Enhanced convergence guarantees of score-based generative models in $\mathcal{W}_2$-distance beyond log-concavity

Xiaoyan Zhang
Central South University
x.y.zhang@csu.edu.cn

Chenxu Pang
Central South University
c.x.pang@csu.edu.cn

Xiaojie Wang*
Central South University
x.j.wang7@csu.edu.cn

Abstract

Although score-based generative models (SGMs) have achieved remarkable success in real-world sample generation tasks, it is still far from sufficient to understand them from a mathematical perspective. In this paper, we aim to provide enhanced convergence guarantees for two existing SGMs in $\mathcal{W}_2$-distance beyond log-concavity. More precisely, we develop a novel framework of discretization error analysis for Euler-Maruyama (EM) and Poisson midpoint (PM) time discretization schemes for SGMs. Under a non-log-concavity condition, we show that $\tilde{\mathcal{O}}(\sqrt{d}/\epsilon)$ iterations suffice to approximate the target distributions in ϵ-accuracy for the classical EM-based SGM, considerably improving upon the existing iteration complexity $\tilde{\mathcal{O}}(d/\epsilon^2)$. Notably, the novel framework of error analysis also enables us to establish an $\tilde{\mathcal{O}}(\sqrt{d}/\epsilon^{2/3})$ iteration complexity for the PM-based SGM in ϵ-accuracy, significantly outperforming current state-of-the-art Wasserstein convergence guarantees for SDE-based diffusion samplers.

1 Introduction

Diffusion models such as denoising diffusion probabilistic models (DDPMs) [20] have achieved state-of-the-art performance in data generation across a wide range of applications, including image generation [15, 20, 43], natural language processing [38, 33], inverse problems [11, 6] and many other areas. In many cases, they are shown to outperform other generative models such as generative adversarial networks (GANs) [19], variational autoencoders (VAEs) [24], and normalizing flows [39].

Through stochastic differential equations (SDEs), Song et al. [45] originally formulated a general theoretical framework for diffusion models, referred to as score-based generative models (SGMs), consisting of two processes: the forward process and the reverse process. Via the density flow $\{p_t\}_{t\geq 0}$, the forward process gradually (as $t \to \infty$) transforms the target distribution $p_0 = \mu$ into a tractable prior distribution p_∞, typically the standard Gaussian law. This step is commonly referred to as the diffusion process. Along these diffusion trajectories, the gradient of the log-density function (i.e., $\nabla \log p_t$), known as the score function and generally not available in a closed form, is estimated using score matching techniques, such as denoising score matching [49] or sliced score matching [44]. By leveraging the learned score function $s_{\theta^\star}$, one can rely on a time-discretization scheme (e.g., the classical Euler-Maruyama, EM) for the reverse process to convert random noise back into new samples that approximately follow the target distribution μ. Despite the remarkable empirical success of SGMs across various applications, their theoretical understanding is still far from enough and the present paper attempts to provide enhanced convergence guarantees of SGMs in $\mathcal{W}_2$-distance beyond log-concavity.

1.1 Related works on convergence guarantees of SGMs

*Corresponding author

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

In recent years, a growing body of research has focused on the theoretical understanding of SDE-based SGMs, see, e.g., [3, 8, 10, 28, 26, 22, 30, 29, 34, 27, 32, 53]. In early works [3, 34], convergence guarantees often suffer from exponential dependencies on parameters such as dimension and smoothness constant. This was overcome by [10], where a $\tilde{\mathcal{O}}(d/\epsilon^2)$ iteration complexity was obtained in total variation (TV) distance under the spatial Lipschitz condition on the score functions $\nabla \log p_t(x)$ (with respect to x). Further, the authors of [8] established $\tilde{\mathcal{O}}(d/\epsilon^2)$ and $\tilde{\mathcal{O}}(d^2/\epsilon^2)$ iteration complexity bounds in KL divergence under the spatial Lipschitz condition on $\nabla \log p_t(x)$ and $\nabla \log p_0(x)$, respectively. These error bounds scale polynomially and even have a logarithmic dependence on relevant problem parameters, significantly improving prior works. More recently in [32], the authors showed $\tilde{\mathcal{O}}(d/\epsilon)$ iterations for the SDE-based sampler suffice to achieve a given precision level ϵ in TV distance, under the assumption that the target distribution has finite moments.

