Higher-order Diffusion Sampling via Chebyshev Interpolation and Gauss–Seidel Iterations

Anonymous authors
Paper under double-blind review

Abstract

Higher-order ODE solvers have shown strong empirical promise for accelerating diffusion models through the probability flow ODE, but rigorous non-asymptotic guarantees for such acceleration remain limited. In this paper, we develop a Chebyshev–Gauss–Seidel higher-order sampler and establish a non-asymptotic convergence guarantee that allows the approximation order to grow logarithmically with the number of outer iterations. In the exact-score setting, up to logarithmic factors, the proposed sampler requires at most $d^{1+o_T(1)}\varepsilon^{-1/K_1}$ score function evaluations to approximate the target distribution on $\mathbb{R}^d$ within total variation distance ε, where $o_T(1) \to 0$ as $T \to \infty$ and $K_1 > 0$ is a sufficiently large constant. The analysis assumes only a polynomial second-moment bound on the target distribution, thereby relaxing the bounded-support condition imposed in existing higher-order theory. Moreover, the guarantee is robust to score and Jacobian estimation errors and does not require higher-order smoothness assumptions on the score estimates. Numerical experiments on anisotropic Gaussian mixture benchmarks support the predicted improvement in the accuracy–cost tradeoff under finite score-evaluation budgets.

1 Introduction

1.1 Diffusion models

Diffusion models have emerged as a central paradigm in modern generative modeling, achieving strong empirical performance across image synthesis, text generation, audio generation, and related tasks (Sohl-Dickstein et al., 2015; Song & Ermon, 2019; Ho et al., 2020; Song et al., 2021a;b; Dhariwal & Nichol, 2021; Rombach et al., 2022; Saharia et al., 2022). Two of the most influential diffusion frameworks are denoising diffusion probabilistic models (DDPM) (Ho et al., 2020) and denoising diffusion implicit models (DDIM) (Song et al., 2021a). These methods generate high-quality samples by approximately reversing a progressive noising process. In contrast to alternative generative paradigms, such as generative adversarial networks (GANs) (Goodfellow et al., 2014), variational autoencoders (VAEs) (Kingma & Welling, 2014), and normalizing flows (Rezende & Mohamed, 2015), which typically permit sample generation in a single forward pass or with substantially fewer iterative steps, diffusion samplers require a sequence of reverse-time updates, each of which generally involves evaluating a pretrained neural network for denoising or score estimation.

More specifically, a diffusion model is built on two stochastic processes in $\mathbb{R}^d$. The first is a forward process

$$X_0 \xrightarrow{\text{add noise}} X_1 \xrightarrow{\text{add noise}} \cdots \xrightarrow{\text{add noise}} X_T,$$

which begins with a sample drawn from the target data distribution and gradually transforms it into a noise-like distribution according to a prescribed variance schedule $\{\beta_t\}_{t=1}^T$; see, e.g., (Ho et al., 2020; Song et al., 2021b). When T is sufficiently large, the distribution of X_T is typically close to a standard Gaussian distribution. The goal of diffusion generative modeling is then to construct a reverse process

$$Y_T \xrightarrow{\text{denoise}} Y_{T-1} \xrightarrow{\text{denoise}} \cdots \xrightarrow{\text{denoise}} Y_0,$$

which starts from pure noise and progressively recovers a sample whose distribution is close to that of the target data, ideally so that

$$Y_t \overset{d}{\approx} X_t, \qquad t = T, \dots, 0.$$

The reverse-time dynamics are determined by the score functions of the forward marginals, thereby linking diffusion sampling to score-based generative modeling and reverse-time diffusion theory (Song & Ermon, 2019; Anderson, 1982; Haussmann & Pardoux, 1986; Hyvärinen, 2005; Vincent, 2011). Within this framework, DDPM is commonly viewed as a discretization of the reverse-time stochastic dynamics (Ho et al., 2020), whereas DDIM is closely connected to the deterministic probability flow ODE (Song et al., 2021a;b). Both formulations require repeated evaluation of pretrained score or denoising networks during sampling. Consequently, even after training, generation can remain computationally expensive, making it a central goal of accelerated diffusion sampling to reduce the number of score function evaluations while maintaining sampling accuracy.

1.2 Training-free acceleration and the higher-order theory gap

Training-free acceleration has become a central approach to fast diffusion sampling. Rather than introducing an additional distillation or consistency-training stage, it keeps the pretrained score estimates fixed and improves the accuracy–cost tradeoff through more effective discretizations of the reverse-time dynamics. For probability flow ODE samplers, this naturally leads to higher-order ODE discretizations. Methods such as DPM-Solver (Lu et al., 2022), DEIS (Zhang & Chen, 2023), UniPC (Zhao et al., 2023), and DPM-Solver++ for guided sampling (Lu et al., 2025) have demonstrated substantial empirical speedups while maintaining high sample quality; see also classifier-free guidance (Ho & Salimans, 2022). These empirical successes make higher-order probability flow ODE sampling a compelling target for rigorous non-asymptotic analysis.

The theoretical understanding of such acceleration, however, remains relatively limited. For SDE-based diffusion sampling, polynomial-time guarantees under weak assumptions were first established in (Chen et al., 2023c; Lee et al., 2022; 2023; Benton et al., 2024). For deterministic samplers based on the probability flow ODE, Chen et al. (2023b) gave the first provable convergence guarantee, and Li et al. (2024b) later established a sharp first-order benchmark. Since the present work focuses on training-free higher-order acceleration, we summarize in Table 1 the deterministic probability flow ODE guarantees most relevant to our setting, with all displayed iteration complexities suppressing logarithmic factors. In particular, Runge–Kutta analyses (Huang et al., 2025) improve the dependence on ε, but require compact support and higher-order smoothness assumptions. Meanwhile, Li et al. (2025) proves acceleration under comparatively weak distributional and score-estimation assumptions, but still treats the approximation order K as a constant and assumes bounded support of the target distribution. This suggests that the current complexity theory for higher-order probability flow ODE sampling remains improvable under weak assumptions, which is precisely the motivation for this paper.

Table 1: Comparison of deterministic probability flow ODE sampling guarantees in total variation. Here s_τ and ∇s_τ denote the estimated score function and its Jacobian, while $s^\star_\tau$ and $\nabla s^\star_\tau$ denote their exact counterparts, D represents the radius of the data support and L bounds certain higher-order derivatives of the score estimates.

Paper	Target Assumption	Score / Regularity Assumption	Iteration Complexity	Higher-order Solver
Li et al. (2024b)	$\mathbb{P}(\|X_0\|_2 \le T^{c_R}) = 1$	$s_\tau \approx s^\star_\tau,\ \nabla s_\tau \approx \nabla s^\star_\tau$	$\max\{d^2, d/\varepsilon\}$	×
Huang et al. (2025)	$\mathbb{P}(\|X_0\|_2 \le D) = 1$	$s_\tau \approx s^\star_\tau, \quad s_\tau \in C^{p+1}$	$(LDd)^{1+1/p}\varepsilon^{-1/p}$	✓ p-th order
Li et al. (2025)	$\mathbb{P}(\|X_0\|_2 \le T^{c_R}) = 1$	$s_\tau \approx s^\star_\tau,\ \nabla s_\tau \approx \nabla s^\star_\tau$	$\max\{d^2,\ d^{1+2/K}\varepsilon^{-1/K}\}$	✓ K-th order (fixed K)
this work	$\mathbb{E}\|X_0\|_2^2 \le T^{2c_R}$	$s_\tau \approx s^\star_\tau,\ \nabla s_\tau \approx \nabla s^\star_\tau$	$d^{1+o_T(1)}\varepsilon^{-1/K_1}$	✓ K-th order ($K \asymp \log T$)

1.3 Main contributions

As noted above, existing complexity guarantees for higher-order probability flow ODE sampling rely on two key restrictions: the interpolation order K on each reverse-time interval is treated as a fixed constant, and the target distribution is assumed to satisfy the bounded-support condition $\mathbb{P}(\|X_0\|_2 \leq T^{c_R}) = 1$. Under these assumptions, the iteration complexity required to achieve ε-accuracy in total variation scales as $d^{1+2/K}\varepsilon^{-1/K}$. In this paper, we show that this complexity can be further improved under the substantially weaker moment condition $\mathbb{E}\|X_0\|_2^2 \leq T^{2c_R}$. Specifically, by developing a Chebyshev–Gauss–Seidel higher-order sampler within the probability flow ODE framework of Li et al. (2025), we obtain the improved complexity bound $d^{1+o_T(1)}\varepsilon^{-1/K_1}$, where K_1 is a sufficiently large constant.

Our approach has two main ingredients. First, we replace the equi-spaced nodes used in Li et al. (2025) with Chebyshev–Lobatto nodes, which allow the interpolation order K to grow as $\log T$ while preserving control of error propagation. Second, we replace the Jacobi iteration in Li et al. (2025) with a Gauss–Seidel-type refinement scheme to approximate the probability flow ODE solution more effectively. As a result, the number of score function evaluations required to achieve a prescribed total variation accuracy is reduced to

$$d^{1+o_T(1)}\varepsilon^{-1/K_1},$$

up to logarithmic factors, where $o_T(1) \to 0$ as $T \to \infty$.

In addition, our theory is robust to inexact score estimation: it requires only score and Jacobian accuracy along the iterates and does not impose higher-order smoothness assumptions on the learned score. Finally, numerical experiments on anisotropic Gaussian mixture benchmarks support the predicted improvement in the accuracy–cost tradeoff.

1.4 Other related work

Beyond the probability flow ODE literature discussed above, non-asymptotic convergence theory for score-based diffusion sampling has been developed for DDPM-type samplers and reverse-SDE methods under increasingly general assumptions on the data distribution and score estimation error. These works establish guarantees in total variation, Kullback–Leibler divergence, and Wasserstein distance (Lee et al., 2022; 2023; Chen et al., 2023c;a; Benton et al., 2024; Bruno et al., 2025; Conforti et al., 2025; Gao et al., 2025).

For probability flow ODE samplers, recent studies have established polynomial-time convergence with a corrector step, nearly dimension-linear convergence for discrete-time implementations, prediction–correction guarantees, Wasserstein convergence, and adaptation to intrinsic low-dimensional structure (Chen et al., 2023b; Li et al., 2024b; Pedrotti et al., 2024; Gao & Zhu, 2025; Tang & Yan, 2025; Beyler & Bach, 2025). Provable acceleration has further been investigated through training-free accelerated DDIM/DDPM samplers, higher-order Runge–Kutta discretizations of the probability flow ODE, higher-order Lagrange interpolation with successive refinement, operator-splitting samplers, and first-order forward-value evaluation (Li et al., 2024a; Huang et al., 2025; Li et al., 2025; Liu et al., 2026; Jiao et al., 2025). Recent work has also established high-accuracy diffusion sampling guarantees with polylogarithmic dependence on the inverse accuracy under a variety of algorithmic and structural assumptions (Gatmiry et al., 2026; Chen et al., 2026).

1.5 Notations

Throughout this paper, we write $f = O(g)$ or $f \lesssim g$ if $|f| \leq C|g|$ for some universal constant $C > 0$. We write $f \asymp g$ if both $f \lesssim g$ and $g \lesssim f$ hold. The notation $\widetilde{O}(\cdot)$ suppresses logarithmic factors in the relevant parameters, and $o_T(1)$ denotes a quantity that tends to zero as $T \to \infty$. For any two probability measures P and Q, their total variation distance is defined by $\mathrm{TV}(P, Q) := \frac{1}{2}\int |\mathrm{d}P - \mathrm{d}Q|$. In addition, $p_X(\cdot)$ and $p_{X|Y}(\cdot \mid \cdot)$ denote the probability density functions of X and of X conditional on Y, respectively. For any matrix A, we use $\|A\|$ and $\|A\|_{\mathrm{F}}$ to denote its spectral norm and Frobenius norm, respectively. Finally, for any vector-valued function f, we let $\frac{\partial f}{\partial x}$ denote the Jacobian matrix of f.

1.6 Organization

The remainder of this paper is organized as follows. Section 2 presents the preliminaries on diffusion models and the probability flow ODE, and Section 3 introduces the proposed Chebyshev–Gauss–Seidel higher-order sampler. Section 4 states the main result, and Section 5 contains its proof. Section 6 reports the numerical experiments, while Section 7 concludes the paper with further discussion. The appendix collects the auxiliary lemmas used in the analysis, as well as the detailed proofs of the lemmas used in Section 5.

2 Preliminaries

In this section, we review the basic framework of diffusion generative models based on (Stein) score functions. The goal of a generative model is to produce samples whose distribution is close to an unknown target distribution p_{data} on $\mathbb{R}^d$, given access to data drawn from p_{data}. A diffusion generative model typically involves two stochastic processes: a forward process and a reverse process, which we describe below.

Forward process. Starting from an initial sample X_0 drawn from the target distribution p_{data} on $\mathbb{R}^d$, the forward process is defined by

$$X_t = \sqrt{\alpha_t}\, X_{t-1} + \sqrt{1-\alpha_t}\, W_t, \qquad t = 1, \dots, T, \tag{1}$$

where $0 < \alpha_t < 1$ are prescribed noise-scheduling parameters, and $\{W_t\}_{t=1}^T$ is a sequence of independent standard Gaussian random vectors in $\mathbb{R}^d$, that is,

$$W_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_d).$$

Define

$$\bar{\alpha}_t := \prod_{k=1}^{t} \alpha_k, \qquad t = 1, \dots, T. \tag{2}$$

Then, for each $t = 1, \dots, T$, it is straightforward to verify that

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

In particular, when $\bar{\alpha}_T$ is sufficiently small, the distribution of X_T is close to $\mathcal{N}(0, I_d)$ for a broad class of data distributions:

$$\mathrm{Law}(X_T) \approx \mathcal{N}(0, I_d). \tag{4}$$

Diffusion models are closely connected to the framework of stochastic differential equations (SDEs), and it is therefore useful to introduce a continuous-time version $(\bar{X}_\tau)_{\tau \in [0,1]}$ of the forward diffusion process. Specifically, consider the SDE

$$\mathrm{d}\bar{X}_\tau = -\frac{1}{2(1-\tau)} \bar{X}_\tau \,\mathrm{d}\tau + \frac{1}{\sqrt{1-\tau}}\, \mathrm{d}B_\tau, \qquad \bar{X}_0 \sim p_{\mathrm{data}}, \qquad 0 \le \tau < 1, \tag{5}$$

where B_τ is a standard Brownian motion in $\mathbb{R}^d$. One can verify that the solution to this SDE satisfies

$$\bar{X}_\tau = \sqrt{1-\tau}\, X_0 + \sqrt{\tau}\, Z, \qquad X_0 \sim p_{\mathrm{data}}, \quad Z \sim \mathcal{N}(0, I_d), \tag{6}$$

where X_0 and Z are independent. In particular,

$$\bar{X}_{1-\bar{\alpha}_t} \overset{\mathrm{d}}{=} X_t, \qquad t = 1, \dots, T. \tag{7}$$

Reverse process. A central goal of diffusion models is to construct a time-reversed process whose marginal distributions coincide with, or closely approximate, those of the forward process. More precisely, we seek a reverse-time process

$$Y_T \to Y_{T-1} \to \cdots \to Y_1$$

such that

$$Y_t \overset{\mathrm{d}}{\approx} X_t, \qquad t = 1, \dots, T.$$

Using the relation between X_t and $\bar{X}_\tau$ in (7), we can design a discrete reverse-time process $\{Y_t\}_{t=1}^T$ through the time-reversal of the continuous-time process $\bar{X}_\tau$. Specifically, for the forward process (5), classical results on reverse-time SDEs (Song et al., 2021b) show that the corresponding probability flow ODE is given as follows: for any $\tau_0 \in (0,1)$, the process $\{Y_\tau^{\mathrm{ode}}\}_{\tau\in[0,\tau_0)}$ satisfies

$$\mathrm{d}Y_\tau^{\mathrm{ode}} = -\frac{1}{2(1-\tau)}\left(Y_\tau^{\mathrm{ode}} + \nabla \log p_{\bar{X}_\tau}\left(Y_\tau^{\mathrm{ode}}\right)\right)\mathrm{d}\tau, \qquad Y_{\tau_0}^{\mathrm{ode}} \sim \bar{X}_{\tau_0}, \qquad 0 \le \tau < \tau_0,$$

or equivalently,

$$\mathrm{d}\left(\frac{Y_\tau^{\mathrm{ode}}}{\sqrt{1-\tau}}\right) = -\frac{1}{2(1-\tau)^{3/2}}\nabla \log p_{\bar{X}_\tau}\left(Y_\tau^{\mathrm{ode}}\right)\mathrm{d}\tau, \qquad Y_{\tau_0}^{\mathrm{ode}} \sim \bar{X}_{\tau_0}, \qquad 0 \le \tau < \tau_0, \tag{8}$$

and has the same marginal distributions as $\{\bar{X}_\tau\}_{\tau\in[0,\tau_0)}$.

Importantly, the reverse process depends only on the gradient of the log-density $\nabla \log p_{\bar{X}_\tau}$, known as the *score function* of $p_{\bar{X}_\tau}$. For the continuous-time forward process $\{\bar{X}_\tau\}_{\tau\in[0,1]}$, the exact score function at time $\tau \in (0,1]$ is defined by

$$s_\tau^\star(x) := \nabla \log p_{\bar{X}_\tau}(x), \qquad x \in \mathbb{R}^d. \tag{9}$$

If these score functions were available, one could initialize $Y_T \sim \mathcal{N}(0, I_d)$ in view of (4), and then generate samples by numerically solving the probability flow ODE (8). In practice, however, the true score functions are unknown and must be learned from data, typically using neural networks trained on samples from the target distribution (Ho et al., 2020; Song et al., 2021a). Consequently, the practical reverse process $Y_T \to Y_{T-1} \to \cdots \to Y_1$ is obtained by approximately solving (8) using estimated score functions, giving rise to a broad class of diffusion sampling algorithms (Chen et al., 2023b; Li & Cai, 2024; Li et al., 2024b; 2025).

3 Gauss–Seidel Refinement with Higher-Order Diffusion Sampling

In this section, we present the main ideas and the detailed procedure of our proposed algorithm. At a high level, the method is a Gauss–Seidel-type higher-order iterative scheme for approximately solving the ODE (8). To motivate the construction, recall that $X_t \overset{\mathrm{d}}{=} \bar{X}_{1-\bar{\alpha}_t}$. According to the probability flow ODE (8), the ideal reverse update from Y_t to Y_{t-1} is given by $Y_{t-1} = Y_{1-\bar{\alpha}_{t-1}}^{\mathrm{ode}}$, where Y_τ^{ode} denotes the solution to (8) with terminal condition $Y_{1-\bar{\alpha}_t}^{\mathrm{ode}} = Y_t$. In particular, we have

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

Directly implementing this update is, however, computationally prohibitive, since the integral depends on a continuum of score evaluations, whereas in practice we only have access to a finite collection of estimated score functions. To address this issue, on each interval $\left[1-\bar{\alpha}_t,\, 1-\bar{\alpha}_{t-1}\right]$ we approximate $s_\tau^\star\left(Y_\tau^{\mathrm{ode}}\right)$ by polynomial interpolation. In this paper, we employ a higher-order approximation based on K score evaluations at Chebyshev–Lobatto nodes, where $K \ge 2$. Specifically, for the interval $\left[1-\bar{\alpha}_t,\, 1-\bar{\alpha}_{t-1}\right]$, define

$$\tau_{t,0} = 1-\bar{\alpha}_t, \qquad \tau_{t,K-1} = 1-\bar{\alpha}_{t-1}, \tag{11}$$

and choose the Chebyshev–Lobatto nodes by

$$\tau_{t,j} = \frac{\tau_{t,0}+\tau_{t,K-1}}{2} + \frac{\tau_{t,0}-\tau_{t,K-1}}{2}\cos\left(\frac{j\pi}{K-1}\right), \qquad 0 \le j \le K-1. \tag{12}$$

Then

$$\tau_{t,K-1} < \tau_{t,K-2} < \cdots < \tau_{t,0}.$$

Given the collection of points

$$\left(\tau_{t,j},\, (1-\tau_{t,j})^{-3/2} s_{\tau_{t,j}}^\star\left(Y_{\tau_{t,j}}^{\mathrm{ode}}\right)\right), \qquad 0 \le j \le K-1,$$

we approximate $(1-\tau)^{-3/2}s^{\star}_{\tau}\big(Y^{\mathrm{ode}}_{\tau}\big)$ by the degree-$(K-1)$ Lagrange interpolating polynomial through these K points, namely,

$$\frac{1}{(1-\tau)^{3/2}}s^{\star}_{\tau}\big(Y^{\mathrm{ode}}_{\tau}\big) \approx \sum_{j=0}^{K-1}\psi_{t,j}(\tau)\frac{s^{\star}_{\tau_{t,j}}\Big(Y^{\mathrm{ode}}_{\tau_{t,j}}\Big)}{(1-\tau_{t,j})^{3/2}}, \qquad \forall \tau \in [\tau_{t,K-1},\tau_{t,0}], \tag{13}$$

where $\psi_{t,j}(\tau)$ are the Lagrange basis polynomials defined by

$$\psi_{t,j}(\tau) := \frac{\prod_{j'\neq j}(\tau-\tau_{t,j'})}{\prod_{j'\neq j}(\tau_{t,j}-\tau_{t,j'})}, \qquad 0\le j\le K-1. \tag{14}$$

A closer inspection shows that the approximation in (13) still cannot be used directly, because the values $\{Y^{\mathrm{ode}}_{\tau_{t,j}}\}_{j=0}^{K-1}$ are generally unavailable. To overcome this difficulty, we propose a *Gauss–Seidel refinement* procedure that alternates between estimating $\{Y^{\mathrm{ode}}_{\tau_{t,j}}\}_{j=0}^{K-1}$ and approximating $(1-\tau)^{-3/2}s^{\star}_{\tau}\big(Y^{\mathrm{ode}}_{\tau}\big)$. More precisely, starting from an initial sequence $\{x^{(0)}_{\tau_{t,j}}\}_{j=0}^{K-1}$ as an approximation to $\{Y^{\mathrm{ode}}_{\tau_{t,j}}\}_{j=0}^{K-1}$, the Gauss–Seidel scheme updates this estimate to $\{x^{(n+1)}_{\tau_{t,j}}\}_{j=0}^{K-1}$ at iteration $n=0,1,\dots,N-1$ as follows: for each $j=1,\dots,K-1$,

$$\frac{x^{(n+1)}_{\tau_{t,j}}}{\sqrt{1-\tau_{t,j}}} = \frac{x^{(0)}_{\tau_{t,0}}}{\sqrt{1-\tau_{t,0}}} + \sum_{k=0}^{j-1}\frac{\gamma_{t,k}(\tau_{t,j})}{2(1-\tau_{t,k})^{3/2}}s_{\tau_{t,k}}\left(x^{(n+1)}_{\tau_{t,k}}\right) + \sum_{k=j}^{K-1}\frac{\gamma_{t,k}(\tau_{t,j})}{2(1-\tau_{t,k})^{3/2}}s_{\tau_{t,k}}\left(x^{(n)}_{\tau_{t,k}}\right), \tag{15}$$

where $x^{(n)}_{\tau_{t,0}} = x^{(0)}_{\tau_{t,0}}$ for all $n=1,\dots,N$, $s_{\tau_{t,k}}$ is the estimated score function, and

$$\gamma_{t,k}(\tau_{t,j}) := \int_{\tau_{t,j}}^{\tau_{t,0}}\psi_{t,k}(\tau)\,\mathrm{d}\tau. \tag{16}$$

We now summarize our higher-order method for diffusion models; the detailed procedure is given in Algorithm 1.

(i) **Initialization.** Sample $Y_T\sim\mathcal{N}(0,I_d)$, and set $x^{(0)}_{\tau_{T,j}} := Y_T$ for all $j=0,\dots,K-1$, where $\tau_{T,j}$ are defined in (12).

(ii) **Iterative update rule.** For $t=T,T-1,\dots,2$, compute $\{x^{(n+1)}_{\tau_{t,j}}\}_{j=0}^{K-1}$ according to the update rule (15) for $n=0,1,\dots,N-1$. After N iterations at step t, set $Y_{t-1} := x^{(N)}_{\tau_{t,K-1}}$ and initialize the next step by setting $x^{(0)}_{\tau_{t-1,j}} := Y_{t-1}$ for all $j=0,\dots,K-1$.

Algorithm 1 Gauss–Seidel Refinement with Higher-Order Diffusion Sampling

Require: T,K,N, interpolation nodes $\{\tau_{t,i}\}$, coefficients $\{\gamma_{t,j}(\tau_{t,i})\}$, score estimator $s_\tau(\cdot)$
Ensure: Y_1

```
1: Sample Y_T ~ N(0, I_d)
2: for t = T, T−1, ..., 2 do
3:     Set x^{(0)}_{τ_{t,j}} ← Y_t for all j = 0, ..., K−1
4:     for n = 0, ..., N−1 do
5:         Compute x^{(n+1)}_{τ_{t,j}} via (15) sequentially for j = 1, ..., K−1
6:     end for
7:     Set Y_{t−1} ← x^{(N)}_{τ_{t,K−1}}
8: end for
9: return Y_1
```

Algorithm 1 shows that, once the initial value Y_T is sampled, all intermediate iterates $\{x^{(n)}_{\tau_{t,i}}\}$ are fully determined by the prescribed update rule together with the available score estimator. The total number of score function evaluations is $O(TK(N+1))$. Since, as shown later, $K\le c\log T$ and $N=\lceil C_3K\log T\rceil$, it follows that the overall number of score function evaluations is $\widetilde{O}(T)$.

4 Main Results

In this section, we establish convergence guarantees for Algorithm 1. Throughout, let $q_t := \text{distribution}\,(X_t)$ denote the target distribution at time t, where X_t is the forward process initialized from the data distribution, and let $p_t := \text{distribution}\,(Y_t)$ denote the distribution at time t generated by our algorithm. We begin by introducing several assumptions on the learning-rate schedule, the target distribution, and the estimated score functions.

