma 8) and we omit the details here for the sake of brevity.