During the preparation of the present manuscript, we were aware of two interesting papers [22, 53] on convergence guarantees of SDE-based samplers. More accurately, the authors of [22] obtained an improved iteration complexity of order $\tilde{\mathcal{O}}(\sqrt{d}/\epsilon)$ in TV distance and KL divergence under the assumption that the score function is spatially Lipschitz. Also, the authors of [53] proposed a denoising diffusion randomized midpoint (DDRaM) method for the time discretization of the reverse process and derived a $\tilde{\mathcal{O}}(\sqrt{d}/\epsilon)$ iteration complexity in KL divergence, under the condition that the estimated score function is spatial Lipschitz (not uniform-in-time). Besides, we would like to mention a series of recent works [30, 29, 31, 9, 21], providing theoretical error bounds in TV distance for ODE-based diffusion samplers.

Convergence in Wasserstein distance with and without log-concavity. Note that the above-mentioned convergence guarantees are all obtained in TV distance and KL divergence. In contrast, convergence guarantees in Wasserstein distance are less studied in the literature. To analyze the error in $\mathcal{W}_2$-distance, previous works often imposed the log-concavity condition on the target distribution (see, e.g., [5, 17, 18, 47, 52] and references therein), which can be also utilized to ensure the above frequently used Lipschitz condition on the score function.

Nevertheless, the log-concavity condition is too restrictive and precludes complex and multi-modal real-world target distributions. To remedy it, some researchers try to analyze the error in $\mathcal{W}_2$- distance under certain non-log-concavity conditions such as the weak log-concavity [42, 25] and the log-Sobolev inequality (LSI) condition [27]. The authors of [4] established convergence guarantees under semi log-concavity and strong log-concavity at infinity allowing for potentially discontinuous gradients, obtaining iteration complexities of $\tilde{\mathcal{O}}(d/\epsilon^2)$ in the general setting and $\tilde{\mathcal{O}}(d/\epsilon)$ under additional regularity assumptions on the estimated score function in $\mathcal{W}_2$-distance. In particular, for a class of weakly log-concave target distributions, the authors of [42] derived iteration complexity $\tilde{\mathcal{O}}(d/\epsilon^2)$ for the EM scheme to achieve a given precision level ϵ in $\mathcal{W}_2$-distance, in the setting that the target distribution satisfies the weak log-concavity condition. For a general class of probability flow ODEs, the authors of [25] established an $\tilde{\mathcal{O}}(\sqrt{d}/\epsilon)$ iteration complexity in $\mathcal{W}_2$-distance under the weak log-concavity condition on the target distribution.

Bearing the above error bounds in mind, the following interesting question thus arises:

Under the weak log-concavity condition, can the classical EM scheme for the SDE-based SGMs achieve an $\tilde{\mathcal{O}}(\sqrt{d}/\epsilon)$ iteration complexity in $\mathcal{W}_2$-distance?

This paper provides an affirmative answer to this question. As illustrated by Theorem 3.5, an $\tilde{\mathcal{O}}(\sqrt{d}/\epsilon)$ iteration complexity in $\mathcal{W}_2$-distance can be derived for the EM method under the weak log-concavity condition. This matches the best known iteration complexity bound in any metric for the EM discretization of the reverse SDE.