4.1 Assumptions

For the forward process given in (1), the learning rates $\{\alpha_t\}_{t=1}^T$ are chosen as follows:

$$\beta_1 = 1 - \alpha_1 = T^{-c_0}, \tag{17}$$

and for $t \geq 2$,

$$\beta_t = 1 - \alpha_t = \frac{c_1 \log T}{T} \min\left\{\beta_1\left(1 + \frac{c_1 \log T}{T}\right)^t, 1\right\}, \tag{18}$$

where $c_0, c_1 > 0$ are sufficiently large numerical constants. This choice of learning-rate schedule is standard in recent diffusion theory; see, for example, (Benton et al., 2024; Li & Yan, 2025; Li et al., 2025; Huang et al., 2026). Next, we impose a mild assumption on the target distribution p_{data}.

Assumption 1 (Target data distribution)**.** *The target distribution p_{data} has a bounded second moment, namely,*

$$\mathbb{E}_{X_0 \sim p_{\text{data}}}\left[\|X_0\|_2^2\right] \leq T^{2c_R}$$

for some sufficiently large constant $c_R > 0$.

Remark 1. *Assumption 1 requires only that the second moment of the target distribution grow at most polynomially with the iteration number T. Since T itself typically scales polynomially with the dimension d, this condition allows the second moment to remain fairly large. In contrast, the closely related work (Li et al., 2025) assumes the bounded-support condition*

$$\mathbb{P}\big(\|X_0\|_2 \leq T^{c_R}\big) = 1,$$

which is strictly stronger than Assumption 1. Furthermore, our assumption is more consistent with standard conditions in sampling theory and more plausible for empirical data encountered in practice.

As described in Section 3, the true score function appearing in the integral is approximated by a finite collection of estimated score functions. Accordingly, the convergence behavior of our proposed algorithm is intrinsically tied to the accuracy of these score estimates.

Assumption 2 (Score accuracy)**.** *Assume that the score estimates satisfy*

$$\frac{1}{T(N+1)K} \sum_{t=1}^{T} \mathbb{E}_{Y_t \sim q_t}\left[\varepsilon_{\text{score},t}^2(Y_t)\right] \leq \varepsilon_{\text{score}}^2,$$

where

$$\varepsilon_{\text{score},t}^2(Y_t) := \sum_{i=0}^{K-1} \sum_{n=0}^{N} \left\| s_{\tau_{t,i}}\left(x_{\tau_{t,i}}^{(n)}\right) - s_{\tau_{t,i}}^\star\left(x_{\tau_{t,i}}^{(n)}\right)\right\|_2^2$$

and $x_{\tau_{t,i}}^{(n)}$ are the points generated by (15), with $x_{\tau_{t,0}}^{(n)} = Y_t$ for all $n = 0, 1, \ldots, N$.

For ODE-based samplers, an ℓ_2 bound on the score estimation error alone is generally insufficient for convergence analysis. We therefore impose an additional assumption on the Jacobian estimation error.

Assumption 3 (Jacobian accuracy)**.** *Assume that, for all $1 \le t \le T$ and $0 \le i \le K-1$, each score function $s_{\tau_{t,i}}(\cdot)$ is continuously differentiable. Moreover, assume that*

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

where

$$\varepsilon^2_{\mathrm{Jacobi},t}(Y_t) := \sum_{i=0}^{K-1}\sum_{n=0}^{N}\left\| \frac{\partial s_{\tau_{t,i}}\left(x^{(n)}_{\tau_{t,i}}\right)}{\partial x} - \frac{\partial s^{\star}_{\tau_{t,i}}\left(x^{(n)}_{\tau_{t,i}}\right)}{\partial x}\right\|^2$$

and $x^{(n)}_{\tau_{t,i}}$ denotes the sequence of points generated by (15)*, with $x^{(n)}_{\tau_{t,0}} = Y_t$ for every $n = 0, 1, \dots, N$.*

4.2 Convergence guarantees

We now state the main convergence result for Algorithm 1.

Theorem 1. *Suppose that Assumptions 1, 2, 3 hold. Then there exist sufficiently large constants $C_2, C_3, K_1, M_0 > 1$ and a sufficiently small constant $c > 0$ such that, whenever*

$$T \ge C_2 d\log^4 T, \qquad N = \lceil C_3 K\log T\rceil, \qquad 2 \le K \le c\log T,$$

the output of Algorithm 1 satisfies

$$\mathrm{TV}(q_1, p_1) \lesssim d^2 K\log^2 T\left(\frac{M_0 K d\log^2 T}{T}\right)^K + d\log^{5/2}T\sqrt{\log K}\,\varepsilon_{\mathrm{Jacobi}} + d^{3/2}\log^{7/2}T\log K\,\varepsilon_{\mathrm{score}} + T^{-K_1}.$$

Remark 2. *Theorem 1 shows that the total variation error consists of four components: the discretization error arising from the higher-order approximation of the continuous reverse process, the Jacobian estimation error, the score estimation error, and the initialization error. We next compare our result with the most closely related existing work.*

1. ***Iteration complexity.*** *Given the Jacobian estimation error $\varepsilon_{\mathrm{Jacobi}}$ and the score estimation error $\varepsilon_{\mathrm{score}}$, for any target accuracy $\varepsilon \ge d\log^{5/2}T\sqrt{\log K}\,\varepsilon_{\mathrm{Jacobi}} + d^{3/2}\log^{7/2}T\log K\,\varepsilon_{\mathrm{score}}$, Algorithm 1 achieves $\mathrm{TV}(q_1, p_1) \lesssim \varepsilon$ within*

$$T = \widetilde{O}\left(\max\left\{d^{1+2/K}\varepsilon^{-1/K}, \varepsilon^{-1/K_1}\right\}\right). \tag{19}$$

   In particular, our analysis allows $K \asymp \log T$. In this regime, $d^{1+2/K} = d^{1+o_T(1)}$, and $\varepsilon^{-1/K} = \varepsilon^{-o_T(1)}$, so that T can be small as

$$\widetilde{O}\big(d^{1+o_T(1)}\varepsilon^{-1/K_1}\big)$$

   which significantly improves upon the best known result $\widetilde{O}\left(\max\left\{d^2,\, d^{1+2/K}\varepsilon^{-1/K}\right\}\right)$ in Li et al. (2025), where K is treated as a fixed constant..

2. ***Mild target distribution and score assumptions.*** *Our convergence guarantee requires only a finite second-moment condition on the target distribution, and therefore applies to a broad class of data distributions. In particular, it avoids the bounded-support assumption $\mathbb{P}(\|X_0\|_2 \le T^{c_R}) = 1$ imposed in Li et al. (2025), as well as other restrictive assumptions such as smoothness or log-concavity. Moreover, compared to other higher-order ODE-based analyses (e.g., (Huang et al., 2025)), our framework does not require bounded higher-order derivatives or higher-order Lipschitz continuity of the score function. Instead, it relies only on first-order accuracy of the score and its Jacobian along the sampler trajectory.*

Remark 3. *We note that the bounds on the Jacobian estimation error and the score estimation error in Theorem 1 are larger than those reported in Li et al. (2025). This difference arises because our analysis only requires $T \gtrsim d\log^4 T$, whereas Li et al. (2025) assumes the stronger condition $T \gtrsim d^2\log^3 T$. In fact, if one imposes the stronger requirement $T \gtrsim d^2\log^4 T$, our bounds on both the Jacobian and score estimation errors match those in Li et al. (2025).*

5 Proof of the main result

In this section, we present the proofs of the main results. Throughout, we assume that

$$\varepsilon_{\text{score}} \lesssim (d^{3/2}\log^{9/2} T)^{-1} \quad \text{and} \quad \varepsilon_{\text{Jacobi}} \lesssim (d\log^3 T)^{-1}. \tag{20}$$

Otherwise, Theorem 1 follows immediately, since the total variation distance is always bounded above by 1. Recall from (4) that $\mathrm{Law}(X_T)$ is close to $\mathcal{N}(0, I_d)$, and that Algorithm 1 is initialized with $Y_T \sim \mathcal{N}(0, I_d)$. The following lemma shows that the distributions of X_T and Y_T are indeed close in total variation distance, with an error that can be controlled explicitly.

Lemma 1 (Li & Cai (2024), Lemma 2)**.** *Suppose that T is sufficiently large, and let $K_1 > 0$ be any fixed constant. Then*

$$\big(\mathrm{TV}(q_T, p_T)\big)^2 \le \frac{1}{2}\,\mathrm{KL}(q_T \| p_T) \lesssim \frac{1}{T^{2K_1}}. \tag{21}$$

5.1 Main steps for proving Theorem 1

Before proceeding, we introduce some notation. Fix an initial value $Y_T \in \mathbb{R}^d$ in Algorithm 1, and for each $1 \le t \le T$, define

$$x_t := x^{(0)}_{\tau_{t,0}},$$

where $x^{(0)}_{\tau_{t,0}}$ is generated by Algorithm 1. On each interval $\big[1-\bar{\alpha}_t,\, 1-\bar{\alpha}_{t-1}\big]$, let $\{x^\star_\tau\}_{\tau\in[\tau_{t,K-1},\tau_{t,0}]}$ denote the exact probability flow ODE trajectory associated with (8), with $\tau_0 := 1-\bar{\alpha}_t$ and initial condition $Y^{\text{ode}}_{\tau_0} := x_t$, and set

$$x^\star_{t-1} := x^\star_{\tau_{t,K-1}}.$$

Likewise, let $Y^\star_{t-1}$ denote the solution to (8) at $\tau = 1-\bar{\alpha}_{t-1}$ with $\tau_0 := 1-\bar{\alpha}_t$ and initial condition $Y^{\text{ode}}_{\tau_0} := Y_t$. With this notation, we now outline the proof strategy, which consists of several steps.

Step 1: controlling the density ratio. With Lemma 1 in hand, the basic idea to establish the convergence rate is to connect the density ratio $\frac{p_{X_{t-1}}(x_{t-1})}{p_{Y_{t-1}}(x_{t-1})}$ at the $(t-1)$-th step with its counterpart part $\frac{p_{X_t}(x_t)}{p_{Y_t}(x_t)}$ at the t-th step. To this end, we begin with the identity

$$\begin{aligned}\frac{p_{X_{t-1}}(x_{t-1})}{p_{Y_{t-1}}(x_{t-1})} &= \frac{p_{\sqrt{\alpha_t}X_{t-1}}(\sqrt{\alpha_t}x_{t-1})}{p_{\sqrt{\alpha_t}X_{t-1}}(\sqrt{\alpha_t}x^\star_{t-1})} \cdot \frac{p_{\sqrt{\alpha_t}X_{t-1}}(\sqrt{\alpha_t}x^\star_{t-1})}{p_{X_t}(x_t)} \\ &\quad \cdot \left[\frac{p_{\sqrt{\alpha_t}Y_{t-1}}(\sqrt{\alpha_t}x_{t-1})}{p_{\sqrt{\alpha_t}Y^\star_{t-1}}(\sqrt{\alpha_t}x^\star_{t-1})} \cdot \frac{p_{\sqrt{\alpha_t}Y^\star_{t-1}}(\sqrt{\alpha_t}x^\star_{t-1})}{p_{Y_t}(x_t)}\right]^{-1} \cdot \frac{p_{X_t}(x_t)}{p_{Y_t}(x_t)}.\end{aligned} \tag{22}$$

Since $Y^\star_{t-1}$ is obtained by transporting Y_t through the exact probability-flow ODE (10) on $[\tau_{t,K-1}, \tau_{t,0}]$, it implies

$$\frac{p_{\sqrt{\alpha_t}Y^\star_{t-1}}(\sqrt{\alpha_t}x^\star_{t-1})}{p_{Y_t}(x_t)} = \frac{p_{\sqrt{\alpha_t}X_{t-1}}(\sqrt{\alpha_t}x^\star_{t-1})}{p_{X_t}(x_t)}. \tag{23}$$

Substituting (23) into (22) gives

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

It has been shown by Li et al. (2025, Lemma 6) that the right hand side can be well controlled by the distance between x_{t-1} and $x^\star_{t-1}$ and their Jacobian matrices, as stated below.

Lemma 2 (Li et al. (2025), Lemma 6)**.** *For each $t = 2, \ldots, T$, it holds*

(a)

$$\frac{p_{\sqrt{\alpha_t}X_{t-1}}(\sqrt{\alpha_t}\,x_{t-1})}{p_{\sqrt{\alpha_t}X_{t-1}}(\sqrt{\alpha_t}\,x^\star_{t-1})} = \exp\left(O\left(\frac{\|x_{t-1}-x^\star_{t-1}\|_2^2}{1-\bar{\alpha}_{t-1}} + \sqrt{\frac{d\log T}{1-\bar{\alpha}_{t-1}}}\|x_{t-1}-x^\star_{t-1}\|_2\right)\right). \tag{24}$$

(b) Assume that

$$\|(J^{(N)}_{\tau_{t,K-1}})^{-1}\| \lesssim 1, \tag{25}$$

then

$$\frac{p_{\sqrt{\alpha_t}Y^\star_{t-1}}\left(\sqrt{\alpha_t}\, x^\star_{t-1}\right)}{p_{\sqrt{\alpha_t}Y_{t-1}}\left(\sqrt{\alpha_t}\, x_{t-1}\right)} = \exp\Big(O\Big(d\,\|J^{(N)}_{\tau_{t,K-1}} - J^\star_{\tau_{t,K-1}}\|\Big)\Big). \tag{26}$$

Here, $J^{(n)}_{\tau_{t,i}} := \frac{\partial(x^{(n)}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}})}{\partial(x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}})}$, $J^\star_{\tau_{t,i}} := \frac{\partial(x^\star_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}})}{\partial(x^\star_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}})}$ *and we regard* $x^{(n)}_{\tau_{t,i}}$ *and* $x^\star_{\tau_{t,i}}$ *as functions of* $x_{\tau_{t,0}}$*, since the Gauss–Seidel iterates and the exact ODE trajectory are determined by the rescaled initial value* $x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}}$*.*

In view of Lemma 2, we obtain

$$\frac{p_{X_{t-1}}(x_{t-1})}{p_{Y_{t-1}}(x_{t-1})} = \exp\left(O\left(\frac{\|x_{t-1}-x^\star_{t-1}\|_2^2}{1-\bar\alpha_{t-1}} + \sqrt{\frac{d\log T}{1-\bar\alpha_{t-1}}}\|x_{t-1}-x^\star_{t-1}\|_2 + d\,\|J^{(N)}_{\tau_{t,K-1}} - J^\star_{\tau_{t,K-1}}\|\right)\right)\frac{p_{X_t}(x_t)}{p_{Y_t}(x_t)}. \tag{27}$$

By Algorithm 1, we know $x_{t-1} = x^{(0)}_{\tau_{t-1,0}} = x^{(N)}_{\tau_{t,K-1}}$. Therefore, it suffices to control

$$\|x^{(N)}_{\tau_{t,K-1}} - x^\star_{\tau_{t,K-1}}\|_2 \qquad \text{and} \qquad \|J^{(N)}_{\tau_{t,K-1}} - J^\star_{\tau_{t,K-1}}\|.$$

A key observation is that, for the probability ODE flow, once Y_T in Algorithm 1 is fixed, all quantities $x^{(n)}_{\tau_{t,i}}$, $x^\star_{\tau_{t,i}}$, and $x^\star_\tau$ are fully determined. In the initialization stage, we take $Y_T \sim \mathcal{N}(0, I_d)$. We will show that for "typical" points $Y_T \in \mathbb{R}^d$, the above two quantities are sufficiently small. We say that a point $Y_T \in \mathbb{R}^d$ is *typical* if

$$Y_T \in E_t := E_{1,t} \cap E_{2,t} \cap E_{3,t}, \tag{28}$$

where the events $E_{1,t}$, $E_{2,t}$, and $E_{3,t}$ are defined as follows:

$$E_{1,t} := \left\{ Y_T : \frac{d\sqrt{\log K}\log T}{T}\sqrt{\sum_{i=0}^{K-1}\sum_{n=0}^{N}\left(\varepsilon^{(n)}_{\text{Jacobi},t,i}(x^{(n)}_{\tau_{t,i}})\right)^2} + \frac{d^{3/2}\log K\log^2 T}{T}\sqrt{\sum_{i=0}^{K-1}\sum_{n=0}^{N}\left(\varepsilon^{(n)}_{\text{score},t,i}(x^{(n)}_{\tau_{t,i}})\right)^2} \le c_3 \right\}$$

$$E_{2,t} := \left\{ Y_T \in \mathbb{R}^d : -\log p_{\bar X_{\tau_{t,i}}}\left(\lambda x^{(n)}_{\tau_{t,i}} + (1-\lambda)x^\star_{\tau_{t,i}}\right) \le C_4 d\log T, \text{for all } 0 \le i \le K-1,\ 0 \le n \le N,\ \lambda \in [0,1] \right\}$$

$$E_{3,t} := \left\{ Y_T \in \mathbb{R}^d : -\log p_{\bar X_\tau}(x^\star_\tau) \le C_4 d\log T, \quad \text{for all} \quad \tau \in [\tau_{t,K-1}, \tau_{t,0}] \right\}.$$

Here, $C_4 > 0$ is a sufficiently large absolute constant, $c_3 > 0$ is a sufficiently small constant, and we denote

$$\varepsilon^{(n)}_{\text{score},t,i}(x^{(n)}_{\tau_{t,i}}) := \left\| s_{\tau_{t,i}}\left(x^{(n)}_{\tau_{t,i}}\right) - s^\star_{\tau_{t,i}}\left(x^{(n)}_{\tau_{t,i}}\right)\right\|_2, \quad \varepsilon^{(n)}_{\text{Jacobi},t,i}(x^{(n)}_{\tau_{t,i}}) := \left\| \frac{\partial s_{\tau_{t,i}}\left(x^{(n)}_{\tau_{t,i}}\right)}{\partial x} - \frac{\partial s^\star_{\tau_{t,i}}\left(x^{(n)}_{\tau_{t,i}}\right)}{\partial x}\right\|. \tag{29}$$

The next lemma shows that under the event E_t, the terms $\|x_{t-1} - x^\star_{\tau_{t,K-1}}\|_2$ are well controlled.

Lemma 3. *For each t, on the event E_t, when $K \lesssim \log T$ and $T \ge C_2 d\log^4 T$ for some sufficiently large constant $C_2 > 0$, it holds*

$$\|x^{(N)}_{\tau_{t,i}} - x^\star_{\tau_{t,i}}\|_2^2 \lesssim \log K\, \tau_{t,0}^2 \frac{\log^2 T}{T^2}\sum_{n=0}^{N}\sum_{j=0}^{K-1}\left(\varepsilon^{(n)}_{\text{score},t,j}(x^{(n)}_{\tau_{t,j}})\right)^2 + \frac{d\tau_{t,0}\log^3 T}{T^2}\left(\frac{M_0 K d\log^2 T}{T}\right)^{2K}$$

for all $0 \le i \le K-1$ and where $M_0 > 1$ is a large enough constant.

Proof. See Section B.1. □

Similarly, under the event E_t, the terms $\|J^{(N)}_{\tau_{t,i}} - J^\star_{\tau_{t,i}}\|^2$ are well controlled.

Lemma 4. *For each t, on the event E_t, when $K \lesssim \log T$ and $T \geq C_2 d \log^4 T$ for some sufficiently large constant $C_2 > 0$, the Jacobian satisfies*

$$\left\|(J^{(N)}_{\tau_{t,K-1}})^{-1}\right\| \lesssim 1 \tag{30}$$

and

$$\begin{aligned}\|J^{(N)}_{\tau_{t,i}} - J^\star_{\tau_{t,i}}\|^2 \lesssim K \log^2 K \, \frac{d^3 \log^7 T}{T^4} \sum_{n=0}^{N} \sum_{j=0}^{K-1} \left(\varepsilon^{(n)}_{\mathrm{score},t,j}(x^{(n)}_{\tau_{t,j}})\right)^2 \\ + \log K \tau_{t,0}^2 \frac{\log^2 T}{T^2} \sum_{n=0}^{N} \sum_{j=0}^{K-1} \left(\varepsilon^{(n)}_{\mathrm{Jacobi},t,j}(x^{(n)}_{\tau_{t,j}})\right)^2 + \left(\frac{M_0 K d \log^2 T}{T}\right)^{2K+2}\end{aligned} \tag{31}$$

for all $0 \leq i \leq K-1$ and where $M_0 > 1$ is a large enough constant.

Proof. See Section B.2. □

Putting Lemma 3 and Lemma 4 into (27), we obtain that on the Event E_t, it holds

$$\frac{p_{X_{t-1}}(x_{t-1})}{p_{Y_{t-1}}(x_{t-1})} = \exp\left(O\left(\widetilde{\xi}_t(x_t) + d\left(\frac{M_0 K d \log^2 T}{T}\right)^{K+1}\right)\right) \frac{p_{X_t}(x_t)}{p_{Y_t}(x_t)}, \tag{32}$$

provided that $T \gtrsim d \log^4 T$, where

$$\widetilde{\xi}_t(x_t) := \frac{d\sqrt{\log K}\log T}{T} \sqrt{\sum_{i=0}^{K-1} \sum_{n=0}^{N} \left(\varepsilon^{(n)}_{\mathrm{Jacobi},t,i}(x^{(n)}_{\tau_{t,i}})\right)^2} + \frac{d^{3/2} \log K \log^2 T}{T} \sqrt{\sum_{i=0}^{K-1} \sum_{n=0}^{N} \left(\varepsilon^{(n)}_{\mathrm{score},t,i}(x^{(n)}_{\tau_{t,i}})\right)^2}. \tag{33}$$

Step 2: decomposing the total variation distance. Define

$$\mathcal{E} := \left\{x \in \mathbb{R}^d : \; q_1(x) > \max\{p_1(x), \exp(-c_6 d \log T)\}\right\}. \tag{34}$$

Informally, the set $\mathcal{E}$ collects those typical points $x \in \mathbb{R}^d$ at which $q_1(x)$ is not exponentially small. The next lemma shows that the total variation distance between q_1 and p_1 is primarily controlled by their discrepancy on $\mathcal{E}$.

Lemma 5. *Assume that Assumption 1 holds. Then, for any fixed constant $K_1 > 0$, if c_6 in (34) is chosen sufficiently large, it holds that*

$$\mathrm{TV}(q_1, p_1) \leq \mathbb{E}_{Y_1 \sim p_1}\left[\left(\frac{q_1(Y_1)}{p_1(Y_1)} - 1\right) \mathbf{1}\{Y_1 \in \mathcal{E}\}\right] + C T^{-2K_1} + \exp(-c_2 d \log T),$$

where $C > 0$ is an absolute constant and $c_2 > 0$ is a constant depending on c_6, c_R.

Proof. The proof is deferred to Appendix B.3. □

Define

$$\widetilde{S}_t(Y_T) := \sum_{k=2}^{t} \widetilde{\xi}_k(x_k), \qquad t \geq 2, \qquad \widetilde{S}_1(Y_T) := 0, \tag{35}$$

where $\{x_k\}$ are generated by Algorithm 1 with initial sample Y_T, and $\widetilde{\xi}_k(x_k)$ is defined in (33). Next, define

$$\mathcal{I}_1 := \left\{Y_T \in \mathbb{R}^d : \widetilde{S}_T(Y_T) \leq c_3\right\}, \tag{36}$$

where $c_3 > 0$ is a sufficiently small constant. Thus, $\mathcal{I}_1$ represents the set of initial points for which the accumulated score error along the backward trajectory is well controlled. With the above notation, Lemma 5 immediately yields

$$\begin{aligned}
\mathrm{TV}(q_1,p_1) \quad &\leq \quad \mathbb{E}_{Y_1\sim p_1}\left[\left(\frac{q_1\left(Y_1\right)}{p_1\left(Y_1\right)}-1\right)\mathbf{1}\left\{Y_1\in\mathcal{E}\right\}\right]+CT^{-2K_1}+\exp(-c_2\,d\log T)\\
&= \quad \underbrace{\mathbb{E}_{Y_T\sim p_T}\left[\left(\frac{q_1(Y_1)}{p_1(Y_1)}-1\right)\mathbf{1}\{Y_1\in\mathcal{E},\ Y_T\in\mathcal{I}_1\}\right]}_{\alpha_1}+\underbrace{\mathbb{E}_{Y_T\sim p_T}\left[\left(\frac{q_1(Y_1)}{p_1(Y_1)}-1\right)\mathbf{1}\{Y_1\in\mathcal{E},\ Y_T\in\mathcal{I}_2\}\right]}_{\alpha_2}\\
&\quad\ +CT^{-2K_1}+\exp(-c_2\,d\log T),
\end{aligned} \tag{37}$$

where the second equality follows from the fact that Y_1 is determined entirely by Y_T through the deterministic update rule, and $\mathcal{I}_2 := \mathbb{R}^d\setminus\mathcal{I}_1$. As we will show later, for any $Y_T\in\mathcal{I}_1$ such that $x_1\in\mathcal{E}$, or equivalently $Y_1\in\mathcal{E}$, one can conclude that Y_T is a typical point in E_t. In this case, we may apply (32) to control α_1. On the other hand, for any $Y_T\in\mathcal{I}_2$ with $Y_1\in\mathcal{E}$, we will show that its contribution to the total variation distance is negligible, so that α_2 is also well controlled.

Step 3: bounding α_1 and α_2. For any fixed point $Y_T\in\mathbb{R}^d$, define

$$\tau(Y_T) := \max\left\{2\leq t\leq T+1:\ \widetilde{S}_{t-1}(Y_T)\leq c_3\right\}. \tag{38}$$

The next lemma shows that we can control the density ratio up to the $(\tau(Y_T)-1)$-th iteration.

Lemma 6. *For any $Y_T\in\mathcal{I}_1$ such that $x_1\in\mathcal{E}$, the following hold. For every $2\leq t\leq\tau(Y_T)-1$,*

$$\frac{q_1(x_1)}{p_1(x_1)}=\left(1+O\left(\sum_{t<\tau(Y_T)}\widetilde{\xi}_t(x_t)+d^2\log^2 T\,K\left(\frac{M_0Kd\log^2 T}{T}\right)^K\right)\right)\frac{q_{\tau(Y_T)-1}(x_{\tau(Y_T)-1})}{p_{\tau(Y_T)-1}(x_{\tau(Y_T)-1})}, \tag{39}$$

and

$$\frac{q_\ell(x_\ell)}{2p_\ell(x_\ell)}\leq\frac{q_1(x_1)}{p_1(x_1)}\leq\frac{2q_\ell(x_\ell)}{p_\ell(x_\ell)},\qquad\forall\,\ell\leq\tau(Y_T)-1. \tag{40}$$

Here, $M_0>1$ is a sufficiently large constant.

Proof. The result follows from Lemmas 14 and 15 by the same argument as in the proof of Lemma 10 of Li et al. (2025). We therefore omit the details for brevity. □

By definition, for every $Y_T\in\mathcal{I}_1$, we have $\tau(Y_T)=T+1$. Combining this with Lemma 6, we obtain

$$\begin{aligned}
\alpha_1 &= \mathbb{E}_{Y_T\sim p_T}\Bigg[\Bigg(\Big(1+\sum_{t=2}^T\widetilde{\xi}_t(x_t)+d^2\log^2 TK\Big(\frac{M_0Kd\log^2 T}{T}\Big)^K\Big)\frac{q_T(Y_T)}{p_T(Y_T)}-1\Bigg)\mathbf{1}\{Y_1\in\mathcal{E},\ Y_T\in\mathcal{I}_1\}\Bigg]\\
&= \int\Bigg[\Bigg(\Big(1+\sum_{t=2}^T\widetilde{\xi}_t(x_t)+d^2\log^2 TK\Big(\frac{M_0Kd\log^2 T}{T}\Big)^K\Big)q_T(x_T)-p_T(x_T)\Bigg)\mathbf{1}\{x_1\in\mathcal{E},\ x_T\in\mathcal{I}_1\}\Bigg]\mathrm{d}x_T\\
&\leq \int|q_T(x_T)-p_T(x_T)|\,\mathrm{d}x_T+\int\sum_{t=2}^T\widetilde{\xi}_t(x_t)\,q_T(x_T)\,\mathbf{1}\{x_1\in\mathcal{E},\ x_T\in\mathcal{I}_1\}\,\mathrm{d}x_T+d^2\log^2 TK\left(\frac{M_0Kd\log^2 T}{T}\right)^K.
\end{aligned}$$

For the first term on the right-hand side, Lemma 1 yields

$$\int|q_T(x_T)-p_T(x_T)|\,\mathrm{d}x_T=2\,\mathrm{TV}\left(q_T,p_T\right)\lesssim\frac{1}{T^{K_1}}.$$

For the second term, observe that

$$\begin{aligned}
&\int \widetilde{S}_T(x_T)\, q_T(x_T)\, \mathbf{1}\{x_1 \in \mathcal{E},\ x_T \in \mathcal{I}_1\}\, \mathrm{d}x_T \\
=&\sum_{t=2}^{T} \mathbb{E}_{Y_T\sim p_T}\left[\frac{\log T}{T}\left(d\sqrt{\log K}\,\varepsilon_{\mathrm{Jacobi},t}(Y_t) + d^{3/2}\log K\log T\,\varepsilon_{\mathrm{score},t}(Y_t)\right)\frac{q_T(Y_T)}{p_T(Y_T)}\mathbf{1}\{Y_1 \in \mathcal{E},\ Y_T \in \mathcal{I}_1\}\right] \\
\overset{(40)}{\le}& 4\sum_{t=2}^{T} \mathbb{E}_{Y_T\sim p_T}\left[\frac{\log T}{T}\left(d\sqrt{\log K}\,\varepsilon_{\mathrm{Jacobi},t}(Y_t) + d^{3/2}\log K\log T\,\varepsilon_{\mathrm{score},t}(Y_t)\right)\frac{q_t(Y_t)}{p_t(Y_t)}\mathbf{1}\{Y_1 \in \mathcal{E},\ Y_T \in \mathcal{I}_1\}\right] \\
\le& 4\sum_{t=2}^{T} \mathbb{E}_{Y_t\sim p_t}\left[\frac{\log T}{T}\left(d\sqrt{\log K}\,\varepsilon_{\mathrm{Jacobi},t}(Y_t) + d^{3/2}\log K\log T\,\varepsilon_{\mathrm{score},t}(Y_t)\right)\frac{q_t(Y_t)}{p_t(Y_t)}\right] \\
=& 4\sum_{t=2}^{T} \mathbb{E}_{Y_t\sim q_t}\left[\frac{\log T}{T}\left(d\sqrt{\log K}\,\varepsilon_{\mathrm{Jacobi},t}(Y_t) + d^{3/2}\log K\log T\,\varepsilon_{\mathrm{score},t}(Y_t)\right)\right] \\
\lesssim&\ \ d\log^{5/2}T\sqrt{\log K}\,\varepsilon_{\mathrm{Jacobi}} + d^{3/2}\log^{7/2}T\log K\varepsilon_{\mathrm{score}},
\end{aligned}$$

where the last inequality follows from Assumptions 2 and 3, together with the bounds

$$\begin{aligned}
\frac{1}{T}\sum_t \mathbb{E}_{Y_t\sim q_t}\left[\varepsilon_{\mathrm{score},t}(Y_t)\right] &\le \sqrt{\frac{1}{T}\sum_t \mathbb{E}_{Y_t\sim q_t}\left[\varepsilon^2_{\mathrm{score},t}(Y_t)\right]} \lesssim \sqrt{(N+1)K}\,\varepsilon_{\mathrm{score}} \asymp \varepsilon_{\mathrm{score}}\log^{3/2}T, \\
\frac{1}{T}\sum_t \mathbb{E}_{Y_t\sim q_t}\left[\varepsilon_{\mathrm{Jacobi},t}(Y_t)\right] &\le \sqrt{\frac{1}{T}\sum_t \mathbb{E}_{Y_t\sim q_t}\left[\varepsilon^2_{\mathrm{Jacobi},t}(Y_t)\right]} \lesssim \sqrt{(N+1)K}\,\varepsilon_{\mathrm{Jacobi}} \asymp \varepsilon_{\mathrm{Jacobi}}\log^{3/2}T.
\end{aligned}$$

Combining the above estimates, we arrive at

$$\alpha_1 \lesssim \frac{1}{T^{K_1}} + d^2\log^2 TK\left(\frac{M_0Kd\log^2 T}{T}\right)^K + d\log^{5/2}T\sqrt{\log K}\,\varepsilon_{\mathrm{Jacobi}} + d^{3/2}\log^{7/2}T\log K\varepsilon_{\mathrm{score}}.$$

Using techniques similar to those in Li et al. (2024b), the term α_2 can also be controlled effectively, as stated in the following lemma.

Lemma 7. *Suppose that Assumptions 1, 2, and 3 hold. If $K \lesssim \log T$ and $T \ge C_2 d\log^4 T$, where $M_0, K_1 > 0$ are sufficiently large constants and c_2 is the constant appearing in Lemma 5, then*

$$\alpha_2 \lesssim d^2\log^2 TK\left(\frac{M_0Kd\log^2 T}{T}\right)^K + d\log^{5/2}T\sqrt{\log K}\,\varepsilon_{\mathrm{Jacobi}} + d^{3/2}\log^{7/2}T\log K\varepsilon_{\mathrm{score}} + e^{-c_2 d\log T} + \frac{1}{T^{K_1}}.$$

Proof. The proof follows arguments similar to those of Lemma 11 in Li et al. (2024b), with minor modifications. Specifically, the proof in Li et al. (2024b) relies on a bounded-support assumption on X_0. However, a careful inspection shows that this condition can be replaced by Assumption 1 combined with Markov's inequality. This yields the similar result up to an additional negligible error $T^{-K_1} + \exp(-c_2 d\log T)$. We therefore omit the proof for brevity. □

Step 4: controlling $\mathrm{TV}(q_1, p_1)$**.** Substituting the bounds for α_1 and α_2 into (37), we obtain the desired result

$$\begin{aligned}
&\mathrm{TV}(q_1,p_1) \\
\lesssim&\ \ d^2\log^2 TK\left(\frac{M_0Kd\log^2 T}{T}\right)^K + d\log^{5/2}T\sqrt{\log K}\,\varepsilon_{\mathrm{Jacobi}} + d^{3/2}\log^{7/2}T\log K\varepsilon_{\mathrm{score}} + e^{-c_2 d\log T} + \frac{1}{T^{K_1}} \\
\asymp&\ \ d^2\log^2 TK\left(\frac{M_0Kd\log^2 T}{T}\right)^K + d\log^{5/2}T\sqrt{\log K}\,\varepsilon_{\mathrm{Jacobi}} + d^{3/2}\log^{7/2}T\log K\varepsilon_{\mathrm{score}} + \frac{1}{T^{K_1}}.
\end{aligned}$$

6 Numerical experiments

In this section, we numerically compare our method with the Li-HEROISM baseline (Li et al., 2025). Following Huang et al. (2025), we consider a finite Gaussian mixture target distribution of the form

$$p_{\mathrm{data}}(x) = \sum_{\ell=1}^{M} w_\ell \mathcal{N}(x; m_\ell, C_\ell), \qquad \sum_{\ell=1}^{M} w_\ell = 1, \tag{41}$$

where M denotes the number of mixture components, $x \in \mathbb{R}^d$, $m_\ell \in \mathbb{R}^d$, and $C_\ell \in \mathbb{R}^{d\times d}$. Under the forward process (6), the marginal density of $\bar{X}_\tau$ remains a Gaussian mixture, as observed in Huang et al. (2025):

$$p_{\bar{X}_\tau}(x) = \sum_{\ell=1}^{M} w_\ell \mathcal{N}\left(x; \sqrt{1-\tau}\, m_\ell,\, (1-\tau)C_\ell + \tau I_d\right). \tag{42}$$

Consequently, for each $x \in \mathbb{R}^d$, the exact score function used in the experiments is available in closed form:

$$s^\star_\tau(x) = -\sum_{\ell=1}^{M} \frac{w_\ell \mathcal{N}\left(x; \sqrt{1-\tau}\, m_\ell,\, (1-\tau)C_\ell + \tau I_d\right)}{p_{\bar{X}_\tau}(x)} \left((1-\tau)C_\ell + \tau I_d\right)^{-1} \left(x - \sqrt{1-\tau}\, m_\ell\right). \tag{43}$$

In practice, the score function is typically trained by a neural network via score matching on progressively corrupted data. In the present numerical study, however, we bypass the training stage and instead evaluate the samplers using an imperfect analytical score parameterized by a scalar δ:

$$s_\tau(x) = s^\star_\tau(x) + \delta\eta(x). \tag{44}$$

Following the artificial-error protocol of Huang et al. (2025), we consider the perturbations

$$\eta_{\mathrm{const}}(x) = \frac{1}{\sqrt{d}}\mathbf{1}, \qquad \eta_{\mathrm{lin}}(x) = \frac{x - m_0}{\sqrt{d}}, \qquad \eta_{\sin}(x) = \frac{\sin(x) \odot (x - m_0)}{\sqrt{d}} \in \mathbb{R}^d,$$

where $m_0 = \mathbb{E}[X_0]$, the sine function is applied coordinatewise, $\odot$ denotes coordinatewise multiplication, and $\mathbf{1} \in \mathbb{R}^d$ is the all-ones vector.

The reverse dynamics are governed by the probability flow ODE (8). We use the same two-phase schedule (17)–(18) as in the theoretical analysis, with $c_0 = 1$ and $c_1 = 0.5$, and linearly rescale the raw endpoints $\tau_t = 1 - \bar{\alpha}_t$ to the fixed numerical interval $[\tau_{\min}, \tau_{\max}] = [10^{-3}, 0.999]$. Thus, $\tau_{\max} = 0.999$ is used as a numerical approximation to the fully noised endpoint $\tau = 1$, whereas $\tau_{\min} = 10^{-3}$ serves as a numerical approximation to the data endpoint $\tau = 0$. To initialize the reverse sampler, we draw J particles from $Y_T \sim \mathcal{N}(0, I_d)$ at $\tau_{\max}$ and evolve the deterministic reverse process down to $\tau_{\min}$.

With T outer iterations, K interpolation nodes, N refinement rounds, and cached score evaluations, the total numbers of score function evaluations are

$$\mathrm{SFE}_{\mathrm{Li}} = T\big(1 + (K-1)N\big), \qquad \mathrm{SFE}_{\mathrm{Ours}} = T\big(K + (K-1)N\big). \tag{45}$$

Here, SFE denotes the total number of score function evaluations over all outer iterations. For each fixed time point τ, one batched evaluation of $s_\tau(\cdot)$ over all J particles is counted as a single score function evaluation. The goal of the experiments is to compare the finite-budget performance of the two methods under the same target distribution, score perturbation, and total score-evaluation budget.

6.1 Test $d = 1$

We first consider the one-dimensional three-component Gaussian mixture (41) with

$$w = [0.1; 0.4; 0.5], \qquad m = [-6.0; 4.0; 6.0], \qquad C = [0.25; 0.25; 0.25].$$

This target follows the one-dimensional setting of Huang et al. (2025) and is used to assess whether the samplers can accurately recover the locations and relative weights of well-separated modes. We sample $J = 5 \times 10^4$ particles from $\mathcal{N}(0, 1)$, fix $K = 6$ and $N = 3$, and vary the number of outer iterations T. Under the SFE convention in (45), this yields $\mathrm{SFE}_{\mathrm{Li}} = 16T$ and $\mathrm{SFE}_{\mathrm{Ours}} = 21T$.

For density visualization, we estimate the empirical density of the generated samples at $\tau_{\min}$, denoted by $\widehat{p}_{\tau_{\min}}$, using a kernel density estimator (KDE) with bandwidth selected according to Silverman's rule (Silverman, 2018). Since $\tau_{\min} = 10^{-3}$ approximates the data endpoint, $p_{\bar{X}_{\tau_{\min}}}$ and $\widehat{p}_{\tau_{\min}}$ serve as numerical proxies for q_1 and p_1, respectively. Following the evaluation protocol of Huang et al. (2025), we approximate the total variation distance on $[-10, 10]$ using a composite midpoint quadrature with 1000 subintervals.

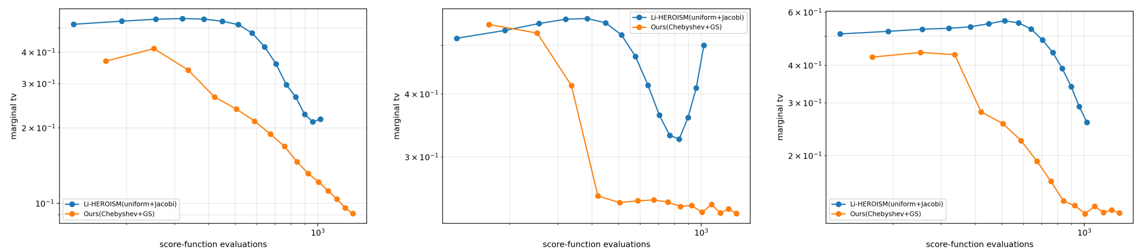

Figure 1: One-dimensional total variation (TV) comparison with $\delta = 0.05$. From left to right: η_{const}, η_{lin}, and η_{sin} score perturbations. Each panel plots TV distance versus total SFE for Li-HEROISM and the proposed GS sampler.

Figure 1 shows that, under the same SFE budget, the proposed Gauss-Seidel sampler achieves a faster reduction in total variation distance than Li-HEROISM across all three score-perturbation settings. Figure 2 compares the estimated densities with the reference density $p_{\bar{X}_{\tau_{\min}}}$. Table 2 further shows that our method often attains smaller total variation errors even when Li-HEROISM is allowed twice as many SFEs, demonstrating the efficiency of the proposed sampler.

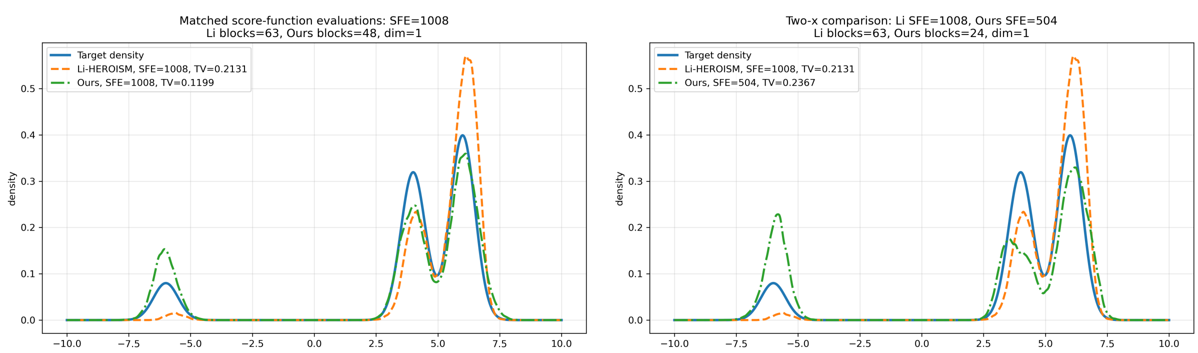

Figure 2: One-dimensional density comparison under the η_{const} score perturbation with $\delta = 0.05$. Left: matched SFE. Right: Li-HEROISM with twice the SFE of the proposed Gauss-Seidel sampler.

Table 2: One-dimensional experiment with η_{const} score perturbation, where Li-HEROISM is allowed twice as many SFEs as the proposed Gauss-Seidel method. The perturbation is $\eta(x) = \mathbf{1}/\sqrt{d}$ with $\delta = 0.05$.

T_{Ours}	$\mathrm{SFE}_{\mathrm{Ours}}$	T_{Li}	$\mathrm{SFE}_{\mathrm{Li}}$	$\mathrm{TV}_{\mathrm{Li}}$	$\mathrm{TV}_{\mathrm{Ours}}$
8	**168**	21	**336**	0.5446	**0.3689**
16	**336**	42	**672**	0.3901	**0.3401**
24	**504**	63	**1008**	**0.2131**	0.2367

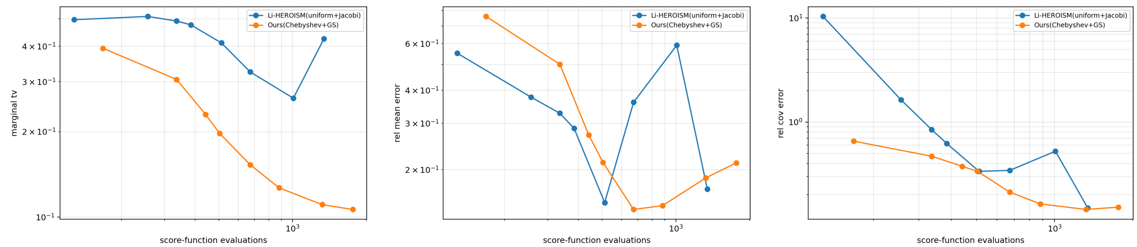

Figure 3: High-dimensional experiment for $d = 128$ under the η_{lin} artificial score perturbation with $\delta = 0.05$. Error versus total SFE. Left: first-coordinate marginal total variation (TV). Middle: full-dimensional relative mean error. Right: full-dimensional relative covariance error.

6.2 Test $d = 128$

Following the high-dimensional Gaussian-mixture setting in Huang et al. (2025), we next test a $d = 128$ anisotropic Gaussian mixture (41) with $M = 5$ components. The mixture weights w_ℓ are sampled from Uniform$(0, 1)$ and normalized. The distribution means and covariance matrices are generated by

$$m_\ell \sim \mathcal{N}(0, 3^2 I_d), \qquad C_\ell = \frac{1}{8}\left(\frac{W_\ell^\top W_\ell}{d} + I_d\right), \qquad (W_\ell)_{ij} \sim \mathcal{N}(0, 1), \qquad W_\ell \in \mathbb{R}^{d\times d}$$

with $d = 128$. We use the same algorithmic setup as in the one-dimensional test, except that the number of particles is $J = 2 \times 10^4$. Since full-dimensional total variation (TV) estimation from particles is unreliable in high dimension, we follow Huang et al. (2025) and compute the first-coordinate marginal TV, estimated with the same KDE bandwidth and fixed-grid quadrature rule as in the one-dimensional test. To complement this one-dimensional marginal metric, we also compute the full-dimensional relative mean error and relative covariance error:

$$\mathrm{Err}_{\text{mean}} = \frac{\|\widehat{m} - m_{\tau_{\min}}\|_2}{\|m_{\tau_{\min}}\|_2}, \qquad \mathrm{Err}_{\text{cov}} = \frac{\|\widehat{\Sigma} - \Sigma_{\tau_{\min}}\|_{\mathrm{F}}}{\|\Sigma_{\tau_{\min}}\|_{\mathrm{F}}}.$$

Here $\widehat{m} \in \mathbb{R}^d$ and $\widehat{\Sigma} \in \mathbb{R}^{d\times d}$ are the empirical mean and covariance computed directly from the generated particles, while $m_{\tau_{\min}}$ and $\Sigma_{\tau_{\min}}$ are the exact mean and covariance of $p_{\bar{X}_{\tau_{\min}}}$. Figure 3 shows that the proposed sampler reduces the marginal TV more steadily and substantially faster than Li-HEROISM, with more stable covariance behavior. Although the relative mean error is mixed, it is only an auxiliary moment diagnostic. Table 3 further shows that our method attains smaller marginal TV in all tested cases even when Li-HEROISM uses twice as many SFEs, demonstrating its finite-budget efficiency.

Table 3: High-dimensional experiment when Li-HEROISM uses twice as many score function evaluations as the proposed Gauss–Seidel sampler. The score perturbation is η_{lin} with $\delta = 0.05$.

T_{Ours}	SFE_{Ours}	T_{Li}	SFE_{Li}	$\text{TV}_{1,\text{Li}}$	$\text{TV}_{1,\text{Ours}}$	$\text{MeanErr}_{\text{Li}}$	$\text{MeanErr}_{\text{Ours}}$	$\text{CovErr}_{\text{Li}}$ / $\text{CovErr}_{\text{Ours}}$
8	168	21	336	0.4911	**0.3936**	0.3271	0.7598	0.8456 / **0.6556**
16	336	42	672	0.3248	**0.3056**	0.3600	0.5010	0.3434 / 0.4695
24	504	63	1008	0.2624	**0.1971**	0.5911	**0.2132**	0.5244 / **0.3365**
32	672	84	1344	0.4251	**0.1528**	0.1688	**0.1415**	0.1491 / 0.2121

7 Discussion and future directions

In this paper, we develop a Chebyshev–Gauss–Seidel higher-order sampler for training-free acceleration of diffusion models through the probability flow ODE, and establish a convergence guarantee that allows the approximation order K to grow with the number of iterations T. In the exact-score setting, by taking

$K \asymp \log T$, the proposed sampler achieves a target total variation accuracy ε using at most

$$\widetilde{O}\Big(d^{1+o_T(1)}\varepsilon^{-1/K_1}\Big)$$

score function evaluations, where $o_T(1) \to 0$ as $T \to \infty$ and $K_1 > 0$ is an absolute constant. Our guarantee holds under a polynomial second-moment condition on the target distribution and accommodates inexact score estimation, without imposing higher-order smoothness assumptions on the score estimates.

Several questions remain open. First, it would be interesting to understand whether the intrinsic low-dimensional structure of the target distribution can be leveraged to further accelerate the proposed sampler. Second, although our analysis avoids higher-order smoothness assumptions on the score estimates, it still requires control of the Jacobian estimation error; an important direction for future work is to determine whether this requirement can be weakened. Finally, our numerical results indicate that interpolation nodes and successive-refinement schemes play a significant role in the performance of higher-order samplers. This naturally raises the question of whether a unified non-asymptotic theory can rigorously explain the impact of these algorithmic design choices.

References

Milton Abramowitz and Irene A. Stegun (eds.). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55 of *Applied Mathematics Series*. National Bureau of Standards, Washington, DC, 1964.

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

Keith Ball. An elementary introduction to modern convex geometry. In Silvio Levy (ed.), *Flavors of Geometry*, volume 31 of *Mathematical Sciences Research Institute Publications*, pp. 1–58. Cambridge University Press, Cambridge, 1997.

Joe Benton, Valentin De Bortoli, Arnaud Doucet, and George Deligiannidis. Nearly d-linear convergence bounds for diffusion models via stochastic localization. In *International Conference on Learning Representations*, 2024.

Eliot Beyler and Francis Bach. Convergence of deterministic and stochastic diffusion-model samplers: A simple analysis in wasserstein distance. *arXiv preprint arXiv:2508.03210*, 2025.

Stefano Bruno, Ying Zhang, Dongyoung Lim, Omer Deniz Akyildiz, and Sotirios Sabanis. On diffusion-based generative models and their error bounds: The log-concave case with full convergence estimates. *Transactions on Machine Learning Research*, 2025.

Fan Chen, Sinho Chewi, Constantinos Daskalakis, and Alexander Rakhlin. High-accuracy sampling for diffusion models and log-concave distributions. *arXiv preprint arXiv:2602.01338*, 2026.

Hongrui Chen, Holden Lee, and Jianfeng Lu. Improved analysis of score-based generative modeling: User-friendly bounds under minimal smoothness assumptions. In *International Conference on Machine Learning*, pp. 4735–4763, 2023a.

Sitan Chen, Sinho Chewi, Holden Lee, Yuanzhi Li, Jianfeng Lu, and Adil Salim. The probability flow ODE is provably fast. *Advances in Neural Information Processing Systems*, 36:68552–68575, 2023b.

Sitan Chen, Sinho Chewi, Jerry Li, Yuanzhi Li, Adil Salim, and Anru R. Zhang. Sampling is as easy as learning the score: Theory for diffusion models with minimal data assumptions. In *International Conference on Learning Representations*, 2023c.

Giovanni Conforti, Alain Durmus, and Marta Gentiloni Silveri. KL convergence guarantees for score diffusion models under minimal data assumptions. *SIAM Journal on Mathematics of Data Science*, 7(1):86–109, 2025. doi: 10.1137/23M1613670.

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

Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106 (496):1602–1614, 2011. doi: 10.1198/jasa.2011.tm11181.

Xuefeng Gao and Lingjiong Zhu. Convergence analysis for general probability flow ODEs of diffusion models in wasserstein distances. In *International Conference on Artificial Intelligence and Statistics*, pp. 1009–1017, 2025.

Xuefeng Gao, Hoang M. Nguyen, and Lingjiong Zhu. Wasserstein convergence guarantees for a general class of score-based generative models. *Journal of Machine Learning Research*, 26(43):1–54, 2025.

Khashayar Gatmiry, Sitan Chen, and Adil Salim. High-accuracy and dimension-free sampling with diffusions. *arXiv preprint arXiv:2601.10708*, 2026.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27, 2014.