It is worth mentioning that, the authors of [2] derived the optimal $\tilde{\mathcal{O}}(\sqrt{d}/\epsilon)$ iteration complexity for the DDPM (EM) sampler in $\mathcal{W}_2$-distance under a uniform posterior-covariance condition for Gaussian perturbations. More recently in [50], the authors established the $\tilde{\mathcal{O}}(\sqrt{d}/\epsilon)$ iteration complexity in $\mathcal{W}_2$-distance under a Gaussian-tail assumption on the target distribution. Compared to these two works, we rely on different assumptions and proof strategies in the error analysis. Notably, the assumptions and proof techniques used in this paper work for other sampling algorithms, such as the Poisson midpoint method proposed in [23, 46].

It is known that, the EM method still suffers from low sampling speed, requiring a large number of score function evaluations (often neural network evaluations) to generate new samples. A natural accelerated alternative is Runge-Kutta type method [51, 21], but the price is to impose stronger smoothness assumption on the (estimated) score function. Indeed, accelerated convergence guarantees of Runge-Kutta type methods for SGMs in $\mathcal{W}_2$-distance beyond log-concavity are, to the best of our knowledge, still missing in the literature.

To speed up the sampling, in this work we do not rely on Runge-Kutta type methods, but turn to the Poisson midpoint (PM) method (9) originally proposed by [23], where only convergence guarantees in KL divergence of PM-based Langevin Monte Carlo (LMC) were established. Very recently in [46], the authors further obtained a $\tilde{\mathcal{O}}(d^{2/3}/\epsilon^{2/3})$ iteration complexity in $\mathcal{W}_2$-distance for the PM-based overdamped LMC under the strongly log-concave condition and the Lipschitz condition of the drift. Although numerical results in [23] showed that the PM-based SGMs substantially outperform the vanilla DDPM and ODE-based methods, the theoretical convergence guarantee of the PM-based SGM is unfortunately still absent in the literature. Therefore, we ask the following question:

Can an enhanced iteration complexity be obtained for PM-based SGM beyond log-concavity, to validate its outperformance over the other SDE-based samplers?

In this work, we answer this question in the affirmative and establish an $\tilde{\mathcal{O}}(\sqrt{d}/\epsilon^{2/3})$ iteration complexity in $\mathcal{W}_2$-distance (see Theorem 3.9) under the weak log-concavity condition. This thus fills the theoretical gap left in [23].

1.2 Contributions of this work

To conclude, our contribution can be summarized as follows:

- **Refined discretization error analysis.** A refined approach of discretization error analysis is developed to characterize the error propagation of the scheme. This helps us provide enhanced convergence guarantees in $\mathcal{W}_2$-distance for SDE-based SGMs with two existing discretization schemes beyond log-concavity, including the Euler-Maruyama method and Poisson midpoint method.
- **Enhanced iteration complexity for Euler–Maruyama discretization in $\mathcal{W}_2$-distance.** We obtain a $\tilde{\mathcal{O}}(\sqrt{d}/\epsilon)$ iteration complexity of the EM discretization (8) of the reverse SDE that suffices to achieve $\mathcal{W}_2$ error ϵ (see Theorem 3.5), significantly improving the previous $\tilde{\mathcal{O}}(d/\epsilon^2)$ EM bound of [42] under weak log-concavity.
- **Sharper iteration complexity via Poisson midpoint discretization in $\mathcal{W}_2$-distance.** For the Poisson midpoint method (9), we establish an iteration complexity of $\tilde{\mathcal{O}}\big(\sqrt{d}/\epsilon^{2/3}\big)$ for SGMs in $\mathcal{W}_2$-distance (see Theorem 3.9), under weak log-concavity and an additional Lipschitz-in-time condition on the score (Assumption 3.7). To the best of our knowledge, this bound fills the theoretical gap left in [23] and improves upon existing state-of-the-art Wasserstein convergence guarantees for SDE-based diffusion samplers.

In Table 1, we compare our theoretical findings with the existing results, in terms of the total iterations required to achieve a given precision level $\epsilon > 0$. Clearly, our results improve upon existing state-of-the-art Wasserstein convergence guarantees for SDE-based diffusion samplers.

The rest of this article is organized as follows. The next section collects some notations, introduces SGMs and two types of methods throughout this paper. In Section 3, we present assumptions and main results of this paper, whose proofs are given in Appendix. A technical overview of the proofs is given in Section 4. Finally, some concluding remarks are given in the last section.

2 Preliminaries on SGMs

2.1 Notation

Throughout this paper, we use $\mathbb{N}$ to denote the set of all positive integers and let $d \in \mathbb{N}$. For any positive integer m, we denote $[m] := \{1, ..., m\}$ and $[m]_0 := \{0, 1, ..., m\}$. We use $I_d \in \mathbb{R}^{d\times d}$ to denote the identity matrix. Let $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$ denote the Euclidean norm and the inner product of vectors in $\mathbb{R}^d$, respectively. For any matrix $A = (a_{i,j}) \in \mathbb{R}^{d\times d}$, the F-norm is defined as

Table 1: A comparison of iteration complexity for SGMs.

	Algorithm	Metric	Iteration complexity	Log -concavity	Additional assumption[1]
[10]	EM	TV	$\tilde{\mathcal{O}}(d/\epsilon^2)$	No	$\nabla \log p_t(x)$ Lipschitz in x
[8]	EM	KL [2]	$\tilde{\mathcal{O}}(d/\epsilon^2)$	No	$\nabla \log p_t(x)$ Lipschitz in x
	EM	KL [2]	$\tilde{\mathcal{O}}(d^2/\epsilon^2)$	No	$\nabla \log p_0(x)$ Lipschitz in x
[17]	EM	$\mathcal{W}_2$	$\tilde{\mathcal{O}}(d/\epsilon^2)$	Yes	$\nabla \log p_t(x)$ Lipschitz in t
[52]	EM	$\mathcal{W}_2$	$\tilde{\mathcal{O}}(d/\epsilon^2)$	Yes	$\nabla \log p_t(x)$ Lipschitz in t
[53]	DDRaM [3]	KL	$\tilde{\mathcal{O}}(\sqrt{d}/\epsilon)$	No	$\tilde{s}_\theta(t,x)$ Lipschitz in x
[25]	Euler (ODE)	$\mathcal{W}_2$	$\tilde{\mathcal{O}}(\sqrt{d}/\epsilon)$	No	Weak log-concavity $\nabla \log p_t(x)$ Lipschitz in t
[42]	EM	$\mathcal{W}_2$	$\tilde{\mathcal{O}}(d/\epsilon^2)$	No	Weak log-concavity
This work	EM	$\mathcal{W}_2$	$\tilde{\mathcal{O}}(\sqrt{d}/\epsilon)$	No	Weak log-concavity
This work	PM	$\mathcal{W}_2$	$\tilde{\mathcal{O}}(\sqrt{d}/\epsilon^{2/3})$	No	Weak log-concavity $\nabla \log p_t(x)$ Lipschitz in t

[1] Smoothness assumptions on the (approximate) score function.
[2] For error bounds in KL divergence, the given precision level is set to be ϵ^2.
[3] Denoising diffusion randomized midpoint method.

$\|A\|_{\mathrm{F}} := \sqrt{\sum_{i,j=1}^d |a_{i,j}|^2}$ and the operator norm as $\|A\| = \|A\|_{\mathrm{op}} := \sup_{\|v\|=1} \|Av\|$. It is not difficult to see $\|A\| \le \|A\|_{\mathrm{F}} \le \sqrt{d}\,\|A\|$. We denote the density of the standard Gaussian distribution on $\mathbb{R}^d$ by γ^d. For two numbers a and b, we denote $a \vee b := \max\{a, b\}$ and $a \wedge b := \min\{a, b\}$. We write $\tilde{\mathcal{O}}(\cdot)$ to mean that it hides polylogarithmic factors in d and $1/\epsilon$.