Uwe G. Haussmann and Étienne Pardoux. Time reversal of diffusions. *The Annals of Probability*, 14(4): 1188–1205, 1986.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.

Daniel Zhengyu Huang, Jiaoyang Huang, and Zhengjiang Lin. Convergence analysis of probability flow ODE for score-based generative models. *IEEE Transactions on Information Theory*, 71(6):4581–4601, 2025.

Zhihan Huang, Yuting Wei, and Yuxin Chen. Denoising diffusion probabilistic models are optimally adaptive to unknown low dimensionality. *Mathematics of Operations Research*, 2026. doi: 10.1287/moor.2024.0769.

Aapo Hyvärinen. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6:695–709, 2005.

Yuchen Jiao, Na Li, Changxiao Cai, and Gen Li. Are first-order diffusion samplers really slower? a fast forward-value approach. *arXiv preprint arXiv:2512.24927*, 2025.

Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *International Conference on Learning Representations*, 2014.

Holden Lee, Jianfeng Lu, and Yixin Tan. Convergence for score-based generative modeling with polynomial complexity. *Advances in Neural Information Processing Systems*, 35:22870–22882, 2022.

Holden Lee, Jianfeng Lu, and Yixin Tan. Convergence of score-based generative modeling for general data distributions. In *International Conference on Algorithmic Learning Theory*, pp. 946–985, 2023.

Gen Li and Changxiao Cai. Provable acceleration for diffusion models under minimal assumptions. *arXiv preprint arXiv:2410.23285*, 2024.

Gen Li and Yuling Yan. $O(d/T)$ convergence theory for diffusion probabilistic models under minimal assumptions. *Journal of Machine Learning Research*, 26(292):1–55, 2025.

Gen Li, Yu Huang, Timofey Efimov, Yuting Wei, Yuejie Chi, and Yuxin Chen. Accelerating convergence of score-based diffusion models, provably. In *International Conference on Machine Learning*, pp. 27942–27954, 2024a.

Gen Li, Yuting Wei, Yuejie Chi, and Yuxin Chen. A sharp convergence theory for the probability flow ODEs of diffusion models. *arXiv preprint arXiv:2408.02320*, 2024b.

Gen Li, Yuchen Zhou, Yuting Wei, and Yuxin Chen. Faster diffusion models via higher-order approximation. *arXiv preprint arXiv:2506.24042*, 2025.

Peiyi Liu, Zhaoqiang Liu, and Yiqi Gu. Operator splitting based diffusion samplers and improved convergence analysis. *arXiv preprint arXiv:2601.17375*, 2026.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver: A fast ODE solver for diffusion probabilistic model sampling in around 10 steps. *Advances in Neural Information Processing Systems*, 35:5775–5787, 2022.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver++: Fast solver for guided sampling of diffusion probabilistic models. *Machine Intelligence Research*, 22(4):730–751, 2025. doi: 10.1007/s11633-025-1562-4.

John C. Mason and David C. Handscomb. *Chebyshev polynomials*. Chapman and Hall/CRC, 2002.

Francesco Pedrotti, Jan Maas, and Marco Mondelli. Improved convergence of score-based diffusion models via prediction-correction. *Transactions on Machine Learning Research*, 2024.

Danilo Jimenez Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International Conference on Machine Learning*, pp. 1530–1538, 2015.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Bjorn Ommer. High-resolution image synthesis with latent diffusion models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10684–10695, 2022.

Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily Denton, Seyed Kamyar Seyed Ghasemipour, Burcu Karagol Ayan, S. Sara Mahdavi, Raphael Gontijo Lopes, Tim Salimans, Jonathan Ho, David J. Fleet, and Mohammad Norouzi. Photorealistic text-to-image diffusion models with deep language understanding. *Advances in Neural Information Processing Systems*, 35:36479–36494, 2022.

Bernard W. Silverman. *Density estimation for statistics and data analysis*. Routledge, 2018.

Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pp. 2256–2265, 2015.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021a.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in Neural Information Processing Systems*, 32, 2019.

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

Jiaqi Tang and Yuling Yan. Adaptivity and convergence of probability flow ODEs in diffusion generative models. *arXiv preprint arXiv:2501.18863*, 2025.

Lloyd N. Trefethen. *Approximation theory and approximation practice, extended edition*. SIAM, 2019.

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

Qinsheng Zhang and Yongxin Chen. Fast sampling of diffusion models with exponential integrator. In *International Conference on Learning Representations*, 2023.

Wenliang Zhao, Lujia Bai, Yongming Rao, Jie Zhou, and Jiwen Lu. UniPC: A unified predictor-corrector framework for fast sampling of diffusion models. *Advances in Neural Information Processing Systems*, 36: 49842–49869, 2023.

A Auxiliary Lemmas

This section gathers the auxiliary estimates that will be used repeatedly in the theoretical analysis and in the subsequent density-ratio propagation arguments.

Lemma 8. *Assume that $K \geq 2$, and let*

$$z_j = \cos\left(\frac{j\pi}{K-1}\right), \qquad 0 \leq j \leq K-1,$$

denote the Chebyshev–Lobatto nodes on $[-1,1]$. Let $\tau_{t,j}$ be defined as in (12), and let $\psi_{t,j}$ be given by (14). Then

$$\sup_{\tau\in[\tau_{t,K-1},\tau_{t,0}]} \sum_{j=0}^{K-1} |\psi_{t,j}(\tau)| \leq \frac{2}{\pi}\log K + 1.$$

Proof. The proof is given in Section C.1. □

Lemma 9. *For large enough T, one has*

$$\alpha_t \geq 1 - \frac{c_1 \log T}{T} \geq \frac{1}{2}, \qquad 1 \leq t \leq T. \tag{46a}$$

$$\frac{1}{2}\frac{1-\alpha_t}{1-\bar{\alpha}_t} \leq \frac{1}{2}\frac{1-\alpha_t}{\alpha_t - \bar{\alpha}_t} \leq \frac{1-\alpha_t}{1-\bar{\alpha}_{t-1}} \leq \frac{4c_1\log T}{T}, \qquad 2 \leq t \leq T. \tag{46b}$$

$$1 \leq \frac{1-\bar{\alpha}_t}{1-\bar{\alpha}_{t-1}} \leq 1 + \frac{4c_1 \log T}{T}, \qquad 2 \leq t \leq T. \tag{46c}$$

$$\bar{\alpha}_T \leq \frac{1}{T^{C_1}}. \tag{46d}$$

$$\frac{\bar{\alpha}_{t+1}}{1-\bar{\alpha}_{t+1}} \leq \frac{\bar{\alpha}_t}{1-\bar{\alpha}_t} \leq \frac{4\bar{\alpha}_{t+1}}{1-\bar{\alpha}_{t+1}}, \qquad 1 \leq t < T. \tag{46e}$$

$$\left|\frac{\tau_{t,i_1}-\tau_{t,i_2}}{\tau_{t,i_3}(1-\tau_{t,i_4})}\right| \leq 8c_1\frac{\log T}{T}, \qquad 2 \leq t \leq T, \quad 0 \leq i_1, i_2, i_3, i_4 \leq K-1. \tag{46f}$$

$$|\gamma_{t,j}(\tau_{t,i})| \leq 2(\tau_{t,0} - \tau_{t,i}), \qquad 0 \leq i, j \leq K-1. \tag{46g}$$

$$1-\tau_{t,i} \asymp 1-\tau_{t,j}, \qquad \tau_{t,i} \asymp \tau_{t,j}, \qquad 0 \leq i, j \leq K-1, \quad 2 \leq t \leq T. \tag{46h}$$

$$\sum_{j=0}^{K-1} |\gamma_{t,j}(\tau_{t,i})| \leq C_4(\tau_{t,0}-\tau_{t,i})\log K, \qquad 0 \leq i \leq K-1. \tag{46i}$$

$$\sum_{j=0}^{K-1} |A_{ji}^{(t)}| \leq C_4 \frac{(\tau_{t,0}-\tau_{t,i})\log K}{(1-\tau_{t,0})^{3/2}}, \qquad 0 \leq i \leq K-1, \quad 2 \leq t \leq T. \tag{46j}$$

$$\sum_{j=0}^{K-1} |A_{ji}^{(t)}|^2 \leq C_5 \frac{(\tau_{t,0}-\tau_{t,i})^2\log K}{(1-\tau_{t,0})^{3}}, \qquad 0 \leq i \leq K-1, \quad 2 \leq t \leq T. \tag{46k}$$

Here, $\gamma_{t,j}(\tau_{t,i})$ is given in (16), c_1 is defined in (18), $C_1 > 0$ is a sufficiently large constant, $C_4, C_5 > 0$ are universal constants, and $A_{ji}^{(t)} := \frac{1}{2}\gamma_{t,j}(\tau_{t,i})(1-\tau_{t,j})^{-3/2}$ for convenience.

Proof. See Section C.2. □

Define

$$\theta_\tau(x) := \max\left\{-\frac{\log p_{\bar{X}_\tau}(x)}{d\log T}, c_6\right\}, \tag{47}$$

where $c_6 > 0$ is a sufficiently large constant. The next lemma provides a quantitative tail estimate for the conditional distribution of X_0 given the continuous-time forward variable $\bar{X}_\tau$, where $\bar{X}_\tau$ is given in (6).

Lemma 10. *Suppose that $\tau \geq T^{-c_0}$ where c_0 is defined in* (17)*. Assume that Assumption 1 holds. Then for any constant $c_5 \geq 2$, one has*

$$\mathbb{P}\Big(\big\|\sqrt{1-\tau}X_0 - y\big\|_2 > 5c_5\sqrt{\theta_\tau(y)\, d\,\tau \log T}\;\Big|\; \bar{X}_\tau = y\Big) \leq \exp\big(-c_5^2\theta_\tau(y) d \log T\big)\,.$$

Moreover, there exist absolute constants $C_6, C_7, C_8, C_9 > 0$ such that

$$\mathbb{E}\big[\big\|\sqrt{1-\tau}X_0 - y\big\|_2 \,\big|\, \bar{X}_\tau = y\big] \leq C_6\sqrt{\theta_\tau(y)\, d\,\tau \log T}, \tag{48}$$

$$\mathbb{E}\Big[\big\|\sqrt{1-\tau}X_0 - y\big\|_2^2 \,\Big|\, \bar{X}_\tau = y\Big] \leq C_7\, \theta_\tau(y)\, d\,\tau \log T, \tag{49}$$

$$\mathbb{E}\Big[\big\|\sqrt{1-\tau}X_0 - y\big\|_2^3 \,\Big|\, \bar{X}_\tau = y\Big] \leq C_8\big(\theta_\tau(y)\, d\,\tau \log T\big)^{3/2}, \tag{50}$$

$$\mathbb{E}\Big[\big\|\sqrt{1-\tau}X_0 - y\big\|_2^4 \,\Big|\, \bar{X}_\tau = y\Big] \leq C_9\big(\theta_\tau(y)\, d\,\tau \log T\big)^{2}. \tag{51}$$

Proof. See Section C.3. □

By Tweedie's formula (Efron, 2011), the score function (9) can also be expressed as

$$s_\tau^\star(x) = -\frac{1}{\tau}\left(x - \sqrt{1-\tau}\,\mathbb{E}\left[X_0 \mid \bar{X}_\tau = x\right]\right), \tag{52}$$

which further implies

$$s_\tau^\star(\sqrt{1-\tau}x) = -\frac{\sqrt{1-\tau}}{\tau}\int_{x_0} p_{X_0|\bar{X}_\tau}\left(x_0 \mid \sqrt{1-\tau}x\right)(x - x_0)\,\mathrm{d}x_0.$$

Based on Lemma 10, it is straightforward to obtain the following results.

Lemma 11. *For any $\tau \geq T^{-c_0}$ where c_0 is defined in* (17)*, one has*

$$\|s_\tau^\star(x)\|_2 \lesssim \sqrt{\frac{d\,\theta_\tau(x)\log T}{\tau}}, \tag{53}$$

$$\left\|\frac{\partial s_\tau^\star(x)}{\partial x}\right\| \lesssim \frac{d\,\theta_\tau(x)\log T}{\tau}. \tag{54}$$

Moreover, under the condition

$$-\log p_{\bar{X}_\tau}\big(\lambda x_1 + (1-\lambda)x_2\big) \lesssim d\log T, \qquad \forall \lambda \in [0,1],$$

one has

$$\left\|\frac{\partial s_\tau^\star(x_1)}{\partial x} - \frac{\partial s_\tau^\star(x_2)}{\partial x}\right\| \lesssim \sqrt{\frac{d^3\log^3 T}{\tau^3}}\,\|x_1 - x_2\|_2. \tag{55}$$

Proof. The first estimate follows directly from Tweedie's formula and the first-moment bound in Lemma 10. The second estimate follows from the (52) together with the second-moment bound in Lemma 10. The third estimate is standard and follows, for instance, from the argument of Li et al. (2024b, Claim (88)). □

Lemma 12 (Li et al. (2025), Lemma 4)**.** *Suppose that $-\log p_{\bar{X}_{\tau_e}}(x_{\tau_e}^\star) \leq \theta d \log T$ for some $T^{-c_0} \leq \tau_e < 1$ where c_0 is defined in* (17) *and some $\theta > 1$, then for every τ' satisfying $|\tau' - \tau_e| \leq c_4\tau_e(1-\tau_e)$, one has*

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

Here, $\bar{X}_\tau$ is given in (6) *and $c_4 > 0$ is some sufficiently small constant.*

Lemma 13. *Assume that $\tau \in [\tau_{t,K-1}, \tau_{t,0}]$ and $-\log p_{\bar{X}_\tau}(x^\star_\tau) \le \theta\, d \log T$ for some $\theta \ge c_6$ where c_6 is defined in* (47). *Set*

$$u^\star_\tau := \frac{x^\star_\tau}{\sqrt{1-\tau}} \quad \text{and} \quad J^\star_\tau := \frac{\partial u^\star_\tau}{\partial u^\star_{\tau_{t,0}}} = \frac{\partial(x^\star_\tau/\sqrt{1-\tau})}{\partial(x^\star_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}})}.$$

Then there exist universal constants $M, C_\mathrm{s}, C_\mathrm{J} > 0$, independent of k, K, τ, d, T, θ where $M = \max\{M_1, M_2\}$ in proof, such that for every integer $0 \le k \le K$,

$$\left\| \frac{\partial^k}{\partial \tau^k} \left(\frac{s^\star_\tau(x^\star_\tau)}{(1-\tau)^{3/2}} \right) \right\|_2 \le C_\mathrm{s}\,(MK)^k k! \sqrt{\frac{d\theta \log T}{\tau(1-\tau)^3}} \left(\frac{d\theta \log T}{\tau(1-\tau)} \right)^k, \tag{56}$$

$$\left\| \frac{\partial^k}{\partial \tau^k} \left[\frac{1}{(1-\tau)^{3/2}} \frac{\partial s^\star_\tau(x^\star_\tau)}{\partial u^\star_\tau} J^\star_\tau \right] \right\| \le C_\mathrm{J}\,(MK)^{k+1} k! \left(\frac{d\theta \log T}{\tau(1-\tau)} \right)^{k+1}. \tag{57}$$

Proof. See Section C.4. ☐

Lemma 14. *Let $\theta_t := \theta_{\tau_{t,0}}(x_{\tau_{t,0}})$, where $\theta_{\tau_{t,0}}(x_{\tau_{t,0}})$ is defined in* (47). *Assume that*

$$C_{10} \left\{ \frac{\theta_t d \log^2 T}{T} + \frac{\sqrt{\theta_t d \log^3 T \log K \sum_{m,j} \left(\varepsilon^{(m)}_{\mathrm{score},t,j}\left(x^{(m)}_{\tau_{t,j}} \right) \right)^2}}{T} \right\} \le 1 \tag{58}$$

for some sufficiently large constant $C_{10} > 0$, where $\sum_{m,j} := \sum_{m=0}^{N} \sum_{j=0}^{K-1}$ and $\varepsilon^{(m)}_{\mathrm{score},t,j}(\cdot)$ is defined in (29)*. Then for all $0 \le i \le K-1$, $0 \le n \le N-1$, and all $\lambda \in [0,1]$,*

$$-\log p_{\bar{X}_{\tau_{t,i}}} \left(\lambda x^{(n+1)}_{\tau_{t,i}} + (1-\lambda) x^\star_{\tau_{t,i}} \right) \le 2.1\, d\theta_t \log T, \tag{59}$$

and

$$\log \frac{p_{\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,i}}}\bar{X}_{\tau_{t,i}}} \left(\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,i}}}\, x^{(n+1)}_{\tau_{t,i}} \right)}{p_{\bar{X}_{\tau_{t,0}}}(x_{\tau_{t,0}})} \le \frac{4c_1 d \log T}{T} + C_{10} \left\{ \frac{d^2 \theta_t^2 \log^4 TK}{T^2} + \frac{\sqrt{d\theta_t \log^3 T\, \log K \sum_{m,j} \left(\varepsilon^{(m)}_{\mathrm{score},t,j}\left(x^{(m)}_{\tau_{t,j}} \right) \right)^2}}{T} \right\}.$$

Proof. See Section C.5. ☐

Lemma 15. *Recall that q_t is the distribution of X_t, where X_t in* (7)*. If $-\log q_1(x_1) \le c_6 d \log T$ where c_6 is defined in* (47) *and $T \ge C_2 d \log^4 T$, then for all integers $1 \le \ell < \tau(x_T)$ where $\tau(x_T)$ is defined in* (38)*, one has*

$$-\log q_\ell(x_\ell) \le 2c_6 d \log T, \tag{60}$$

provided that $c_6 > 5c_1$.

Proof. The proof of Lemma 15 follows the same argument as that of Lemma 9 in Li et al. (2025), with the only modification being that we invoke Lemma 14 in place of the corresponding estimate used there. This change requires the conditions $T \ge C_2 d \log^4 T$ and $K \lesssim \log T$ to obtain the desired conclusion. We omit the details for brevity. ☐

B Proofs of lemmas in Section 5

For $\tau_{t-1,0} \le \tau < \tau_{t,0}$, we let $x^\star_\tau$ denote the solution of (8) with the initial condition $x^\star_{\tau_{t,0}} = x_{\tau_{t,0}}$. In particular, with iteration (15) we obtain

$$x^{(n)}_{\tau_{t,0}} = x^\star_{\tau_{t,0}} = x_{\tau_{t,0}}, \qquad \text{for all } n \ge 0.$$

For convenience, we define $A^{(t)}_{ji} := \frac{1}{2} \gamma_{t,j}(\tau_{t,i}) (1-\tau_{t,j})^{-3/2}$, where $\gamma_{t,j}(\tau_{t,i})$ is defined in (16).

B.1 Proof of Lemma 3

Proof. Fix $0 \le i \le K-1$. By the Gauss–Seidel update (15),

$$\frac{x_{\tau_{t,i}}^{(n+1)}}{\sqrt{1-\tau_{t,i}}} = \frac{x_{\tau_{t,0}}}{\sqrt{1-\tau_{t,0}}} + \sum_{j<i} A_{ji}^{(t)}\, s_{\tau_{t,j}}\left(x_{\tau_{t,j}}^{(n+1)}\right) + \sum_{j\ge i} A_{ji}^{(t)}\, s_{\tau_{t,j}}\left(x_{\tau_{t,j}}^{(n)}\right).$$

On the other hand, by integrating the rescaled probability flow ODE along the exact path, we obtain

$$\frac{x_{\tau_{t,i}}^{\star}}{\sqrt{1-\tau_{t,i}}} = \frac{x_{\tau_{t,0}}^{\star}}{\sqrt{1-\tau_{t,0}}} + \int_{\tau_{t,i}}^{\tau_{t,0}} \frac{s_\tau^\star(x_\tau^\star)}{2(1-\tau)^{3/2}} d\tau.$$

Subtracting the above two identities yields

$$\begin{aligned} x_{\tau_{t,i}}^{(n+1)} - x_{\tau_{t,i}}^{\star} &= \underbrace{\sqrt{1-\tau_{t,i}}\left[\sum_{j<i} A_{ji}^{(t)}\left(s_{\tau_{t,j}}\left(x_{\tau_{t,j}}^{(n+1)}\right) - s_{\tau_{t,j}}^\star\left(x_{\tau_{t,j}}^\star\right)\right) + \sum_{j\ge i} A_{ji}^{(t)}\left(s_{\tau_{t,j}}\left(x_{\tau_{t,j}}^{(n)}\right) - s_{\tau_{t,j}}^\star\left(x_{\tau_{t,j}}^\star\right)\right)\right]}_{:=G_1} \\ &\quad + \underbrace{\sqrt{1-\tau_{t,i}}\left[\sum_{j=0}^{K-1} A_{ji}^{(t)} s_{\tau_{t,j}}^\star(x_{\tau_{t,j}}^\star) - \int_{\tau_{t,i}}^{\tau_{t,0}} \frac{s_\tau^\star(x_\tau^\star)}{2(1-\tau)^{3/2}} d\tau\right]}_{:=G_2}. \end{aligned} \tag{61}$$

Bound for G_1. A simple calculation gives

$$\begin{aligned} G_1 &= \underbrace{\sqrt{1-\tau_{t,i}}\left[\sum_{j<i} A_{ji}^{(t)}\left(s_{\tau_{t,j}}\left(x_{\tau_{t,j}}^{(n+1)}\right) - s_{\tau_{t,j}}^\star\left(x_{\tau_{t,j}}^{(n+1)}\right)\right) + \sum_{j\ge i} A_{ji}^{(t)}\left(s_{\tau_{t,j}}\left(x_{\tau_{t,j}}^{(n)}\right) - s_{\tau_{t,j}}^\star\left(x_{\tau_{t,j}}^{(n)}\right)\right)\right]}_{:=G_{11}} \\ &\quad + \underbrace{\sqrt{1-\tau_{t,i}}\left[\sum_{j<i} A_{ji}^{(t)}\left(s_{\tau_{t,j}}^\star\left(x_{\tau_{t,j}}^{(n+1)}\right) - s_{\tau_{t,j}}^\star\left(x_{\tau_{t,j}}^\star\right)\right) + \sum_{j\ge i} A_{ji}^{(t)}\left(s_{\tau_{t,j}}^\star\left(x_{\tau_{t,j}}^{(n)}\right) - s_{\tau_{t,j}}^\star\left(x_{\tau_{t,j}}^\star\right)\right)\right]}_{:=G_{12}}. \end{aligned} \tag{62}$$

For the term G_{11}, the triangle inequality and Cauchy–Schwarz yield

$$\begin{aligned} \|G_{11}\|_2^2 &\le (1-\tau_{t,i})\Big(\sum_{j=0}^{K-1} |A_{ji}^{(t)}|^2\Big)\left[\sum_{j<i}\left(\varepsilon_{\text{score},t,j}^{(n+1)}(x_{\tau_{t,j}}^{(n+1)})\right)^2 + \sum_{j\ge i}\left(\varepsilon_{\text{score},t,j}^{(n)}(x_{\tau_{t,j}}^{(n)})\right)^2\right] \\ &\lesssim \log K \cdot \frac{(\tau_{t,0}-\tau_{t,i})^2}{(1-\tau_{t,0})^2}\left[\sum_{j=0}^{K-1}\left(\varepsilon_{\text{score},t,j}^{(n)}(x_{\tau_{t,j}}^{(n)})\right)^2 + \sum_{j=0}^{K-1}\left(\varepsilon_{\text{score},t,j}^{(n+1)}(x_{\tau_{t,j}}^{(n+1)})\right)^2\right] \\ &\lesssim \log K\, \tau_{t,i}^2 \frac{\log^2 T}{T^2}\left[\sum_{j=0}^{K-1}\left(\varepsilon_{\text{score},t,j}^{(n)}(x_{\tau_{t,j}}^{(n)})\right)^2 + \sum_{j=0}^{K-1}\left(\varepsilon_{\text{score},t,j}^{(n+1)}(x_{\tau_{t,j}}^{(n+1)})\right)^2\right], \end{aligned}$$

where the second inequality follows from (46k) and the fact $1-\tau_{t,i} \asymp 1-\tau_{t,0}$, the last inequality comes from (46f) that by setting $i_1 = 0$, $i_2 = i$, $i_3 = i$, and $i_4 = 0$.

For the term G_{12}, on the event E_t, it holds

$$\begin{aligned}\|G_{12}\|_2^2 \leq& 2\left(1-\tau_{t,i}\right)\left(\left\|\sum_{j<i} A_{ji}^{(t)}\left(s^\star_{\tau_{t,j}}\left(x^{(n+1)}_{\tau_{t,j}}\right)-s^\star_{\tau_{t,j}}\left(x^\star_{\tau_{t,j}}\right)\right)\right\|_2^2+\left\|\sum_{j\geq i} A_{ji}^{(t)}\left(s^\star_{\tau_{t,j}}\left(x^{(n)}_{\tau_{t,j}}\right)-s^\star_{\tau_{t,j}}\left(x^\star_{\tau_{t,j}}\right)\right)\right\|_2^2\right)\\ \lesssim&(1-\tau_{t,i})\left(\sum_{j=0}^{K-1}\left|A_{ji}^{(t)}\right|\right)^2\left(\frac{d\log T}{\tau_{t,0}}\right)^2\left(\max_{0\leq\ell\leq i-1}\left\|x^{(n+1)}_{\tau_{t,\ell}}-x^\star_{\tau_{t,\ell}}\right\|_2^2+\max_{0\leq\ell\leq K-1}\left\|x^{(n)}_{\tau_{t,\ell}}-x^\star_{\tau_{t,\ell}}\right\|_2^2\right)\\ \lesssim&\log^2 K\left(\frac{d\log^2 T}{T}\right)^2\left(\max_{0\leq\ell\leq i-1}\left\|x^{(n+1)}_{\tau_{t,\ell}}-x^\star_{\tau_{t,\ell}}\right\|_2^2+\max_{0\leq\ell\leq K-1}\left\|x^{(n)}_{\tau_{t,\ell}}-x^\star_{\tau_{t,\ell}}\right\|_2^2\right),\end{aligned}$$

where the third line follows from the typical region that $-\log p_{\bar{X}_{\tau_{t,j}}}\left(\lambda x^{(n)}_{\tau_{t,j}}+(1-\lambda)x^\star_{\tau_{t,j}}\right)\leq C_4 d\log T$, for all j,n together with Lemma 11, namely $\left\|s^\star_{\tau_{t,j}}\left(x^{(n)}_{\tau_{t,j}}\right)-s^\star_{\tau_{t,j}}\left(x^\star_{\tau_{t,j}}\right)\right\|_2\leq C\frac{d\log T}{\tau_{t,j}}\left\|x^{(n)}_{\tau_{t,j}}-x^\star_{\tau_{t,j}}\right\|_2$, and the last line holds by $\sum_{j=0}^{K-1}\left|A_{ji}^{(t)}\right|\leq C_4\frac{(\tau_{t,0}-\tau_{t,i})\log K}{(1-\tau_{t,0})^{3/2}}$ in Lemma 9 and $\tau_{t,0}\asymp\tau_{t,i}$.

Bound for G_2. For any $\tau_{t,i}$, the Lagrange remainder bound gives

$$\|G_2\|_2^2\leq(1-\tau_{t,i})\cdot\left(\int_{\tau_{t,i}}^{\tau_{t,0}}\frac{1}{K!}\sup_{\tau_{t,i}\leq\tau\leq\tau_{t,0}}\left\|\frac{\partial^K}{\partial\tau^K}\frac{s^\star_\tau(x^\star_\tau)}{2(1-\tau)^{3/2}}\right\|_2\left|\prod_{j=0}^{K-1}(\tau-\tau_{t,j})\right|d\tau\right)^2. \tag{63}$$

For all $\tau_{t,i}\leq\tau\leq\tau_{t,0}$, according to Lemma 13, it holds

$$\sup_{\tau_{t,i}\leq\tau\leq\tau_{t,0}}\left\|\frac{\partial^K}{\partial\tau^K}\frac{s^\star_\tau(x^\star_\tau)}{2(1-\tau)^{3/2}}\right\|_2\leq C_s(MK)^K K!\sqrt{\frac{d\theta\log T}{\tau_{t,i}(1-\tau_{t,i})^3}}\left(\frac{d\theta\log T}{\tau_{t,i}(1-\tau_{t,i})}\right)^K.$$

Next, set

$$\tau=\frac{\tau_{t,0}+\tau_{t,K-1}}{2}+\frac{\tau_{t,0}-\tau_{t,K-1}}{2}z,\qquad z\in[-1,1],$$

then

$$\prod_{j=0}^{K-1}(\tau-\tau_{t,j})=\left(\frac{\tau_{t,0}-\tau_{t,K-1}}{2}\right)^K\prod_{j=0}^{K-1}(z-z_j),$$

where

$$z_j=\cos\left(\frac{j\pi}{K-1}\right),\qquad 0\leq j\leq K-1.$$

Using the fact that

$$\prod_{j=0}^{K-1}(z-z_j)=2^{2-K}(z^2-1)U_{K-2}(z)\quad\text{and}\quad\sup_{z\in[-1,1]}\left|(z^2-1)U_{K-2}(z)\right|\leq 1,$$

where U_{K-2} is the Chebyshev polynomial of the second kind, we obtain

$$\sup_{\tau_{t,K-1}\leq\tau\leq\tau_{t,0}}\left|\prod_{j=0}^{K-1}(\tau-\tau_{t,j})\right|\leq 4^{1-K}(\tau_{t,0}-\tau_{t,K-1})^K. \tag{64}$$

Substituting the preceding two estimates into (63) yields

$$\begin{aligned}\|G_2\|_2^2 &\lesssim (\tau_{t,0}-\tau_{t,i})^2\frac{d\theta\log T}{\tau_{t,i}(1-\tau_{t,i})^2}\left(\frac{MKd\theta(\tau_{t,0}-\tau_{t,K-1})\log T}{4\tau_{t,i}(1-\tau_{t,i})}\right)^{2K}\\ &\lesssim \frac{d\theta\,\tau_{t,i}\log^3 T}{T^2}\left(\frac{c_1MKd\theta\log^2 T}{4T}\right)^{2K}\\ &\lesssim \frac{d\tau_{t,i}\log^3 T}{T^2}\left(\frac{M_0Kd\log^2 T}{T}\right)^{2K},\end{aligned}$$

where the second inequality comes from the fact that $\tau_{t,0} - \tau_{t,i} \le c_1 \tau_{t,i}(1-\tau_{t,i})\frac{\log T}{T}$ and $\tau_{t,0} - \tau_{t,K-1} \le c_1 \tau_{t,i}(1-\tau_{t,i})\frac{\log T}{T}$, and the last inequality follows from the typical condition that $\theta \le C_\theta$ for an absolute constant $C_\theta > 0$ and $M_0 \ge \frac{c_1 M C_\theta}{4}$.

Putting the estimates of G_{11}, G_{12}, and G_2 into (61), with $\tau_{t,0} \ge \tau_{t,i}, 0 \le i \le K-1$ we arrive at

$$\begin{aligned}
\|x^{(n+1)}_{\tau_{t,i}} - x^\star_{\tau_{t,i}}\|_2^2 \le{}& \bar{C}\log^2 K \left(\frac{d\log^2 T}{T}\right)^2 \left[\max_{0\le \ell\le i-1} \|x^{(n+1)}_{\tau_{t,\ell}} - x^\star_{\tau_{t,\ell}}\|_2^2 + \max_{0\le \ell \le K-1}\|x^{(n)}_{\tau_{t,\ell}} - x^\star_{\tau_{t,\ell}}\|_2^2\right] \\
&+ \bar{C}\log K\,\tau_{t,0}^2 \frac{\log^2 T}{T^2}\left[\sum_{j=0}^{K-1}\left(\varepsilon^{(n)}_{\text{score },t,j}\left(x^{(n)}_{\tau_{t,j}}\right)\right)^2 + \sum_{j=0}^{K-1}\left(\varepsilon^{(n+1)}_{\text{score },t,j}\left(x^{(n+1)}_{\tau_{t,j}}\right)\right)^2\right] \\
&+ \bar{C}\frac{d\tau_{t,0}\log^3 T}{T^2}\left(\frac{M_0 K d\log^2 T}{T}\right)^{2K}
\end{aligned} \tag{65}$$

for any fixed $0 \le i \le K-1$. Here, $\bar{C}$ is an universal constant. Note that $K \le c\log T$ and $T \ge C_2 d\log^4 T$ for a sufficiently large constant $C_2 > 0$. Therefore, it holds

$$\bar{C}\log^2 K\left(\frac{d\log^2 T}{T}\right)^2 \le \frac{1}{4}.$$

By induction, one can easily check that for every $0 \le i \le K-1$,

$$\begin{aligned}
\|x^{(n+1)}_{\tau_{t,i}} - x^\star_{\tau_{t,i}}\|_2^2 \le{}& \frac{1}{3}\max_{0\le\ell\le K-1}\|x^{(n)}_{\tau_{t,\ell}} - x^\star_{\tau_{t,\ell}}\|_2^2 + \frac{4\bar{C}}{3}\log K\,\tau_{t,0}^2\frac{\log^2 T}{T^2}\sum_{m=n}^{n+1}\sum_{j=0}^{K-1}\left(\varepsilon^{(m)}_{\text{score},t,j}(x^{(m)}_{\tau_{t,j}})\right)^2 \\
&+ \frac{4\bar{C}}{3}\frac{d\tau_{t,0}\log^3 T}{T^2}\left(\frac{M_0 K d \log^2 T}{T}\right)^{2K}.
\end{aligned}\tag{66}$$

Indeed, for $i = 0$, (66) holds immediate since $x^{(n+1)}_{\tau_{t,0}} = x^\star_{\tau_{t,0}} = x_{\tau_{t,0}}$. Next, assume that (66) holds for all $0 \le i \le k_0$, where $0 \le k_0 \le K-2$. Using the hypothesis conditions, one can see that (66) holds for $i = k_0+1$. With (66) in place, we obtain

$$\begin{aligned}
\max_{0\le r\le K-1}\|x^{(N)}_{\tau_{t,r}} - x^\star_{\tau_{t,r}}\|_2^2 \le{}& 3^{-N}\max_{0\le r\le K-1}\|x^{(0)}_{\tau_{t,r}} - x^\star_{\tau_{t,r}}\|_2^2 + C\log K\,\tau_{t,0}^2\frac{\log^2 T}{T^2}\sum_{m=0}^{N}\sum_{j=0}^{K-1}\left(\varepsilon^{(m)}_{\text{score},t,j}(x^{(m)}_{\tau_{t,j}})\right)^2 \\
&+ C\frac{d\tau_{t,0}\log^3 T}{T^2}\left(\frac{M_0 K d\log^2 T}{T}\right)^{2K}.
\end{aligned}\tag{67}$$

Here, $C = 3\bar{C}$. Finally, as shown by Li et al. (2025), one has

$$\|x^\star_{\tau_{t,i}} - x^{(0)}_{\tau_{t,i}}\|_2^2 \le \left(\bar{\alpha}_{t-1}\sqrt{1-\alpha_t}\sqrt{\frac{1-\alpha_t}{1-\bar{\alpha}_{t-1}}}\sqrt{d\log T}\right)^2 \le \frac{4c_1^2 d\log^3 T}{T^2}, \qquad 0 \le i \le K-1.$$

Since $T \ge C_2 d\log^4 T$, $K \le c\log T$, and $N \ge C_3 K\log T$ for a sufficiently large constant $C_3 > 0$, it holds

$$3^{-N}\max_{0\le r\le K-1}\|x^{(0)}_{\tau_{t,r}} - x^\star_{\tau_{t,r}}\|_2^2 \le C\frac{d\tau_{t,0}\log^3 T}{T^2}\left(\frac{M_0 K d\log^2 T}{T}\right)^{2K}.$$

Substituting this into (67) yields

$$\max_{0\le r\le K-1}\|x^{(N)}_{\tau_{t,r}} - x^\star_{\tau_{t,r}}\|_2^2 \le C\log K\,\tau_{t,0}^2\frac{\log^2 T}{T^2}\sum_{m=0}^{N}\sum_{j=0}^{K-1}\left(\varepsilon^{(m)}_{\text{score},t,j}(x^{(m)}_{\tau_{t,j}})\right)^2 + C\frac{d\tau_{t,0}\log^3 T}{T^2}\left(\frac{M_0 K d\log^2 T}{T}\right)^{2K}.$$

This completes the proof. $\square$

B.2 Proof of Lemma 4

Proof. Recall that

$$J^{(n)}_{\tau_{t,i}} := \frac{\partial(x^{(n)}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}})}{\partial(x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}})}, \qquad J^{\star}_{\tau_{t,i}} := \frac{\partial(x^{\star}_{\tau_{t,i}}/\sqrt{1-\tau_{t,i}})}{\partial(x^{\star}_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}})}.$$

It then follows from (15) that

$$J^{(n+1)}_{\tau_{t,i}} = I + \sum_{j<i} A^{(t)}_{ji} \frac{\partial s_{\tau_{t,j}}\left(x^{(n+1)}_{\tau_{t,j}}\right)}{\partial\left(x^{(n+1)}_{\tau_{t,j}}/\sqrt{1-\tau_{t,j}}\right)} J^{(n+1)}_{\tau_{t,j}} + \sum_{j\geq i} A^{(t)}_{ji} \frac{\partial s_{\tau_{t,j}}\left(x^{(n)}_{\tau_{t,j}}\right)}{\partial\left(x^{(n)}_{\tau_{t,j}}/\sqrt{1-\tau_{t,j}}\right)} J^{(n)}_{\tau_{t,j}}. \tag{68}$$

Similarly, note that the exact rescaled probability flow ODE is

$$\frac{x^{\star}_{\tau_{t,i}}}{\sqrt{1-\tau_{t,i}}} = \frac{x^{\star}_{\tau_{t,0}}}{\sqrt{1-\tau_{t,0}}} - \int_{\tau_{t,0}}^{\tau_{t,i}} \frac{s^{\star}_{\tau}(x^{\star}_{\tau})}{2(1-\tau)^{3/2}}\, d\tau.$$

Therefore,

$$J^{\star}_{\tau_{t,i}} = I - \int_{\tau_{t,0}}^{\tau_{t,i}} \frac{1}{2(1-\tau)^{3/2}} \frac{\partial s^{\star}_{\tau}(x^{\star}_{\tau})}{\partial\left(x^{\star}_{\tau}/\sqrt{1-\tau}\right)} J^{\star}_{\tau}\, d\tau. \tag{69}$$

Combining the two previous equations, we can obtain

$$\begin{aligned}
& J^{(n+1)}_{\tau_{t,i}} - J^{\star}_{\tau_{t,i}} \\
= \; & \sum_{j<i} A^{(t)}_{ji} \left[\frac{\partial s_{\tau_{t,j}}(x^{(n+1)}_{\tau_{t,j}})}{\partial(x^{(n+1)}_{\tau_{t,j}}/\sqrt{1-\tau_{t,j}})} J^{(n+1)}_{\tau_{t,j}} - \frac{\partial s^{\star}_{\tau_{t,j}}(x^{\star}_{\tau_{t,j}})}{\partial(x^{\star}_{\tau_{t,j}}/\sqrt{1-\tau_{t,j}})} J^{\star}_{\tau_{t,j}} \right] \\
& + \sum_{j\geq i} A^{(t)}_{ji} \left[\frac{\partial s_{\tau_{t,j}}(x^{(n)}_{\tau_{t,j}})}{\partial(x^{(n)}_{\tau_{t,j}}/\sqrt{1-\tau_{t,j}})} J^{(n)}_{\tau_{t,j}} - \frac{\partial s^{\star}_{\tau_{t,j}}(x^{\star}_{\tau_{t,j}})}{\partial(x^{\star}_{\tau_{t,j}}/\sqrt{1-\tau_{t,j}})} J^{\star}_{\tau_{t,j}} \right] \\
& + \int_{\tau_{t,0}}^{\tau_{t,i}} \left[\frac{1}{2(1-\tau)^{3/2}} \frac{\partial s^{\star}_{\tau}(x^{\star}_{\tau})}{\partial(x^{\star}_{\tau}/\sqrt{1-\tau})} J^{\star}_{\tau} - \sum_{j=0}^{K-1} \psi_{t,j}(\tau) \frac{1}{2(1-\tau_{t,j})^{3/2}} \frac{\partial s^{\star}_{\tau_{t,j}}(x^{\star}_{\tau_{t,j}})}{\partial(x^{\star}_{\tau_{t,j}}/\sqrt{1-\tau_{t,j}})} J^{\star}_{\tau_{t,j}} \right] d\tau \\
= \; & \underbrace{\sum_{j<i} A^{(t)}_{ji} \left(\frac{\partial s_{\tau_{t,j}}(x^{(n+1)}_{\tau_{t,j}})}{\partial(x^{(n+1)}_{\tau_{t,j}}/\sqrt{1-\tau_{t,j}})} - \frac{\partial s^{\star}_{\tau_{t,j}}(x^{\star}_{\tau_{t,j}})}{\partial(x^{\star}_{\tau_{t,j}}/\sqrt{1-\tau_{t,j}})} \right) J^{(n+1)}_{\tau_{t,j}}}_{H_1} \\
& + \underbrace{\sum_{j<i} A^{(t)}_{ji} \frac{\partial s^{\star}_{\tau_{t,j}}\left(x^{\star}_{\tau_{t,j}}\right)}{\partial\left(x^{\star}_{\tau_{t,j}}/\sqrt{1-\tau_{t,j}}\right)} \left(J^{(n+1)}_{\tau_{t,j}} - J^{\star}_{\tau_{t,j}}\right)}_{H_2} \\
& + \underbrace{\sum_{j\geq i} A^{(t)}_{ji} \left(\frac{\partial s_{\tau_{t,j}}\left(x^{(n)}_{\tau_{t,j}}\right)}{\partial\left(x^{(n)}_{\tau_{t,j}}/\sqrt{1-\tau_{t,j}}\right)} - \frac{\partial s^{\star}_{\tau_{t,j}}\left(x^{\star}_{\tau_{t,j}}\right)}{\partial\left(x^{\star}_{\tau_{t,j}}/\sqrt{1-\tau_{t,j}}\right)} \right) J^{(n)}_{\tau_{t,j}}}_{H_3} \\
& + \underbrace{\sum_{j\geq i} A^{(t)}_{ji} \frac{\partial s^{\star}_{\tau_{t,j}}\left(x^{\star}_{\tau_{t,j}}\right)}{\partial\left(x^{\star}_{\tau_{t,j}}/\sqrt{1-\tau_{t,j}}\right)} \left(J^{(n)}_{\tau_{t,j}} - J^{\star}_{\tau_{t,j}}\right)}_{H_4} \\
& + \underbrace{\int_{\tau_{t,0}}^{\tau_{t,i}} \left[\frac{1}{2(1-\tau)^{3/2}} \frac{\partial s^{\star}_{\tau}(x^{\star}_{\tau})}{\partial(x^{\star}_{\tau}/\sqrt{1-\tau})} J^{\star}_{\tau} - \sum_{j=0}^{K-1} \psi_{t,j}(\tau) \frac{1}{2(1-\tau_{t,j})^{3/2}} \frac{\partial s^{\star}_{\tau_{t,j}}(x^{\star}_{\tau_{t,j}})}{\partial(x^{\star}_{\tau_{t,j}}/\sqrt{1-\tau_{t,j}})} J^{\star}_{\tau_{t,j}} \right] d\tau}_{H_5}.
\end{aligned}$$

To begin with, we claim that on the event E_t, for all $0 \le n \le N$ and $0 \le j \le K-1$, it holds

$$\|J_{\tau_{t,j}}^{(n)}\| \le \widetilde{C} \tag{70}$$

for some universal constant $\widetilde{C} > 0$.

Bound for H_1 and H_3. Observe that

$$H_1 = \underbrace{\sum_{j<i} A_{ji}^{(t)} \left(\frac{\partial s_{\tau_{t,j}}\left(x_{\tau_{t,j}}^{(n+1)}\right)}{\partial\left(x_{\tau_{t,j}}^{(n+1)}/\sqrt{1-\tau_{t,j}}\right)} - \frac{\partial s_{\tau_{t,j}}^{\star}\left(x_{\tau_{t,j}}^{(n+1)}\right)}{\partial\left(x_{\tau_{t,j}}^{(n+1)}/\sqrt{1-\tau_{t,j}}\right)} \right) J_{\tau_{t,j}}^{(n+1)}}_{H_{11}}$$
$$+ \underbrace{\sum_{j<i} A_{ji}^{(t)} \left(\frac{\partial s_{\tau_{t,j}}^{\star}(x_{\tau_{t,j}}^{(n+1)})}{\partial(x_{\tau_{t,j}}^{(n+1)}/\sqrt{1-\tau_{t,j}})} - \frac{\partial s_{\tau_{t,j}}^{\star}(x_{\tau_{t,j}}^{\star})}{\partial(x_{\tau_{t,j}}^{\star}/\sqrt{1-\tau_{t,j}})} \right) J_{\tau_{t,j}}^{(n+1)}}_{H_{12}},$$

Note that from (70) that $\|J_{\tau_{t,j}}^{(n)}\| \le \widetilde{C}$. This gives

$$\begin{aligned} \|H_{11}\|^2 &\lesssim \left(\sum_{j=0}^{K-1} \left|A_{ji}^{(t)}\right|^2\right)(1-\tau_{t,0}) \left[\sum_{j<i}\left(\varepsilon_{\text{Jacobi }, t, j}^{(n+1)}\left(x_{\tau_{t,j}}^{(n+1)}\right)\right)^2 + \sum_{j\ge i}\left(\varepsilon_{\text{Jacobi }, t, j}^{(n)}\left(x_{\tau_{t,j}}^{(n)}\right)\right)^2\right] \\ &\le C\log K \tau_{t,i}^2 \frac{\log^2 T}{T^2} \sum_{m=n}^{n+1}\sum_{\ell=0}^{K-1}\left(\varepsilon_{\text{Jacobi},t,\ell}^{(m)}(x_{\tau_{t,\ell}}^{(m)})\right)^2, \end{aligned}$$

where the first inequality follows from the Cauchy-Schwarz inequality and the second inequality follows by (46k) in Lemma 9. For the term H_{12}, since E_t holds, Lemma 11 yields

$$\left\| \frac{\partial s_{\tau_{t,j}}^{\star}(x)}{\partial(x/\sqrt{1-\tau_{t,j}})} - \frac{\partial s_{\tau_{t,j}}^{\star}(y)}{\partial(y/\sqrt{1-\tau_{t,j}})} \right\| \le C\sqrt{1-\tau_{t,j}}\sqrt{\frac{d^3\log^3 T}{\tau_{t,j}^3}}\,\|x-y\|.$$

Therefore, we have

$$\begin{aligned} \|H_{12}\|^2 &\lesssim \left(\sum_{j=0}^{K-1}|A_{ji}^{(t)}|^2\right)\sum_{j<i}(1-\tau_{t,j})\frac{d^3\log^3 T}{\tau_{t,j}^3}\|x_{\tau_{t,j}}^{(n+1)} - x_{\tau_{t,j}}^{\star}\|_2^2 \\ &\lesssim \log K\,\frac{\log^2 T}{T^2}\frac{d^3\log^3 T}{\tau_{t,K-1}}\sum_{j<i}\|x_{\tau_{t,j}}^{(n+1)} - x_{\tau_{t,j}}^{\star}\|_2^2, \end{aligned}$$

Applying Lemma 3, one has

$$\begin{aligned} \|H_{12}\|^2 \le C'\frac{d^3\log^5 TK\log K}{T^2\tau_{t,K-1}} &\left(\frac{1}{3^n}4c_1^2\frac{d\log^3 T}{T^2} + C\log K\tau_{t,0}^2\frac{\log^2 T}{T^2}\sum_{m=0}^{n+1}\sum_{j=0}^{K-1}\left(\varepsilon_{\text{score},t,j}^{(m)}\left(x_{\tau_{t,j}}^{(m)}\right)\right)^2\right. \\ &\left. +C\frac{d\tau_{t,0}\log^3 T}{T^2}\left(\frac{M_0 K d\log^2 T}{T}\right)^{2K}\right). \end{aligned}$$

Combining the bounds for H_{11} and H_{12}, we get an upper bound for H_1. A similar bound for H_3 can be given, and we omit it.

Bound for H_2 and H_4. Again by Lemma 11 and the fact that $\tau_{t,j} \asymp \tau_{t,i}$, one has

$$\left\| \frac{\partial s^\star_{\tau_{t,j}}(x^\star_{\tau_{t,j}})}{\partial(x^\star_{\tau_{t,j}}/\sqrt{1-\tau_{t,j}})} \right\| \lesssim \sqrt{1-\tau_{t,i}} \frac{d \log T}{\tau_{t,i}}.$$

Therefore,

$$\begin{aligned} \|H_2\|^2 &\leq C' \left(\sum_{j<i} |A^{(t)}_{ji}|^2 \right) \sum_{j<i} (1-\tau_{t,j}) \frac{d^2 \log^2 T}{\tau^2_{t,j}} \|J^{(n+1)}_{\tau_{t,j}} - J^\star_{\tau_{t,j}}\|^2 \\ &\leq C' \log K \left(\frac{d \log^2 T}{T} \right)^2 K \max_{0 \leq j \leq i-1} \|J^{(n+1)}_{\tau_{t,j}} - J^\star_{\tau_{t,j}}\|^2, \end{aligned}$$

Similarly, we have

$$\|H_4\|^2 \leq C' \log K \left(\frac{d \log^2 T}{T} \right)^2 K \max_{0 \leq j \leq K-1} \|J^{(n)}_{\tau_{t,j}} - J^\star_{\tau_{t,j}}\|^2.$$

Bound for H_5. According to the Lagrange interpolation, one has

$$\begin{aligned} \|H_5\|^2 \quad &\leq \quad \left(\int_{\tau_{t,i}}^{\tau_{t,0}} \frac{1}{K!} \sup_{\tau_{t,i} \leq \tau \leq \tau_{t,0}} \left\| \frac{\partial^K}{\partial \tau^K} \frac{\partial s^\star_\tau(x^\star_\tau)}{\partial(x^\star_\tau/\sqrt{1-\tau})} J^\star_\tau \right\| \left| \prod_{j=0}^{K-1} (\tau - \tau_{t,j}) \right| d\tau \right)^2 \\ &\lesssim \quad \left(\frac{M_0 K d \log^2 T}{T} \right)^{2K+2}, \end{aligned}$$

where the second inequality comes from (64), the fact that $\tau \asymp \tau_{t,i}$ and $1-\tau \asymp 1-\tau_{t,i}$, and Lemma 13 that

$$\sup_{\tau_{t,K-1} \leq \tau \leq \tau_{t,0}} \left\| \frac{\partial^K}{\partial \tau^K} \left[\frac{1}{2(1-\tau)^{3/2}} \frac{\partial s^\star_\tau(x^\star_\tau)}{\partial(x^\star_\tau/\sqrt{1-\tau})} J^\star_\tau \right] \right\| \leq C_{\mathrm{J}} (MK)^{K+1} K! \left(\frac{d\theta \log T}{\tau_{t,i}(1-\tau_{t,i})} \right)^{K+1}.$$

Collect all the bounds. Combining the estimates for H_1, H_2, H_3, H_4 and H_5 with $\tau_{t,0} \geq \tau_{t,i}, 0 \leq i \leq K-1$, we obtain that for any fixed $0 \leq i \leq K-1$, it holds

$$\begin{aligned} \|J^{(n+1)}_{\tau_{t,i}} - J^\star_{\tau_{t,i}}\|^2 \leq C' K \log K \left(\frac{d \log^2 T}{T} \right)^2 & \left[\max_{0 \leq j \leq i-1} \|J^{(n+1)}_{\tau_{t,j}} - J^\star_{\tau_{t,j}}\|^2 + \max_{0 \leq j \leq K-1} \|J^{(n)}_{\tau_{t,j}} - J^\star_{\tau_{t,j}}\|^2 \right] \\ + C' \frac{d^3 \log^5 T K \log K}{T^2 \tau_{t,K-1}} & \left(\frac{1}{3^n} 4 c_1^2 \frac{d \log^3 T}{T^2} + C \log K \tau^2_{t,0} \frac{\log^2 T}{T^2} \sum_{m=0}^{n+1} \sum_{j=0}^{K-1} \left(\varepsilon^{(m)}_{\mathrm{score},t,j} \left(x^{(m)}_{\tau_{t,j}} \right) \right)^2 \right. \\ & \left. + C \frac{d \tau_{t,0} \log^3 T}{T^2} \left(\frac{M_0 K d \log^2 T}{T} \right)^{2K} \right) + C' \left(\frac{M_0 K d \log^2 T}{T} \right)^{2K+2} \\ + C' \log K \tau^2_{t,0} \frac{\log^2 T}{T^2} & \left[\sum_{\ell=0}^{K-1} \left(\varepsilon^{(n)}_{\mathrm{Jacobi},t,\ell} (x^{(n)}_{\tau_{t,\ell}}) \right)^2 + \sum_{\ell=0}^{K-1} \left(\varepsilon^{(n+1)}_{\mathrm{Jacobi},t,\ell} (x^{(n+1)}_{\tau_{t,\ell}}) \right)^2 \right]. \end{aligned}$$

By the same arguments to (66) and (67), we obtain

$$\begin{aligned} \max_{0 \leq i \leq K-1} \|J^{(N)}_{\tau_{t,i}} - J^\star_{\tau_{t,i}}\|^2 &\leq 7^{-N} \max_{0 \leq i \leq K-1} \|J^{(0)}_{\tau_{t,i}} - J^\star_{\tau_{t,i}}\|^2 + C \frac{d^3 \tau_{t,0} \log^7 T \log^2 K}{T^4} \sum_{m=0}^{N} \sum_{j=0}^{K-1} \left(\varepsilon^{(m)}_{\mathrm{score},t,j} \left(x^{(m)}_{\tau_{t,j}} \right) \right)^2 \\ &+ 2C \log K \, \tau^2_{t,0} \frac{\log^2 T}{T^2} \sum_{m=0}^{N} \sum_{j=0}^{K-1} \left(\varepsilon^{(m)}_{\mathrm{Jacobi},t,j} \left(x^{(m)}_{\tau_{t,j}} \right) \right)^2 + C \left(\frac{M_0 K d \log^2 T}{T} \right)^{2K+2}. \end{aligned} \tag{71}$$

We claim that

$$\max_{0\le j\le K-1}\left\|J^{(0)}_{\tau_{t,j}}-J^{\star}_{\tau_{t,j}}\right\|^2\le C\left(\frac{dK\log^2 T}{T}\right)^2. \tag{72}$$

Note that $T\ge C_2 d\log^4 T$, $K\le c\log T$ and $N\ge C_3 K\log T$ for some sufficiently large constant $C_3>0$. Combining (71) and (72) gives

$$\begin{aligned}\|J^{(N)}_{\tau_{t,i}}-J^{\star}_{\tau_{t,i}}\|^2 &\le CK\log^2 K\,\frac{d^3\log^7 T}{T^4}\sum_{n=0}^{N}\sum_{j=0}^{K-1}\Big(\varepsilon^{(n)}_{\mathrm{score},t,j}(x^{(n)}_{\tau_{t,j}})\Big)^2\\ &\quad + C\log K\tau_{t,0}^2\frac{\log^2 T}{T^2}\sum_{n=0}^{N}\sum_{j=0}^{K-1}\Big(\varepsilon^{(n)}_{\mathrm{Jacobi},t,j}(x^{(n)}_{\tau_{t,j}})\Big)^2+C\left(\frac{M_0Kd\log^2 T}{T}\right)^{2K+2}.\end{aligned}\tag{73}$$

This proves (31). It remains to prove the claims (70) and (72).

Proof of claim (72). Note that $x^{(0)}_{\tau_{t,j}}=x_{\tau_{t,0}},\quad 0\le j\le K-1$. It immediately gives

$$J^{(0)}_{\tau_{t,j}}=\frac{\partial(x^{(0)}_{\tau_{t,j}}/\sqrt{1-\tau_{t,j}})}{\partial(x_{\tau_{t,0}}/\sqrt{1-\tau_{t,0}})}=\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,j}}}\,I.$$

Using $1-\tau_{t,j}\asymp 1-\tau_{t,i}$, one has

$$\|J^{(0)}_{\tau_{t,j}}\|\le c_7,\qquad \|J^{(0)}_{\tau_{t,j}}-I\|=\left|\frac{1-\tau_{t,0}}{1-\tau_{t,j}}-1\right|=\frac{\tau_{t,0}-\tau_{t,j}}{1-\tau_{t,j}}\le c_8\frac{\log T}{T}, \tag{74}$$

where $c_7,c_8>0$ are universal constants. For convenience, define

$$A(\tau):=\frac{1}{2(1-\tau)^{3/2}}\frac{\partial s^{\star}_{\tau}(x^{\star}_{\tau})}{\partial(x^{\star}_{\tau}/\sqrt{1-\tau})}.$$

Similar to (69), we have

$$J^{\star}_{\tau}=I+\int_{\tau}^{\tau_{t,0}}A(u)J^{\star}_u\,du,\qquad \tau\in[\tau_{t,K-1},\tau_{t,0}]. \tag{75}$$

According to Lemma 13 with $k=0$, one has

$$\|A(u)\|\le C_J M\frac{dK\log T}{u(1-u)},\qquad u\in[\tau_{t,K-1},\tau_{t,0}].$$

Therefore, we obtain, for every $\tau\in[\tau_{t,K-1},\tau_{t,0}]$,

$$\|J^{\star}_{\tau}\|\le 1+\int_{\tau}^{\tau_{t,0}}C_JM\frac{dK\log T}{u(1-u)}\|J^{\star}_u\|\,du\le\exp\!\left(C_JM\int_{\tau}^{\tau_{t,0}}\frac{dK\log T}{u(1-u)}\,du\right),$$

where the last inequality comes from Gronwall's inequality. Using the fact that $u\asymp\tau_{t,0}$ and $1-u\asymp 1-\tau_{t,0}$ on $[\tau_{t,K-1},\tau_{t,0}]$, together with (46f) in Lemma 9, we have

$$\int_{\tau}^{\tau_{t,0}}\frac{dK\log T}{u(1-u)}\,du\le c_1\frac{dK\log^2 T}{T}.$$

Since $T\ge C_2 d\log^4 T$ and $K\le c\log T$, it implies

$$\sup_{\tau\in[\tau_{t,K-1},\tau_{t,0}]}\|J^{\star}_{\tau}\|\le\exp\left(C_JMc_1\frac{dK\log^2 T}{T}\right)\le c_8. \tag{76}$$

Substituting (76) into (75) gives

$$\|J^{\star}_{\tau_{t,j}}-I\|\le\int_{\tau_{t,j}}^{\tau_{t,0}}\|A(u)\|\,\|J^{\star}_u\|\,du\le c_8\int_{\tau_{t,j}}^{\tau_{t,0}}C_JM\frac{dK\log T}{u(1-u)}\,du\le c_9\frac{dK\log^2 T}{T}\le\frac{1}{2} \tag{77}$$

for all $0 \le j \le K-1$, where c_9 is an universal constant. Finally, by the triangle inequality and using (74),

$$\|J^{(0)}_{\tau_{t,j}} - J^{\star}_{\tau_{t,j}}\| \le \|J^{(0)}_{\tau_{t,j}} - I\| + \|J^{\star}_{\tau_{t,j}} - I\| \le c_8 \frac{\log T}{T} + c_9 \frac{dK \log^2 T}{T} \le c_{10} \frac{dK \log^2 T}{T}$$

holds for all $0 \le j \le K-1$, where $c_{10} > 0$ is an universal constant. This proves the claim (72).

Proof of claim (70)**.** The proof is inductive in nature. When $n = 0$, through (74), it holds $\|J^{(0)}_{\tau_{t,i}}\| \le \widetilde{C}$ for any $0 \le i \le K-1$. Next, assume that $\|J^{(n_0)}_{\tau_{t,i}}\| \le \widetilde{C}$ holds for all $0 \le i \le K-1$. We then prove $\|J^{(n_0+1)}_{\tau_{t,i}}\| \le \widetilde{C}$ holds for all $0 \le i \le K-1$ by induction again. For the case $i = 0$, we know $J^{(n_0+1)}_{\tau_{t,0}} = I$. Now assume that $\|J^{(n_0+1)}_{\tau_{t,i}}\| \le \widetilde{C}$ holds for all $1 \le i \le k_0$. According to (68) and using the same approaches as before, we have

$$\begin{aligned}
\|J^{(n_0+1)}_{\tau_{t,k_0+1}} - I\| &\le \left\| \sum_{j<k_0+1} A^{(t)}_{ji} \frac{\partial s^{\star}_{\tau_{t,j}}(x^{(n_0+1)}_{\tau_{t,j}})}{\partial (x^{(n_0+1)}_{\tau_{t,j}} / \sqrt{1-\tau_{t,j}})} J^{(n_0+1)}_{\tau_{t,j}} + \sum_{j \ge k_0+1} A^{(t)}_{ji} \frac{\partial s^{\star}_{\tau_{t,j}}(x^{(n_0)}_{\tau_{t,j}})}{\partial (x^{(n_0)}_{\tau_{t,j}} / \sqrt{1-\tau_{t,j}})} J^{(n_0)}_{\tau_{t,j}} \right\| \\
&\quad + \left\| \sum_{j<k_0+1} A^{(t)}_{ji} \left(\frac{\partial s_{\tau_{t,j}}(x^{(n_0+1)}_{\tau_{t,j}})}{\partial (x^{(n_0+1)}_{\tau_{t,j}} / \sqrt{1-\tau_{t,j}})} - \frac{\partial s^{\star}_{\tau_{t,j}}(x^{(n_0+1)}_{\tau_{t,j}})}{\partial (x^{(n_0+1)}_{\tau_{t,j}} / \sqrt{1-\tau_{t,j}})} \right) J^{(n_0+1)}_{\tau_{t,j}} \right\| \\
&\quad + \left\| \sum_{j \ge k_0+1} A^{(t)}_{ji} \left(\frac{\partial s_{\tau_{t,j}}(x^{(n_0)}_{\tau_{t,j}})}{\partial (x^{(n_0)}_{\tau_{t,j}} / \sqrt{1-\tau_{t,j}})} - \frac{\partial s^{\star}_{\tau_{t,j}}(x^{(n_0)}_{\tau_{t,j}})}{\partial (x^{(n_0)}_{\tau_{t,j}} / \sqrt{1-\tau_{t,j}})} \right) J^{(n_0)}_{\tau_{t,j}} \right\| \\
&\lesssim \frac{d \log^2 T \log K}{T} + \sqrt{\log K}\, \tau_{t,i} \frac{\log T}{T} \left(\sum_{m=n_0}^{n_0+1} \sum_{j=0}^{K-1} \left(\varepsilon^{(m)}_{\mathrm{Jacobi},t,j}(x^{(m)}_{\tau_{t,j}}) \right)^2 \right)^{1/2}.
\end{aligned}$$

On the event E_t and note that $T \ge C_2 d \log^4 T$, one has $\|J^{(n_0+1)}_{\tau_{t,k_0+1}} - I\| \le c'$ for a sufficiently small absolute constant $c' > 0$, which gives $\|J^{(n_0+1)}_{\tau_{t,k_0+1}}\| \le \widetilde{C}$. By induction, we have

$$\|J^{(n)}_{\tau_{t,i}}\| \le \widetilde{C}, \qquad 0 \le i \le K-1, \quad 0 \le n \le N,$$

which gives claim (70).

Next, we turn to prove the second part of Lemma 4, namely, to prove (30). From (73), and on the event E_t, we have

$$\|J^{(N)}_{\tau_{t,K-1}} - J^{\star}_{\tau_{t,K-1}}\| \le \frac{1}{4}$$

Combining this with (77), we obtain

$$\|J^{(N)}_{\tau_{t,K-1}} - I\| \le \|J^{(N)}_{\tau_{t,K-1}} - J^{\star}_{\tau_{t,K-1}}\| + \|J^{\star}_{\tau_{t,K-1}} - I\| \le \frac{3}{4}.$$

Hence $J^{(N)}_{\tau_{t,K-1}}$ is invertible, and by the Neumann-series bound,

$$\left\| \left(J^{(N)}_{\tau_{t,K-1}} \right)^{-1} \right\| \le \frac{1}{1 - \|J^{(N)}_{\tau_{t,K-1}} - I\|} \le 4.$$

This proves (30) and completes the proof.

$\square$

B.3 Proof of Lemma 5

Proof. From the definition of the total variation distance and set $\mathcal{E}$ in (34), we have

$$\begin{aligned}
\mathrm{TV}(q_1,p_1) &= \int_{\{x\in\mathbb{R}^d: q_1(x)>p_1(x)\}} (q_1(x)-p_1(x))\,dx \\
&= \int_{\mathcal{E}} (q_1(x)-p_1(x))\,dx + \int_{\{x\in\mathbb{R}^d: p_1(x)<q_1(x)\le\exp(-c_6 d\log T)\}} (q_1(x)-p_1(x))\,dx \\
&\le \int_{\mathcal{E}} (q_1(x)-p_1(x))\,dx + \int_{\{x\in\mathbb{R}^d: q_1(x)\le\exp(-c_6 d\log T)\}} q_1(x)\,dx \\
&=: \mathbb{E}_{Y_1\sim p_1}\left[\left(\frac{q_1(Y_1)}{p_1(Y_1)}-1\right)\mathbf{1}\{Y_1\in\mathcal{E}\}\right] + r_T.
\end{aligned}$$

For the term r_T, let $K_1>0$ be any fixed constant, one has

$$\begin{aligned}
r_T &\le \int_{\{x\in\mathbb{R}^d:\ \|x\|_2\le T^{c_R+K_1}\sqrt{d},\ q_1(x)\le\exp(-c_6 d\log T)\}} q_1(x)\,dx + \int_{\{x\in\mathbb{R}^d:\ \|x\|_2> T^{c_R+K_1}\sqrt{d}\}} q_1(x)\,dx \\
&\le \exp(-c_6 d\log T)\ \mathrm{Vol}\Big(\Big\{x\in\mathbb{R}^d:\|x\|_2\le T^{c_R+K_1}\sqrt{d}\Big\}\Big) + \mathbb{P}_{X_1\sim q_1}\Big(\|X_1\|_2>T^{c_R+K_1}\sqrt{d}\Big), \qquad (78)
\end{aligned}$$

where c_R is the constant defined in Assumption 1, and $\mathrm{Vol}(\cdot)$ denotes the d-dimensional Lebesgue measure. Recall that

$$X_1=\sqrt{\bar\alpha_1}\,X_0+\sqrt{1-\bar\alpha_1}\,Z,\qquad Z\sim\mathcal{N}(0,I_d),$$

and hence

$$\mathbb{E}\big[\|X_1\|_2^2\big]\le 2\bar\alpha_1\mathbb{E}\big[\|X_0\|_2^2\big]+2(1-\bar\alpha_1)\mathbb{E}\big[\|Z\|_2^2\big]\le 2T^{2c_R}+2d,$$

where we used Assumption 1 in the last inequality. Therefore, by Markov's inequality,

$$\mathbb{P}_{X_1\sim q_1}\Big(\|X_1\|_2>T^{c_R+K_1}\sqrt{d}\Big)\le\frac{\mathbb{E}\big[\|X_1\|_2^2\big]}{d\cdot T^{2c_R+2K_1}}\le\frac{2T^{2c_R}+2d}{d\cdot T^{2c_R+2K_1}}\lesssim T^{-2K_1}. \qquad (79)$$

By the standard formula for the volume of the d-dimensional Euclidean ball,

$$\mathrm{Vol}\left(\{x\in\mathbb{R}^d:\|x\|_2\le r\}\right)=\frac{\pi^{d/2}}{\Gamma(d/2+1)}r^d,$$

see, e.g., (Ball, 1997, Lecture 1). Moreover, by the Stirling-type lower bound for the Gamma function, see (Abramowitz & Stegun, 1964, p. 257, Eq. (6.1.38)), there exists an absolute constant $c>0$ such that

$$\Gamma(z+1)\ge c\sqrt{z}\left(\frac{z}{e}\right)^z,\qquad z\ge 1.$$

Taking $z=d/2$, we obtain, for all $d\ge 2$,

$$\mathrm{Vol}\left(\{x\in\mathbb{R}^d:\|x\|_2\le r\}\right)\le\frac{\sqrt{2}}{c\sqrt{d}}\left(\frac{2\pi e}{d}\right)^{d/2}r^d\le\left(\frac{Cr}{\sqrt{d}}\right)^d.$$

Hence, taking $r=T^{c_R+K_1}\sqrt{d}$, we get

$$\mathrm{Vol}\Big(\{x\in\mathbb{R}^d:\|x\|_2\le T^{c_R+K_1}\sqrt{d}\}\Big)\le\big(CT^{c_R+K_1}\big)^d.$$

Consequently,

$$\begin{aligned}
\exp(-c_6 d\log T)\ \mathrm{Vol}\Big(\{x\in\mathbb{R}^d:\|x\|_2\le T^{c_R+K_1}\sqrt{d}\}\Big) &\le \exp\big(-(c_6-c_R-K_1)d\log T+d\log C\big) \\
&\le \exp(-c_2\,d\log T), \qquad (80)
\end{aligned}$$

where the last inequality follows by choosing c_6 sufficiently large so that $c_6>c_R+K_1+1$, and c_2 is a constant. Putting (79) and (80) into (78), we obtain

$$r_T\le CT^{-2K_1}+\exp(-c_2\,d\log T),$$

where C is a constant. This completes the proof. □

C Proofs of auxiliary lemmas

C.1 Proof of Lemma 8

Proof. Let

$$\phi(z) := \frac{\tau_{t,0} + \tau_{t,K-1}}{2} + \frac{\tau_{t,0} - \tau_{t,K-1}}{2}\, z, \qquad z \in [-1,1].$$

Then ϕ maps $[-1,1]$ bijectively onto $[\tau_{t,K-1}, \tau_{t,0}]$, and $\phi(z_j) = \tau_{t,j}$ for all $0 \le j \le K-1$. Let ℓ_j denote the Lagrange basis polynomials associated with the standard Chebyshev–Lobatto nodes $\{z_j\}_{j=0}^{K-1}$ on $[-1,1]$, namely,

$$\ell_j(z) := \frac{\prod_{j':\, j' \neq j}(z - z_{j'})}{\prod_{j':\, j' \neq j}(z_j - z_{j'})}, \qquad 0 \le j \le K-1. \tag{81}$$

By the definition of $\psi_{t,j}$ as in (14), a direct calculation gives

$$\psi_{t,j}(\phi(z)) = \ell_j(z), \qquad 0 \le j \le K-1. \tag{82}$$

Therefore,

$$\sup_{\tau \in [\tau_{t,K-1}, \tau_{t,0}]} \sum_{j=0}^{K-1} |\psi_{t,j}(\tau)| = \sup_{z \in [-1,1]} \sum_{j=0}^{K-1} |\ell_j(z)|.$$

For the standard Chebyshev–Lobatto nodes on $[-1,1]$, the classical bound

$$\sup_{z \in [-1,1]} \sum_{j=0}^{K-1} |\ell_j(z)| \le \frac{2}{\pi} \log K + 1$$

holds; see (Trefethen, 2019, Chapter 15, Theorem 15.2). This completes the proof. □

C.2 Proof of Lemma 9

Proof. We prove the lemma one by one. First, (46a)–(46e) follow directly from Lemma 1 in Li et al. (2025).

Proof of (46f)**.** For any fixed $2 \le t \le T$, it holds $\tau_{t,K-1} = 1 - \bar{\alpha}_{t-1} \le \tau_{t,i} \le \tau_{t,0} = 1 - \bar{\alpha}_t$ for all $i = 0, 1, \ldots, K-1$. Therefore, for any $0 \le i_1, i_2, i_3, i_4 \le K-1$,

$$\left| \frac{\tau_{t,i_1} - \tau_{t,i_2}}{\tau_{t,i_3}(1 - \tau_{t,i_4})} \right| \le \frac{\bar{\alpha}_{t-1} - \bar{\alpha}_t}{(1 - \bar{\alpha}_{t-1})\bar{\alpha}_t} = \frac{1}{\alpha_t} \cdot \frac{1 - \alpha_t}{1 - \bar{\alpha}_{t-1}} \le 8c_1 \frac{\log T}{T},$$

where the last inequality comes from (46a) and (46b). This proves (46f).

Proof of (46g)**.** In the following, cause $K \ge 2$ and $0 \le j \le K-1$. From the definition of $\gamma_{t,j}(\cdot)$ in (16), one has

$$|\gamma_{t,j}(\tau_{t,i})| \le \int_{\tau_{t,i}}^{\tau_{t,0}} |\psi_{t,j}(\tau)|\, d\tau \le (\tau_{t,0} - \tau_{t,i}) \sup_{\tau \in [\tau_{t,K-1},\, \tau_{t,0}]} |\psi_{t,j}(\tau)| = (\tau_{t,0} - \tau_{t,i}) \sup_{z \in [-1,1]} |\ell_j(z)|, \tag{83}$$

where the last equation comes from (82). It then suffices to upper bound $\sup_{z \in [-1,1]} |\ell_j(z)|$. Let

$$\omega_K(z) := \prod_{j=0}^{K-1} (z - z_j), \quad z_j = \cos\left(\frac{j\pi}{K-1}\right), \qquad 0 \le j \le K-1.$$

Then with (81), we obtain

$$\ell_j(z) = \frac{\omega_K(z)}{\omega_K'(z_j)(z - z_j)}.$$

According to Mason & Handscomb (2002, Chapter 1, §1.2.2), one has

$$\omega_K(z) = 2^{1-(K-1)}(z^2-1)U_{K-2}(z),$$

where U_{K-2} is the Chebyshev polynomial of the second kind, characterized by $U_{K-2}(\cos\theta) = \frac{\sin((K-1)\theta)}{\sin\theta}$, $\theta \in [0,\pi]$. A direct computation gives

$$|\ell_0(z)| = \frac{(1+z)|U_{K-2}(z)|}{2(K-1)}, \qquad |\ell_{K-1}(z)| = \frac{(1-z)|U_{K-2}(z)|}{2(K-1)},$$

and, for $1 \le j \le K-2$,

$$|\ell_j(z)| = \frac{(1-z^2)|U_{K-2}(z)|}{(K-1)|z-z_j|}.$$

Now write $z = \cos\theta, z_j = \cos\theta_j, \theta \in [0,\pi], \theta_j = \frac{j\pi}{K-1}$. Using

$$U_{K-2}(\cos\theta) = \frac{\sin((K-1)\theta)}{\sin\theta},$$

we estimate $\sup_{z\in[-1,1]} |\ell_j(z)|$ separately:

Endpoint case $j=0$. We have

$$|\ell_0(z)| = \frac{(1+\cos\theta)|\sin((K-1)\theta)|}{2(K-1)\sin\theta} \le \frac{(1+\cos\theta)\,(K-1)\sin\theta}{2(K-1)\sin\theta} = \frac{1+\cos\theta}{2} \le 1.$$

Endpoint case $j = K-1$. Similarly,

$$|\ell_{K-1}(z)| = \frac{(1-\cos\theta)|\sin((K-1)\theta)|}{2(K-1)\sin\theta} \le \frac{1-\cos\theta}{2} \le 1.$$

Interior case $1 \le j \le K-2$. In this case,

$$|\ell_j(z)| = \frac{\sin\theta\,|\sin((K-1)\theta)|}{(K-1)|\cos\theta-\cos\theta_j|} \le \frac{\sin\theta\,|\sin(\theta-\theta_j)|}{|\cos\theta-\cos\theta_j|} \le \frac{\sin\theta\left|\cos\frac{\theta-\theta_j}{2}\right|}{\sin\frac{\theta+\theta_j}{2}}.$$

where the first inequality follows from the fact that $(K-1)\theta_j = j\pi$, and hence

$$|\sin((K-1)\theta)| = |\sin((K-1)(\theta-\theta_j))| \le (K-1)\,|\sin(\theta-\theta_j)|.$$

On the other hand, note that

$$\sin\theta + \sin\theta_j = 2\sin\frac{\theta+\theta_j}{2}\cos\frac{\theta-\theta_j}{2} \ge \sin\theta,$$

because $\theta, \theta_j \in [0,\pi]$ implies $\cos\frac{\theta-\theta_j}{2} \ge 0$, $\sin\theta_j \ge 0$. This immediately gives

$$|\ell_j(z)| \le 2\cos^2\frac{\theta-\theta_j}{2} \le 2.$$

Putting it into (83), we have

$$|\gamma_{t,j}(\tau_{t,i})| \le 2(\tau_{t,0} - \tau_{t,i}),$$

which proves (46g)

Proof of (46h). For every $0 \le i \le K-1$ and $2 \le t \le T$, since $\tau_{t,K-1} = 1-\bar{\alpha}_{t-1} \le \tau_{t,i} \le \tau_{t,0} = 1-\bar{\alpha}_t$, we have $\bar{\alpha}_t = 1-\tau_{t,0} \le 1-\tau_{t,i} \le 1-\tau_{t,K-1} = \bar{\alpha}_{t-1}$, and $1-\bar{\alpha}_{t-1} = \tau_{t,K-1} \le \tau_{t,i} \le \tau_{t,0} = 1-\bar{\alpha}_t$. Therefore, it suffices to show

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

which follow directly from (46a) and (46c). This proves (46h).

Proof of (46i). For $K \ge 2$, by Lemma 8, we have

$$\sum_{j=0}^{K-1} |\psi_{t,j}(\tau)| \le \frac{2}{\pi}\log K + 1 \le C_4 \log K.$$

Hence, for any $0 \le i \le K-1$,

$$\sum_{j=0}^{K-1} |\gamma_{t,j}(\tau_{t,i})| \le \int_{\tau_{t,i}}^{\tau_{t,0}} \sum_{j=0}^{K-1} |\psi_{t,j}(\tau)|\, d\tau \le C_4(\tau_{t,0}-\tau_{t,i})\log K,$$

where $\gamma_{t,j}$ is defined in (16). This proves(46i).

Proof of (46j). Using the definition of $A_{ji}^{(t)} = \frac{1}{2}\gamma_{t,j}(\tau_{t,i})(1-\tau_{t,j})^{-3/2}$ together with (46h) and (46i), we obtain

$$\sum_{j=0}^{K-1} |A_{ji}^{(t)}| = \frac{1}{2}\sum_{j=0}^{K-1} \frac{|\gamma_{t,j}(\tau_{t,i})|}{(1-\tau_{t,j})^{3/2}} \le \frac{1}{2}\frac{1}{(1-\tau_{t,0})^{3/2}}\sum_{j=0}^{K-1} |\gamma_{t,j}(\tau_{t,i})| \le C_4 \frac{(\tau_{t,0}-\tau_{t,i})\log K}{(1-\tau_{t,0})^{3/2}}.$$

This proves (46j).

Proof of (46k). According to (46g), one has $|\gamma_{t,j}(\tau_{t,i})| \le 2(\tau_{t,0}-\tau_{t,i})$. Therefore,

$$|A_{ji}^{(t)}| = \frac{1}{2}\frac{|\gamma_{t,j}(\tau_{t,i})|}{(1-\tau_{t,j})^{3/2}} \le C_4 \frac{\tau_{t,0}-\tau_{t,i}}{(1-\tau_{t,0})^{3/2}},$$

which implies

$$\sum_{j=0}^{K-1} |A_{ji}^{(t)}|^2 \le \Big(\max_{0\le j\le K-1} |A_{ji}^{(t)}|\Big)\sum_{j=0}^{K-1} |A_{ji}^{(t)}| \le C_5 \frac{(\tau_{t,0}-\tau_{t,i})^2\log K}{(1-\tau_{t,0})^3}.$$

This proves (46k). □

C.3 Proof of Lemma 10

Proof. For any $r > 0$, Bayes' rule gives

$$\begin{aligned}
\mathbb{P}\big(\big\|\sqrt{1-\tau}X_0 - y\big\|_2 \ge r \,\big|\, \bar{X}_\tau = y\big) &= \frac{\int_{\{\|\sqrt{1-\tau}x_0 - y\|_2 \ge r\}} p_{X_0}(x_0)\, p_{\bar{X}_\tau|X_0}(y \mid x_0)\, dx_0}{p_{\bar{X}_\tau}(y)} \\
&\le \frac{1}{p_{\bar{X}_\tau}(y)} \cdot \frac{1}{(2\pi\tau)^{d/2}}\exp\Big(-\frac{r^2}{2\tau}\Big) \\
&\le \exp\Big(\theta_\tau(y) d\log T + \frac{c_0}{2}d\log T - \frac{r^2}{2\tau}\Big), \qquad (84)
\end{aligned}$$

where the first inequality follows from the fact that

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

and the second inequality comes from the definition of $\theta_\tau(y)$ that $\frac{1}{p_{\bar{X}_\tau}(y)} \le \exp(\theta_\tau(y) d \log T)$ and the fact $\tau \ge T^{-c_0}$. Choosing

$$r = 5c_5 \sqrt{\theta_\tau(y)\, d\, \tau \log T}$$

for some constant $c_5 \ge 2$, then

$$\mathbb{P}\Big(\big\|\sqrt{1-\tau}X_0 - y\big\|_2 \ge 5c_5\sqrt{\theta_\tau(y)\, d\, \tau \log T} \;\Big|\; \bar{X}_\tau = y\Big) \le \exp\bigg(\bigg[1 + \frac{c_0}{2\theta_\tau(y)} - \frac{25}{2}c_5^2\bigg]\theta_\tau(y) d \log T\bigg).$$

Since $\theta_\tau(y) \ge c_6 \ge c_0$, we have

$$\frac{c_0}{2\theta_\tau(y)} \le \frac{1}{2},$$

and therefore

$$1 + \frac{c_0}{2\theta_\tau(y)} - \frac{25}{2}c_5^2 \le \frac{3}{2} - \frac{25}{2}c_5^2 \le -c_5^2 \qquad \text{for all } c_5 \ge 2.$$

Thus

$$\mathbb{P}\Big(\big\|\sqrt{1-\tau}X_0 - y\big\|_2 \ge 5c_5\sqrt{\theta_\tau(y)\, d\, \tau \log T} \;\Big|\; \bar{X}_\tau = y\Big) \le \exp\big(-c_5^2\theta_\tau(y) d \log T\big). \tag{85}$$

This gives (48). Next, note that $\theta_\tau(y) \ge c_6 \ge c_0$. It then follows from (84) that

$$\mathbb{P}\big(\big\|\sqrt{1-\tau}X_0 - y\big\|_2 \ge u \,\big|\, \bar{X}_\tau = y\big) \le \exp\bigg(\frac{3}{2}\,\theta_\tau(y) d \log T - \frac{u^2}{2\tau}\bigg), \qquad \forall u > 0.$$

For every integer $k \ge 1$, the tail-integral identity yields

$$\mathbb{E}\Big[\big\|\sqrt{1-\tau}X_0 - y\big\|_2^k \;\Big|\; \bar{X}_\tau = y\Big] = k\int_0^\infty u^{k-1}\mathbb{P}\big(\big\|\sqrt{1-\tau}X_0 - y\big\|_2 \ge u \,\big|\, \bar{X}_\tau = y\big)\, du.$$

In particular, for $k = 1$, we have

$$\begin{aligned}
& \mathbb{E}\big[\big\|\sqrt{1-\tau}X_0 - y\big\|_2 \,\big|\, \bar{X}_\tau = y\big] \\
\le\;& 10\sqrt{\theta_\tau(y)\, d\, \tau \log T} + \int_{10\sqrt{\theta_\tau(y)\, d\, \tau \log T}}^\infty \mathbb{P}\big(\big\|\sqrt{1-\tau}X_0 - y\big\|_2 \ge u \,\big|\, \bar{X}_\tau = y\big)\, du \\
\le\;& 10\sqrt{\theta_\tau(y)\, d\, \tau \log T} + \int_{10\sqrt{\theta_\tau(y)\, d\, \tau \log T}}^\infty \exp\bigg(\frac{3}{2}\,\theta_\tau(y) d \log T - \frac{u^2}{2\tau}\bigg)\, du \\
\lesssim\;& \sqrt{\theta_\tau(y)\, d\, \tau \log T},
\end{aligned}$$

where the last inequality comes from the Gaussian tail estimate

$$\int_a^\infty e^{-u^2/(2\tau)}\, du \le \frac{\tau}{a}e^{-a^2/(2\tau)}, \qquad \text{for all} \quad a > 0.$$

This proves (48). For the cases $k = 2, 3, 4$, the proof is similar, and so we omit it. $\square$

C.4 Proof of Lemma 13

Proof. The proof follows the argument used for Claims (64a) and (64b) in Appendix A of Li et al. (2025), while keeping track of the dependence on the order k explicitly. Define

$$u_k := \frac{\partial^k}{\partial \tau^k}\left(\frac{s_\tau^\star(x_\tau^\star)}{(1-\tau)^{3/2}}\right), \qquad \theta := \theta_\tau(x_\tau^\star).$$

Then for sufficiently small δ, the Taylor expansion gives

$$\frac{s_{\tau+\delta}^\star\left(x_{\tau+\delta}^\star\right)}{(1-\tau-\delta)^{3/2}} - \frac{s_\tau^\star\left(x_\tau^\star\right)}{(1-\tau)^{3/2}} = \sum_{k=1}^\infty \frac{\delta^k}{k!}u_k.$$

As shown in Li et al. (2025, Equation (74)), and recall $u_\tau^\star := x_\tau^\star/\sqrt{1-\tau}$ one has

$$\frac{s^\star_{\tau+\delta}(x^\star_{\tau+\delta})}{(1-\tau-\delta)^{3/2}} = -\frac{1}{(\tau+\delta)(1-\tau-\delta)}\left(\frac{x^\star_{\tau+\delta}}{\sqrt{1-\tau-\delta}} - \frac{x^\star_\tau}{\sqrt{1-\tau}} + \frac{\int p_{X_0|\bar{X}_\tau}(x_0|x^\star_\tau)\, e^{\Delta}\,(u^\star_\tau - x_0)\, dx_0}{\int p_{X_0|\bar{X}_\tau}(x_0|x^\star_\tau)\, e^{\Delta}\, dx_0}\right), \tag{86}$$

where

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

We next prove that there exist universal constants $M_1 > 1$ and $C_{\mathrm{s}}, \bar{C} > 0$ such that

$$\|u_k\|_2 \leq C_{\mathrm{s}}(M_1K)^k k! \sqrt{\frac{d\theta\log T}{\tau(1-\tau)^3}}\left(\frac{d\theta\log T}{\tau(1-\tau)}\right)^k, \qquad 0\leq k\leq K, \tag{87}$$

and, for every $C \geq 2$,

$$|v_k| \leq \frac{\bar{C}C^2}{M_1K}(M_1K)^k k!\left(\frac{d\theta\log T}{\tau(1-\tau)}\right)^k, \qquad 1\leq k\leq K, \tag{88}$$

provided that

$$\|u^\star_\tau - x_0\|_2 \leq 5C\sqrt{\frac{d\theta\tau\log T}{1-\tau}}.$$

Indeed, for $k = 0$, the results hold directly from Lemma 10 (see Li et al. (2025, Equation (75a)) for more details). Suppose that (87) and (88) hold for all $k \leq k_0$. Then, by the same arguments as in Li et al. (2025), we have

$$\begin{aligned}
&|v_{k_0+1}| \\
\leq\ & (k_0+1)!\Bigg[\tau^{-k_0}\|u^\star_\tau - x_0\|_2^2 + \frac{1}{(k_0+1)!}\frac{1-\tau}{\tau}\|u^\star_\tau - x_0\|_2\|u_{k_0}\|_2 + \|u^\star_\tau - x_0\|_2\sum_{\ell=1}^{k_0}\frac{\tau^{-\ell-1}}{(k_0+1-\ell)!}\|u_{k_0-\ell}\|_2 \\
&+\frac{1-\tau}{4\tau}\sum_{\ell=1}^{k_0}\frac{1}{\ell!(k_0+1-\ell)!}\|u_{\ell-1}\|_2\|u_{k_0-\ell}\|_2 + \sum_{\ell=1}^{k_0-1}\sum_{j=1}^{k_0-\ell}\frac{1}{j!(k_0+1-\ell-j)!}\tau^{-\ell-1}\|u_{j-1}\|_2\|u_{k_0-\ell-j}\|_2\Bigg].
\end{aligned}$$

Using the induction hypothesis (87) and (88), we obtain

$$\begin{aligned}
|v_{k_0+1}| &\leq (k_0+1)!\left[25C^2 + \frac{5C_{\mathrm{s}}C}{k_0+1}(M_1K)^{k_0} + 5C_{\mathrm{s}}C(M_1K)^{k_0} + C_{\mathrm{s}}^2(M_1K)^{k_0-1} + 2C_{\mathrm{s}}^2(M_1K)^{k_0}\right]\left(\frac{d\theta\log T}{\tau(1-\tau)}\right)^{k_0+1} \\
&\leq \frac{\bar{C}C^2}{M_1K}(M_1K)^{k_0+1}(k_0+1)!\left(\frac{d\theta\log T}{\tau(1-\tau)}\right)^{k_0+1},
\end{aligned}$$

provided $\bar{C} > 5(1 + C_s + C_s^2)$ and then $M_1 \geq 100\bar{C}c_7^2$. Therefore, we have verified (88) for $k = k_0+1$. Next, we prove that (87) holds for $k = k_0+1$. Let $e^{\Delta} =: \sum_{k=0}^\infty \frac{\delta^k}{k!}w_k$ denote the Taylor expansion of e^Δ. Then one can show that $w_0 = 1$ and for all $1 \leq k \leq k_0+1$,

$$\begin{aligned}
|w_k| = \left|\sum_{\ell=1}^{k}\frac{1}{\ell!}\sum_{j_1+\cdots+j_\ell=k}\frac{k!}{j_1!\cdots j_\ell!}v_{j_1}\cdots v_{j_\ell}\right| &\leq (M_1K)^k k!\left(\frac{d\theta\log T}{\tau(1-\tau)}\right)^k\sum_{\ell=1}^{k}\frac{1}{\ell!}\binom{k-1}{\ell-1}\left(\frac{\bar{C}C^2}{M_1K}\right)^\ell \\
&\leq \frac{\bar{C}C^2}{M_1K}\left(1+\frac{\bar{C}C^2}{M_1K}\right)^{k-1}(M_1K)^k k!\left(\frac{d\theta\log T}{\tau(1-\tau)}\right)^k \leq \frac{C}{K}(MK)^k k!\left(\frac{d\theta\log T}{\tau(1-\tau)}\right)^k.
\end{aligned} \tag{89}$$

Denote

$$\int_{x_0} p_{X_0|\bar{X}_\tau}(x_0 \mid x^\star_\tau)\exp(\Delta)\,(u^\star_\tau - x_0)\,\mathrm{d}x_0 := \sum_{k=0}^\infty\frac{\delta^k}{k!}a_k, \quad \int_{x_0} p_{X_0|\bar{X}_\tau}(x_0 \mid x^\star_\tau)\exp(\Delta)\mathrm{d}x_0 := \sum_{k=0}^\infty\frac{\delta^k}{k!}b_k.$$

Then for $0 \le k \le k_0+1$, similar to Li et al. (2025), one has

$$\begin{aligned}
a_0 &= \int_{x_0} p_{X_0|\bar{X}_\tau}(x_0 \mid x_\tau^\star)(u_\tau^\star - x_0) w_0 \, \mathrm{d}x_0 = \int_{x_0} p_{X_0|\bar{X}_\tau}(x_0 \mid x_\tau^\star)(u_\tau^\star - x_0) \, \mathrm{d}x_0, \\
\|a_k\|_2 &\le C'(M_1K)^k k! \sqrt{\frac{d\theta\tau\log T}{1-\tau}} \left(\frac{d\theta \log T}{\tau(1-\tau)}\right)^k, \qquad 0 \le k \le k_0+1, \\
b_0 &= 1, \qquad |b_k| \le \frac{C''}{K}(M_1K)^k k! \left(\frac{d\theta\log T}{\tau(1-\tau)}\right)^k, \qquad 1 \le k \le k_0+1.
\end{aligned}$$

Here $C' = O(1+c_7)$ is an absolute constant, and $0 < C'' \le 1/4$. Next define d_k through the Taylor expansion

$$\frac{\sum_{m=0}^\infty \frac{\delta^m}{m!} a_m}{\sum_{\ell=0}^\infty \frac{\delta^\ell}{\ell!} b_\ell} = \frac{\int_{x_0} p_{X_0|\bar{X}_\tau}(x_0 \mid x_\tau^\star)\exp(\Delta)(u_\tau^\star - x_0)\,\mathrm{d}x_0}{\int_{x_0} p_{X_0|\bar{X}_\tau}(x_0 \mid x_\tau^\star)\exp(\Delta)\mathrm{d}x_0} := \sum_{k=0}^\infty \frac{\delta^k}{k!} d_k.$$

Then

$$d_0 = \int_{x_0} p_{X_0|\bar{X}_\tau}(x_0 \mid x_\tau^\star)\left(\frac{x_\tau^\star}{\sqrt{1-\tau}} - x_0\right)\mathrm{d}x_0, \quad d_k = a_k - \sum_{\ell=0}^{k-1}\binom{k}{\ell} d_\ell b_{k-\ell}, \qquad 1 \le k \le k_0+1.$$

Choosing $C''' \ge 2C'$, we obtain

$$\begin{aligned}
\|d_k\|_2 &\le \left(C' + \frac{k}{K}C''C'''\right)(M_1K)^k k! \sqrt{\frac{d\theta\tau\log T}{1-\tau}}\left(\frac{d\theta\log T}{\tau(1-\tau)}\right)^k \\
&\le C'''(M_1K)^k k! \sqrt{\frac{d\theta\tau\log T}{1-\tau}}\left(\frac{d\theta\log T}{\tau(1-\tau)}\right)^k, \quad 0 \le k \le k_0+1. \qquad (90)
\end{aligned}$$

Similarly to (80) in Li et al. (2025), through (86) we have

$$\sum_{k=1}^\infty \frac{\delta^k}{k!} u_k = \sum_{k=1}^\infty \left(\frac{1}{\tau(1-\tau)}\frac{u_{k-1} - 2d_k}{2k!} + \sum_{\ell=1}^k \left[(-\tau)^{-\ell-1} - (1-\tau)^{-\ell-1}\right]\frac{2d_{k-\ell} - u_{k-1-\ell}}{2(k-\ell)!}\right)\delta^k.$$

By hypothesis (87) and (90) hold for $0 \le k < k_0$, we now prove (87) for $k = k_0+1$

$$\begin{aligned}
\|u_{k_0+1}\|_2 &= \left\| \frac{1}{\tau(1-\tau)}\left(\frac{u_{k_0}}{2} - d_{k_0+1}\right) + \sum_{\ell=1}^{k_0+1}\left[(-\tau)^{-\ell-1} - (1-\tau)^{-\ell-1}\right]\frac{(k_0+1)!}{2\,(k_0+1-\ell)!}\left(2d_{k_0+1-\ell} - u_{k_0-\ell}\right)\right\|_2 \\
&\le C'''(M_1K)^{k_0+1}(k_0+1)!\sqrt{\frac{d\theta\log T}{\tau(1-\tau)^3}}\left(\frac{d\theta\log T}{\tau(1-\tau)}\right)^{k_0+1} + \frac{C_\mathrm{s}(M_1K)^{k_0}(k_0+1)!}{2}\sqrt{\frac{d\theta\log T}{\tau(1-\tau)^3}}\left(\frac{d\theta\log T}{\tau(1-\tau)}\right)^{k_0+1} \\
&\quad + \frac{(C''' + C_\mathrm{s})K}{M_1K}(M_1K)^{k_0+1}(k_0+1)!\sqrt{\frac{d\theta\log T}{\tau(1-\tau)^3}}\left(\frac{d\theta\log T}{\tau(1-\tau)}\right)^{k_0} \\
&\le C_\mathrm{s}(M_1K)^{k_0+1}(k_0+1)!\sqrt{\frac{d\theta\log T}{\tau(1-\tau)^3}}\left(\frac{d\theta\log T}{\tau(1-\tau)}\right)^{k_0+1}. \qquad (91)
\end{aligned}$$

where the last inequality follows by choosing $C_\mathrm{s} \ge 2 + C'''$ and $M_1 \ge 100C_\mathrm{s}^2 c_7^2$. Thus we proved (87) holds for any $0 \le k \le K$.

We next prove (57). By Tweedie's formula (Efron, 2011), we have

$$\frac{1}{(1-\tau)^{3/2}}\frac{\partial s_\tau^\star(x_\tau^\star)}{\partial u_\tau^\star} = -\frac{1}{\tau(1-\tau)} I_d + \frac{1}{\tau^2}\operatorname{Cov}\left(\frac{x_\tau^\star}{\sqrt{1-\tau}} - X_0 \,\middle|\, \bar{X}_\tau = x_\tau^\star\right).$$

Denote $m_k := \frac{\partial^k}{\partial \tau^k}\left[\frac{1}{(1-\tau)^{3/2}}\frac{\partial s_\tau^\star(x_\tau^\star)}{\partial u_\tau^\star}\right]$. For sufficiently small δ, we write

$$\sum_{k=1}^{\infty}\frac{\delta^k}{k!}m_k = \frac{1}{(1-\tau-\delta)^{3/2}}\frac{\partial s_{\tau+\delta}^\star(x_{\tau+\delta}^\star)}{\partial u_{\tau+\delta}^\star} - \frac{1}{(1-\tau)^{3/2}}\frac{\partial s_\tau^\star(x_\tau^\star)}{\partial u_\tau^\star},$$

where

$$\frac{1}{(1-\tau-\delta)^{3/2}}\frac{\partial s_{\tau+\delta}^\star(x_{\tau+\delta}^\star)}{\partial u_{\tau+\delta}^\star} = -\frac{I_d}{(\tau+\delta)(1-\tau-\delta)} + \frac{1}{(\tau+\delta)^2}\left[\frac{\int p_{X_0|\bar{X}_\tau}(x_0 \mid x_\tau^\star)e^{\Delta}\left(u_\tau^\star - x_0\right)\left(u_\tau^\star - x_0\right)^\top \mathrm{d}x_0}{\int p_{X_0|\bar{X}_\tau}(x_0 \mid x_\tau^\star)e^{\Delta}\,\mathrm{d}x_0}\right.$$
$$\left. - \left(\frac{\int p_{X_0|\bar{X}_\tau}(x_0 \mid x_\tau^\star)e^{\Delta}\left(u_\tau^\star - x_0\right)\mathrm{d}x_0}{\int p_{X_0|\bar{X}_\tau}(x_0 \mid x_\tau^\star)e^{\Delta}\,\mathrm{d}x_0}\right)\left(\frac{\int p_{X_0|\bar{X}_\tau}(x_0 \mid x_\tau^\star)e^{\Delta}\left(u_\tau^\star - x_0\right)\mathrm{d}x_0}{\int p_{X_0|\bar{X}_\tau}(x_0 \mid x_\tau^\star)e^{\Delta}\,\mathrm{d}x_0}\right)^\top\right],$$

and $\Delta = \sum_{k\geq 1}\frac{\delta^k}{k!}v_k$. Applying the same argument as above, gives

$$\|m_k\| = \left\|\frac{\partial^k}{\partial \tau^k}\left[\frac{1}{(1-\tau)^{3/2}}\frac{\partial s_\tau^\star(x_\tau^\star)}{\partial u_\tau^\star}\right]\right\| \leq \widetilde{C}(M_2K)^k k!\left(\frac{d\theta\log T}{\tau(1-\tau)}\right)^{k+1}, \quad M_2\geq 2, \quad 0\leq k\leq K. \tag{92}$$

On the other hand, recall that $J_\tau^\star := \frac{\partial\left(x_\tau^\star/\sqrt{1-\tau}\right)}{\partial\left(x_{\tau_{t,0}}^\star/\sqrt{1-\tau_{t,0}}\right)}$ and (69), then the exact Jacobian satisfies

$$\frac{d}{d\tau}J_\tau^\star = -\frac{1}{2}\left[\frac{1}{(1-\tau)^{3/2}}\frac{\partial s_\tau^\star(x_\tau^\star)}{\partial u_\tau^\star}\right]J_\tau^\star, \qquad J_{\tau_{t,0}}^\star = I.$$

Because (76) gives $\|J_\tau^\star\|\leq c_8$. Then we choose $\widehat{C}\geq c_8$ and assume that $\left\|\frac{\partial^k}{\partial\tau^k}J_\tau^\star\right\|\leq \widehat{C}k!\,(M_2K)^k\left(\frac{d\theta\log T}{\tau(1-\tau)}\right)^k$ for $0\leq k\leq k_0$. Differentiating the preceding identity $k-1$ times yields

$$\frac{\partial^k}{\partial\tau^k}J_\tau^\star = -\frac{1}{2}\sum_{\ell=0}^{k-1}\binom{k-1}{\ell}m_\ell\frac{\partial^{k-1-\ell}}{\partial\tau^{k-1-\ell}}J_\tau^\star. \tag{93}$$

Using (92) and assumption above, we prove (93) for $k = k_0+1$

$$\begin{aligned}
\left\|\frac{\partial^{k_0+1}}{\partial\tau^{k_0+1}}J_\tau^\star\right\| &\leq \frac{1}{2}\sum_{\ell=0}^{k_0}\binom{k_0}{\ell}\|m_\ell\|\,\left\|\frac{\partial^{k_0-\ell}}{\partial\tau^{k_0-\ell}}J_\tau^\star\right\| \\
&\leq \frac{1}{2}\widetilde{C}\widehat{C}\sum_{\ell=0}^{k_0}\binom{k_0}{\ell}(M_2K)^\ell\ell!\left(\frac{d\theta\log T}{\tau(1-\tau)}\right)^{\ell+1}(M_2K)^{k_0-\ell}(k_0-\ell)!\left(\frac{d\theta\log T}{\tau(1-\tau)}\right)^{k_0-\ell} \\
&\leq \frac{\widetilde{C}}{2M_2K}\widehat{C}(k_0+1)!(M_2K)^{k_0+1}\left(\frac{d\theta\log T}{\tau(1-\tau)}\right)^{k_0+1} \\
&\leq \widehat{C}(k_0+1)!(M_2K)^{k_0+1}\left(\frac{d\theta\log T}{\tau(1-\tau)}\right)^{k_0+1}.
\end{aligned}$$

Finally, applying Leibniz' rule to $\left[\frac{1}{(1-\tau)^{3/2}}\frac{\partial s_\tau^\star(x_\tau^\star)}{\partial u_\tau^\star}\right]J_\tau^\star$ and combining with (92) and the bound of $\left\|\frac{\partial^k}{\partial\tau^k}J_\tau^\star\right\|$ above, for any $0\leq k\leq K$ we get

$$\left\|\frac{\partial^k}{\partial\tau^k}\left[\frac{1}{(1-\tau)^{3/2}}\frac{\partial s_\tau^\star(x_\tau^\star)}{\partial u_\tau^\star}J_\tau^\star\right]\right\| \leq \sum_{\ell=0}^{k}\binom{k}{\ell}\|m_\ell\|\,\|\frac{\partial^{k-\ell}}{\partial\tau^{k-\ell}}J_\tau^\star\|d \leq C_{\mathrm{J}}(M_2K)^{k+1}k!\left(\frac{d\theta\log T}{\tau(1-\tau)}\right)^{k+1},$$

where the last inequality follows from$C_{\mathrm{J}}\geq 2\widetilde{C}\widehat{C}$, $k+1\leq K+1\leq 2K$ and $M_2\geq 2$. This proves (57). ☐

C.5 Proof of Lemma 14

Proof. For convenience, let $\theta_t := \theta_{\tau_{t,0}}(x_{\tau_{t,0}})$. Recalling that $x^\star_\tau$ is the solution of ODE (8) at τ with the initial condition $x^\star_{\tau_{t,0}} = x_{\tau_{t,0}}$, we know from Lemma 12 that

$$-\log p_{\bar{X}_\tau}(x^\star_\tau) \le 2\theta_t d \log T, \qquad \tau_{t,K-1} \le \tau \le \tau_{t,0}.$$

The same argument as in Li et al. (2025, Eq. (91)) gives, for all $0 \le i \le K-1$ and all $\lambda \in [0,1]$,

$$-\log p_{\bar{X}_{\tau_{t,i}}}\left(\lambda x^{(0)}_{\tau_{t,i}} + (1-\lambda)x^\star_{\tau_{t,i}}\right) \le 2.1 d\theta_t \log T. \tag{94}$$

Combining (94) with Lemma 11, we obtain

$$\left\| \frac{\partial s^\star_{\tau_{t,i}}\left(\lambda x^{(0)}_{\tau_{t,i}} + (1-\lambda)x^\star_{\tau_{t,i}}\right)}{\partial x} \right\| \lesssim \frac{d\theta_t \log T}{\tau_{t,i}}, \qquad 0 \le i \le K-1, \quad \lambda \in [0,1]. \tag{95}$$

For any $0 \le n \le N-1$ and $0 \le i \le K-1$, we define

$$u^{(n+1)}_{t,i} := \sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,i}}}\, x^{(n+1)}_{\tau_{t,i}} - x_{\tau_{t,0}} = \sqrt{1-\tau_{t,0}} \left[\sum_{j=0}^{i-1} A^{(t)}_{ji} s_{\tau_{t,j}}\left(x^{(n+1)}_{\tau_{t,j}}\right) + \sum_{j=i}^{K-1} A^{(t)}_{ji} s_{\tau_{t,j}}\left(x^{(n)}_{\tau_{t,j}}\right) \right], \tag{96}$$

where $A^{(t)}_{ji} = \frac{1}{2}\gamma_{t,j}\left(\tau_{t,i}\right)\left(1-\tau_{t,j}\right)^{-3/2}$. Similarly, let

$$u^\star_{t,i} := \sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,i}}}\, x^\star_{\tau_{t,i}} - x_{\tau_{t,0}} = -\sqrt{1-\tau_{t,0}} \int_{\tau_{t,0}}^{\tau_{t,i}} \frac{1}{2(1-\tau)^{3/2}} s^\star_\tau(x^\star_\tau)\, d\tau, \tag{97}$$

where the second inequality follows from (10). Next, we shall prove the following two estimates simultaneously by induction: for each $0 \le n \le N-1$ and each $0 \le i \le K-1$, it holds

$$\left\| u^{(n+1)}_{t,i} - u^\star_{t,i} \right\|_2 \le C_{10} \frac{\tau_{t,0}\log T}{T} \sqrt{\log K \sum_{m,j} \left(\varepsilon^{(m)}_{\text{score},t,j}\left(x^{(m)}_{\tau_{t,j}}\right)\right)^2} + C_{10}\sqrt{\frac{d\theta_t \tau_{t,0} \log^3 T}{T^2}} \left(\frac{d\theta_t \log^2 TK}{T}\right) \tag{98}$$

and for all $\lambda \in [0,1]$,

$$-\log p_{\bar{X}_{\tau_{t,i}}}\left(\lambda x^{(n+1)}_{\tau_{t,i}} + (1-\lambda)x^\star_{\tau_{t,i}}\right) \le 2.1 d\theta_t \log T. \tag{99}$$

Here, $\varepsilon^{(m)}_{\text{score},t,j}(\cdot)$ is defined in (29) and $C_{10} > 0$ is a sufficiently large constant. Moreover, with iteration (15) we have

$$x^{(n)}_{\tau_{t,0}} = x^\star_{\tau_{t,0}} = x_{\tau_{t,0}}, \qquad \text{for all} \quad n \ge 0$$

Thus for $n = 0, i = 0$, we have

$$u^{(1)}_{t,0} = u^\star_{t,0} = 0, \quad \left\| u^{(1)}_{t,0} - u^\star_{t,0} \right\|_2 = 0, \tag{100}$$

$$-\log p_{\bar{X}_{\tau_{t,0}}}\left(\lambda x^{(1)}_{\tau_{t,0}} + (1-\lambda)x^\star_{\tau_{t,0}}\right) = -\log p_{\bar{X}_{\tau_{t,0}}}\left(x^\star_{\tau_{t,0}}\right) \le 2.1 d\theta_t \log T, \tag{101}$$

where (101) holds due to (94). This implies (98) and (99) holds for $n = 0, i = 0$. Similar to (95), we also have

$$\left\| \frac{\partial s^\star_{\tau_{t,0}}\left(\lambda x^{(1)}_{\tau_{t,0}} + (1-\lambda)x^\star_{\tau_{t,0}}\right)}{\partial x} \right\| \lesssim \frac{d\theta_t \log T}{\tau_{t,0}}, \quad \lambda \in [0,1]. \tag{102}$$

Assume that for all $0 \le i \le k_0$ and for all $\lambda \in [0,1]$, (98) and (99) hold, where $0 \le k_0 \le K-2$.

Prove (98) **for** $n = 0$ **and** $i = k_0 + 1$**.** By (96), (97), and the definition of $A_{ji}^{(t)} = \frac{1}{2}\gamma_{t,j}\left(\tau_{t,i}\right)\left(1-\tau_{t,j}\right)^{-3/2}$, we have

$$u_{t,k_0+1}^{(1)} - u_{t,k_0+1}^{\star} = J_{1,k_0+1}^{(1)} + J_{2,k_0+1}^{(1)} + J_{3,k_0+1}^{(1)},$$

where

$$\begin{aligned} J_{1,k_0+1}^{(1)} &= \sqrt{1-\tau_{t,0}}\Bigg[\sum_{j=0}^{k_0} A_{j,k_0+1}^{(t)}\left(s_{\tau_{t,j}}\left(x_{\tau_{t,j}}^{(1)}\right) - s_{\tau_{t,j}}^{\star}\left(x_{\tau_{t,j}}^{(1)}\right)\right) \\ &\qquad + \sum_{j=k_0+1}^{K-1} A_{j,k_0+1}^{(t)}\left(s_{\tau_{t,j}}\left(x_{\tau_{t,j}}^{(0)}\right) - s_{\tau_{t,j}}^{\star}\left(x_{\tau_{t,j}}^{(0)}\right)\right)\Bigg], \\ J_{2,k_0+1}^{(1)} &= \sqrt{1-\tau_{t,0}}\Bigg[\sum_{j=0}^{k_0} A_{j,k_0+1}^{(t)}\left(s_{\tau_{t,j}}^{\star}\left(x_{\tau_{t,j}}^{(1)}\right) - s_{\tau_{t,j}}^{\star}\left(x_{\tau_{t,j}}^{\star}\right)\right) \\ &\qquad + \sum_{j=k_0+1}^{K-1} A_{j,k_0+1}^{(t)}\left(s_{\tau_{t,j}}^{\star}\left(x_{\tau_{t,j}}^{(0)}\right) - s_{\tau_{t,j}}^{\star}\left(x_{\tau_{t,j}}^{\star}\right)\right)\Bigg], \\ J_{3,k_0+1}^{(1)} &= \sqrt{1-\tau_{t,0}}\left[\sum_{j=0}^{K-1} A_{j,k_0+1}^{(t)} s_{\tau_{t,j}}^{\star}\left(x_{\tau_{t,j}}^{\star}\right) - \int_{\tau_{t,k_0+1}}^{\tau_{t,0}} \frac{1}{2(1-\tau)^{3/2}} s_{\tau}^{\star}(x_{\tau}^{\star})\, d\tau\right]. \end{aligned}$$

For the first term $J_{1,k_0+1}^{(1)}$, the Cauchy–Schwarz yields

$$\begin{aligned} \|J_{1,k_0+1}^{(1)}\|_2 &\lesssim \sqrt{1-\tau_{t,0}}\left(\sum_{j=0}^{K-1}|A_{j,k_0+1}^{(t)}|^2\right)^{1/2}\left(\sum_{m=0}^{1}\sum_{j=0}^{K-1}\left(\varepsilon_{\text{score},t,j}^{(m)}\left(x_{\tau_{t,j}}^{(m)}\right)\right)^2\right)^{1/2} \\ &\lesssim \frac{\tau_{t,0}\log T}{T}\sqrt{\log K\sum_{m=0}^{N}\sum_{j=0}^{K-1}\left(\varepsilon_{\text{score},t,j}^{(m)}\left(x_{\tau_{t,j}}^{(m)}\right)\right)^2}, \end{aligned} \tag{103}$$

where the last line follows from (46k) and (46f) by setting $i_1 = 0$, $i_2 = k_0+1$, $i_3 = 0$ and $i_4 = 0$. For the second term $J_{2,k_0+1}^{(1)}$ term, since (99) holds for all $0 \le i \le k_0$, similar to (95), we have

$$\left\|\frac{\partial s_{\tau_{t,i}}^{\star}\left(\lambda x_{\tau_{t,i}}^{(1)} + (1-\lambda)x_{\tau_{t,i}}^{\star}\right)}{\partial x}\right\| \lesssim \frac{d\theta_t \log T}{\tau_{t,i}}, \quad \lambda \in [0,1] \tag{104}$$

for all $0 \le i \le k_0$. Using (104) and (95) we obtain

$$\begin{aligned} \|J_{2,k_0+1}^{(1)}\|_2 &\lesssim \sqrt{1-\tau_{t,0}}\sum_{j=0}^{k_0}|A_{j,k_0+1}^{(t)}|\frac{d\theta_t\log T}{\tau_{t,j}}\left\|x_{\tau_{t,j}}^{(1)} - x_{\tau_{t,j}}^{\star}\right\|_2 \\ &\quad + \sqrt{1-\tau_{t,0}}\sum_{j=k_0+1}^{K-1}|A_{j,k_0+1}^{(t)}|\frac{d\theta_t\log T}{\tau_{t,j}}\left\|x_{\tau_{t,j}}^{(0)} - x_{\tau_{t,j}}^{\star}\right\|_2 \\ &\lesssim \frac{d\theta_t \log^2 T\,\log K}{T}\left[\max_{0\le j\le k_0}\left\|u_{t,j}^{(1)} - u_{t,j}^{\star}\right\|_2 + \max_{k_0+1\le j\le K-1}\left\|x_{\tau_{t,j}}^{(0)} - x_{\tau_{t,j}}^{\star}\right\|_2\right], \end{aligned} \tag{105}$$

where the last line follows from $x_{\tau_{t,j}}^{(m)} - x_{\tau_{t,j}}^{\star} = \sqrt{\frac{1-\tau_{t,j}}{1-\tau_{t,0}}}\left(u_{t,j}^{(m)} - u_{t,j}^{\star}\right)$, $1 \le m \le N$ and (46j) with the fact $\tau_{t,K-1} \le \tau_{t,j}$,$0 \le j \le K-1$, (46f) by setting $i_1 = 0$, $i_2 = k_0+1$, $i_3 = K-1$ and $i_4 = 0$. The first term in (105) is bounded by assumption (98) holds for all $0 \le i \le k_0$. For the second term in (105), the same estimate as in Li et al. (2025, Eq. (90)) gives

$$\max_{k_0+1\le j\le K-1}\left\|x_{\tau_{t,j}}^{(0)} - x_{\tau_{t,j}}^{\star}\right\|_2 \lesssim \frac{\tau_{t,0}-\tau_{t,K-1}}{\sqrt{\tau_{t,0}}}\sqrt{d\theta_t\log T}.$$

Then we obtain

$$\begin{aligned}\|J_{2,k_0+1}^{(1)}\|_2 &\lesssim \frac{d\theta_t \log^2 T\, \log K}{T}\left(C_{10}\frac{\tau_{t,0}\log T}{T}\sqrt{\log K\sum_{m=0}^{N}\sum_{j=0}^{K-1}\left(\varepsilon_{\mathrm{score},t,j}^{(m)}\left(x_{\tau_{t,j}}^{(m)}\right)\right)^2}\right.\\ &\qquad\left.+C_{10}\sqrt{\frac{d\theta_t\tau_{t,0}\log^3 T}{T^2}}\left(\frac{d\theta_t\log^2 TK}{T}\right)+\frac{\tau_{t,0}-\tau_{t,K-1}}{\sqrt{\tau_{t,0}}}\sqrt{d\theta_t\log T}\right)\\ &\lesssim \frac{\tau_{t,0}\log T}{T}\sqrt{\log K\sum_{m=0}^{N}\sum_{j=0}^{K-1}\left(\varepsilon_{\mathrm{score},t,j}^{(m)}\left(x_{\tau_{t,j}}^{(m)}\right)\right)^2}+\sqrt{\frac{d\theta_t\tau_{t,0}\log^3 T}{T^2}}\left(\frac{d\theta_t\log^2 TK}{T}\right). \end{aligned} \tag{106}$$

Here the last inequality follows from $T \geq C_2 d\log^4 T$ and $\frac{\tau_{t,0}-\tau_{t,K-1}}{\sqrt{\tau_{t,0}}}\sqrt{d\theta_t\log T} \leq \frac{(\tau_{t,0}-\tau_{t,K-1})}{(1-\tau_{t,K-1})\tau_{t,0}}\sqrt{d\tau_{t,0}\theta_t\log T}$ with (46f) in Lemma 9. Similar to Bound for G_2 (63) in Section B.1, $J_{3,k_0+1}^{(1)}$ is easily bounded:

$$\|J_{3,k_0+1}^{(1)}\|_2 \lesssim \sqrt{\frac{d\theta_t\tau_{t,0}\log^3 T}{T^2}}\left(\frac{M_0 K d\theta_t\log^2 T}{T}\right)^K. \tag{107}$$

Combining (103), (106), and (107), for sufficiently large T and large enough C_{10} we get

$$\left\|u_{t,k_0+1}^{(1)}-u_{t,k_0+1}^{\star}\right\|_2 \leq C_{10}\frac{\tau_{t,0}\log T}{T}\sqrt{\log K\sum_{m=0}^{N}\sum_{j=0}^{K-1}\left(\varepsilon_{\mathrm{score},t,j}^{(m)}\left(x_{\tau_{t,j}}^{(m)}\right)\right)^2}+C_{10}\sqrt{\frac{d\theta_t\tau_{t,0}\log^3 T}{T^2}}\left(\frac{d\theta_t\log^2 TK}{T}\right). \tag{108}$$

By induction, we prove (98) for any $0 \leq i \leq K-1$ at $n=0$.

Prove (99) for $n = 0$ and $i = k_0 + 1$. We next derive the bound of $-\log p_{\bar{X}_{\tau_{t,k_0+1}}}\left(\lambda x_{\tau_{t,k_0+1}}^{(1)}+(1-\lambda)x_{\tau_{t,k_0+1}}^{\star}\right)$ from (108). We first claim that for $i = k_0+1$ at $n=0$, it holds by large enough C_{10}

$$\begin{aligned}&\left|\log\frac{p_{\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,k_0+1}}}\bar{X}_{\tau_{t,k_0+1}}}\left(\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,k_0+1}}}(\lambda x_{\tau_{t,k_0+1}}^{\star}+(1-\lambda)x_{\tau_{t,k_0+1}}^{(1)})\right)}{p_{\bar{X}_{\tau_{t,0}}}(x_{\tau_{t,0}})}\right|\\ &\qquad\leq C_{10}\left\{\frac{d\theta_t\log^2 T}{T}+\frac{\sqrt{d\theta_t\log^3 T\,\log K\sum_{m,j}\left(\varepsilon_{\mathrm{score},t,j}^{(m)}\left(x_{\tau_{t,j}}^{(m)}\right)\right)^2}}{T}\right\},\end{aligned} \tag{109}$$

By the affine change of variables,

$$\begin{aligned}&p_{\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,k_0+1}}}\bar{X}_{\tau_{t,k_0+1}}}\left(\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,k_0+1}}}(\lambda x_{\tau_{t,k_0+1}}^{\star}+(1-\lambda)x_{\tau_{t,k_0+1}}^{(1)})\right)\\ &\qquad=\left(\frac{1-\tau_{t,0}}{1-\tau_{t,k_0+1}}\right)^{-d/2}p_{\bar{X}_{\tau_{t,k_0+1}}}(\lambda x_{\tau_{t,k_0+1}}^{\star}+(1-\lambda)x_{\tau_{t,k_0+1}}^{(1)}).\end{aligned}$$

Therefore,

$$\begin{aligned}
&-\log p_{\bar{X}_{\tau_{t,k_0+1}}}(\lambda x^{\star}_{\tau_{t,k_0+1}}+(1-\lambda)x^{(1)}_{\tau_{t,k_0+1}}) \\
&=-\log p_{\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,k_0+1}}}\bar{X}_{\tau_{t,k_0+1}}}\left(\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,k_0+1}}}(\lambda x^{\star}_{\tau_{t,k_0+1}}+(1-\lambda)x^{(1)}_{\tau_{t,k_0+1}})\right)-\frac{d}{2}\log\left(\frac{1-\tau_{t,0}}{1-\tau_{t,k_0+1}}\right) \\
&\leq -\log p_{\bar{X}_{\tau_{t,0}}}(x_{\tau_{t,0}})+C_{10}\left\{\frac{d\theta_t\log^2 T}{T}+\frac{\sqrt{d\theta_t\log^3 T\,\log K\sum_{m,j}\left(\varepsilon^{(m)}_{\mathrm{score},t,j}\left(x^{(m)}_{\tau_{t,j}}\right)\right)^2}}{T}\right\}+c_1\frac{d\log T}{T} \\
&\leq 2.1 d\theta_t\log T,
\end{aligned} \tag{110}$$

where the third line follows from $-\frac{d}{2}\log\left(\frac{1-\tau_{t,0}}{1-\tau_{t,k_0+1}}\right)\leq -\frac{d}{2}\log\left(\frac{1-\tau_{t,0}}{1-\tau_{t,K-1}}\right)=-\frac{d}{2}\log\left(\frac{\bar{\alpha}_t}{\bar{\alpha}_{t-1}}\right)=-\frac{d}{2}\log(\alpha_t)=\frac{d}{2}\int_0^{\frac{c_1\log T}{T}}\frac{1}{1-u}du\leq\frac{d}{2}\int_0^{\frac{c_1\log T}{T}}2du\leq\frac{d}{2}\cdot\frac{2c_1\log T}{T}=\frac{c_1 d\log T}{T}$ using (46a) in Lemma 9 and the last line holds by $-\log p_{\bar{X}_{\tau_{t,0}}}\left(x_{\tau_{t,0}}\right)\leq d\theta_t\log T$ with (47) and condition (58) in Lemma 14 for sufficiently large T. This proves that (99) holds for any $0\leq i\leq K-1$ at $n=0$.

Repeat the same arguments above at $n=1,\cdots,N-1$, we can prove (98) and (99) hold for all $0\leq n\leq N-1$ and $0\leq i\leq K-1$. Moreover, we have

$$\log\frac{p_{\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,i}}}\bar{X}_{\tau_{t,i}}}\left(\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,i}}}x^{(n+1)}_{\tau_{t,i}}\right)}{p_{\bar{X}_{\tau_{t,0}}}(x_{\tau_{t,0}})}\leq\frac{4c_1 d\log T}{T}+C_{10}\left\{\frac{d^2\theta_t^2\log^4 TK}{T^2}+\frac{\sqrt{d\theta_t\log^3 T\,\log K\sum_{m,j}\left(\varepsilon^{(m)}_{\mathrm{score},t,j}\left(x^{(m)}_{\tau_{t,j}}\right)\right)^2}}{T}\right\}, \tag{111}$$

holds for all $0\leq n\leq N-1$ and $0\leq i\leq K-1$. This gives the conclusion. It remains to prove claim (109).

Proof of claim (109) As in Li et al. (2025, Eq. (95)–(96)), one has

$$\|u^{\star}_{t,k_0+1}\|_2\leq C\sqrt{\frac{d\theta_t\tau_{t,0}\log^3 T}{T^2}}. \tag{112}$$

Therefore, by (108) and (112),

$$\|u^{(1)}_{t,k_0+1}\|_2\leq C_{10}\left\{\sqrt{\frac{d\theta_t\tau_{t,0}\log^3 T}{T^2}}+\frac{\tau_{t,0}\log T}{T}\sqrt{\log K\sum_{m=0}^{N}\sum_{j=0}^{K-1}\left(\varepsilon^{(m)}_{\mathrm{score},t,j}\left(x^{(m)}_{\tau_{t,j}}\right)\right)^2}\right\}. \tag{113}$$

As in Li et al. (2025, Eq. (93)), we have

$$\begin{aligned}
&\frac{p_{\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,k_0+1}}}\bar{X}_{\tau_{t,k_0+1}}}\left(\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,k_0+1}}}x^{(1)}_{\tau_{t,k_0+1}}\right)}{p_{\bar{X}_{\tau_{t,0}}}(x_{\tau_{t,0}})} \\
&=\left(1+\frac{d}{2}\frac{\tau_{t,0}-\tau_{t,k_0+1}}{(1-\tau_{t,0})\tau_{t,k_0+1}}+O\left(d^2\left(\frac{\tau_{t,0}-\tau_{t,k_0+1}}{(1-\tau_{t,0})\tau_{t,k_0+1}}\right)^2\right)\right)\int_{x_0}p_{X_0|\bar{X}_{\tau_{t,0}}}(x_0\mid x_{\tau_{t,0}}) \\
&\quad\cdot\exp\left(-\frac{(\tau_{t,0}-\tau_{t,k_0+1})\|x_{\tau_{t,0}}-\sqrt{1-\tau_{t,0}}x_0\|_2^2}{2(1-\tau_{t,0})\tau_{t,0}\tau_{t,k_0+1}}-\frac{\|u^{(1)}_{t,k_0+1}\|_2^2+2\left\langle u^{(1)}_{t,k_0+1},x_{\tau_{t,0}}-\sqrt{1-\tau_{t,0}}x_0\right\rangle}{2\frac{(1-\tau_{t,0})\tau_{t,k_0+1}}{1-\tau_{t,k_0+1}}}\right)dx_0.
\end{aligned} \tag{114}$$

Denote for $\ell = 1, 2, \ldots,$

$$E_\ell^{\text{typical}} := \left\{ x_0 : \left\| x_{\tau_{t,0}} - \sqrt{1-\tau_{t,0}} x_0 \right\|_2 \leq 5\ell \sqrt{d\theta_t \tau_{t,0} \log T} \right\}.$$

Then for any $x_0 \in E_\ell^{\text{typical}}$, similar to Li et al. (2025, Eq. (98)), by (46f) and (46h),

$$\frac{(\tau_{t,0} - \tau_{t,k_0+1}) \left\| x_{\tau_{t,0}} - \sqrt{1-\tau_{t,0}} x_0 \right\|_2^2}{2(1-\tau_{t,0})\tau_{t,0}\tau_{t,k_0+1}} \lesssim \ell^2 \frac{d\theta_t \log^2 T}{T}. \tag{115}$$

Also, using (113), (46h), and $\left\| x_{\tau_{t,0}} - \sqrt{1-\tau_{t,0}} x_0 \right\|_2 \leq 5\ell \sqrt{d\theta_t \tau_{t,0} \log T}$, similar to Li et al. (2025, Eq. (99)) we have

$$\left| \frac{\|u_{t,k_0+1}^{(1)}\|_2^2 + 2\left\langle u_{t,k_0+1}^{(1)}, x_{\tau_{t,0}} - \sqrt{1-\tau_{t,0}} x_0 \right\rangle}{2\frac{(1-\tau_{t,0})\tau_{t,k_0+1}}{1-\tau_{t,k_0+1}}} \right|$$
$$\lesssim \ell \frac{d\theta_t \log^2 T}{T} + \ell \frac{\sqrt{d\theta_t \log^3 T \, \log K \sum_{m=0}^N \sum_{j=0}^{K-1} \left(\varepsilon_{\text{score},t,j}^{(m)} \left(x_{\tau_{t,j}}^{(m)} \right) \right)^2}}{T}. \tag{116}$$

Combining (108), (114), (115), and (116), and then repeating the same shell decomposition and Jensen argument as in Li et al. (2025, Eq. (100)–(106)), together with Lemma 13 and (46f), yields

$$\log \frac{p_{\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,k_0+1}}} \bar{X}_{\tau_{t,k_0+1}}} \left(\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,k_0+1}}} x_{\tau_{t,k_0+1}}^{(1)} \right)}{p_{\bar{X}_{\tau_{t,0}}}(x_{\tau_{t,0}})}$$
$$\leq \frac{4c_1 d \log T}{T} + C_{10} \left\{ \frac{d^2\theta_t^2 \log^4 T \, K}{T^2} + \frac{\sqrt{d\theta_t \log^3 T \, \log K \sum_{m,j} \left(\varepsilon_{\text{score},t,j}^{(m)} \left(x_{\tau_{t,j}}^{(m)} \right) \right)^2}}{T} \right\}. \tag{117}$$

Next, fix any $\lambda \in [0, 1]$ and define

$$\widetilde{x}_{\tau_{t,k_0+1}}^{(1)}(\lambda) := \lambda x_{\tau_{t,k_0+1}}^\star + (1-\lambda) x_{\tau_{t,k_0+1}}^{(1)},$$

and

$$\widetilde{u}_{t,k_0+1}^{(1)}(\lambda) := \sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,k_0+1}}} \widetilde{x}_{\tau_{t,k_0+1}}^{(1)}(\lambda) - x_{\tau_{t,0}}.$$

Then

$$\widetilde{u}_{t,k_0+1}^{(1)}(\lambda) = \lambda u_{t,k_0+1}^\star + (1-\lambda) u_{t,k_0+1}^{(1)}.$$

Consequently,

$$\left\| \widetilde{u}_{t,k_0+1}^{(1)}(\lambda) \right\|_2 \leq \left\| u_{t,k_0+1}^\star \right\|_2 + \left\| u_{t,k_0+1}^{(1)} - u_{t,k_0+1}^\star \right\|_2.$$

It follows that the same bounds (115) and (116) remain valid with $u_{t,k_0+1}^{(1)}$ replaced by $\widetilde{u}_{t,k_0+1}^{(1)}(\lambda)$. Repeating the argument used to derive (117), we obtain

$$\left| \log \frac{p_{\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,k_0+1}}} \bar{X}_{\tau_{t,k_0+1}}} \left(\sqrt{\frac{1-\tau_{t,0}}{1-\tau_{t,k_0+1}}} \left(\lambda x_{\tau_{t,k_0+1}}^\star + (1-\lambda) x_{\tau_{t,k_0+1}}^{(1)} \right) \right)}{p_{\bar{X}_{\tau_{t,0}}}(x_{\tau_{t,0}})} \right|$$
$$\leq C_{10} \left\{ \frac{d\theta_t \log^2 T}{T} + \frac{\sqrt{d\theta_t \log^3 T \, \log K \sum_{m,j} \left(\varepsilon_{\text{score},t,j}^{(m)} \left(x_{\tau_{t,j}}^{(m)} \right) \right)^2}}{T} \right\}. \tag{118}$$

This proves the claim (109). $\square$