Let $\mathcal{B}(\mathbb{R}^d)$ be the Borel σ-field on $\mathbb{R}^d$, and let $\mathcal{P}(\mathbb{R}^d)$ denote the set of probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. We write $\mathcal{P}_2(\mathbb{R}^d)$ for the subset of probability measures with finite second moments. By $\mathcal{L}(X)$ we denote the probability distribution of the $\mathbb{R}^d$-valued random variable X. Given $\nu_1, \nu_2 \in \mathcal{P}_2(\mathbb{R}^d)$, let $\mathcal{D}(\nu_1, \nu_2)$ be the set of all couplings of ν_1 and ν_2, i.e., probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ whose first and second marginals are ν_1 and ν_2, respectively. The $\mathcal{W}_2$-distance between ν_1 and ν_2 is then defined by

$$\mathcal{W}_2(\nu_1, \nu_2) := \inf_{\nu \in \mathcal{D}(\nu_1, \nu_2)} \left(\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \|\theta_1 - \theta_2\|^2 \, \mathrm{d}\,\nu\,(\theta_1, \theta_2) \right)^{1/2}. \tag{1}$$

Let $\mathcal{C}\left(\mathbb{R}^d, \mathbb{R}\right)$ denote the set of all continuous functions from $\mathbb{R}^d$ to $\mathbb{R}$. For $k \ge 0$, denote by $\mathcal{C}^k\left(\mathbb{R}^d, \mathbb{R}\right)$ the set of functions from $\mathbb{R}^d$ to $\mathbb{R}$ which have continuous 0-th, ..., k-th order derivatives.

2.2 Score-based generative models

In this subsection, we provide a brief summary of the ideas behind the construction of score-based generative models (SGMs) based on diffusion processes [45]. An SGM consists of a forward process and a reverse process. The forward process progressively transforms the target distribution into a simple reference distribution. In this paper, we use the Ornstein–Uhlenbeck (OU) diffusion as the forward process. Specifically, let $\{\overrightarrow{X}_t\}_{0 \le t \le T}$ solve

$$\mathrm{d}\overrightarrow{X}_t = -\overrightarrow{X}_t \,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B(t), \quad \overrightarrow{X}_0 \sim \mu, \quad 0 \le t \le T, \tag{2}$$

where $\mu(\mathrm{d}x) \propto \exp(-V(x))\,\mathrm{d}x$ is the target distribution and $\{B(t)\}_{0 \le t \le T}$ is an $\mathbb{R}^d$-valued standard Brownian motion. Then the OU process (2) admits the explicit representation as below

$$\overrightarrow{X}_t \overset{\mathrm{d}}{=} \mathrm{e}^{-t}\overrightarrow{X}_0 + \sqrt{1 - \mathrm{e}^{-2t}}\, Z, \quad Z \sim \gamma^d, \tag{3}$$

where Z is independent of $\overrightarrow{X}_0$. Thus, $\overrightarrow{X}_t$ converges exponentially to the standard Gaussian law γ^d.

SGMs generate new samples by simulating the time reversal of (2), which transports the terminal law $\mathcal{L}(\overrightarrow{X}_T)$ back to μ. Let $\{\overleftarrow{X}_t\}_{0\le t\le T}$ denote the corresponding exact reverse process. By the time-reversal formula for diffusions [1, 7], it satisfies

$$\mathrm{d}\overleftarrow{X}_t = \big(\overleftarrow{X}_t + 2\nabla \log p_{T-t}(\overleftarrow{X}_t)\big)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W(t), \quad \overleftarrow{X}_0 \sim \mathcal{L}(\overrightarrow{X}_T), \qquad 0 \le t \le T, \tag{4}$$

where p_t denotes the density of $\mathcal{L}(\overrightarrow{X}_t)$ with respect to Lebesgue measure, and $\{W(t)\}_{0\le t\le T}$ is another $\mathbb{R}^d$-valued standard Brownian motion defined on a filtered probability space $(\Omega_W, \mathcal{F}^W, \{\mathcal{F}_t^W\}_{t\ge 0}, \mathbb{P}_W)$, satisfying the usual conditions. Moreover, such a Brownian motion is assumed to be independent of $\{B(t)\}_{0\le t\le T}$. The drift $\nabla \log p_t$ of the reverse-time SDE is the gradient of the log-density of $\mathcal{L}(\overrightarrow{X}_t)$, also known as the *score function*. Since $\nabla \log \gamma^d(x) = -x$, the backward SDE (4) can be rewritten in the following equivalent form:

$$\mathrm{d}\overleftarrow{X}_t = \big(-\overleftarrow{X}_t + 2\nabla \log \tilde{p}_{T-t}(\overleftarrow{X}_t)\big)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W(t), \quad \overleftarrow{X}_0 \sim \mathcal{L}(\overrightarrow{X}_T), \qquad 0 \le t \le T, \tag{5}$$

where, for $(t, x) \in [0, T] \times \mathbb{R}^d$,

$$\tilde{p}_t(x) := p_t(x)/\gamma^d(x).$$

This time-reversal technique is extensively used in the literature (see, e.g., [7, 13, 42, 48, 47]), owing to its preservation of the forward Ornstein–Uhlenbeck (2) drift $-x$ in the reverse dynamics. However, since the score function $\nabla \log p_t(x)$, which appears in (4), is not available. To obtain a practical sampler, we replace it by a learned approximation $s_\theta : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ from a parametric family $\{s_\theta\}_{\theta\in\Theta}$, with parameter $\theta^\star$ obtained by minimizing the denoising score-matching loss [49]:

$$\theta \to \int_0^T \mathbb{E}\Big[\big\|s_\theta(t, \overrightarrow{X}_t) - \nabla \log p_t(\overrightarrow{X}_t)\big\|^2\Big]\mathrm{d}t. \tag{6}$$

Since $\nabla \log \tilde{p}_t(x) = x + \nabla \log p_t(x)$ for any $x \in \mathbb{R}^d$ and $t \in [0, T]$, therefore, the corresponding estimator of the relative score is then defined by

$$\tilde{s}_{\theta^\star}(t, x) := x + s_{\theta^\star}(t, x).$$

Since $\mathcal{L}(\overrightarrow{X}_T)$ depends on the unknown target μ and is not directly accessible, we initialize the sampler from the stationary law γ^d as shown in (3). Replacing in addition the relative score in (5) by $\tilde{s}_{\theta^\star}$ yields the following practical sampler

$$\mathrm{d}Y_t = \big(-Y_t + 2\tilde{s}_{\theta^\star}(T - t, Y_t)\big)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W(t), \quad Y_0 \sim \gamma^d, \qquad 0 \le t \le T. \tag{7}$$

2.3 Time discretization schemes

For a general learned score, the reverse SDE (7) cannot be simulated exactly. Sampling generation therefore requires a time-discretization scheme for (7). Let π_h be a temporal grid of the form

$$\pi_h := \big\{t_n + \tfrac{ih}{k} = n \cdot h + \tfrac{ih}{k}, \quad n \in [N-1]_0,\, k \in [N], \quad i \in [k]_0 \text{ with } h = T/N\big\}.$$

We consider two types of discretization methods, which are widely used in existing work.

- Euler-Maruyama (EM) method, for any $n \in [N-1]_0$, $N \in \mathbb{N}$:

$$Y_{t_{n+1}} = Y_{t_n} + h\big(-Y_{t_n} + 2\tilde{s}_{\theta^\star}(T - t_n, Y_{t_n})\big) + \sqrt{2}\Delta W_{t_n}, \quad Y_0 \sim \gamma^d, \tag{8}$$

  where $\Delta W_{t_n} := W(t_{n+1}) - W(t_n)$.

- Poisson midpoint (PM) method [46, 23], for any $n \in [N-1]_0$, $i \in [k-1]_0$, $N, k \in \mathbb{N}$:

$$\begin{aligned}
\tilde{Y}_{t_{n,i}} &= Y_{t_n} + \tfrac{ih}{k}\big(-Y_{t_n} + 2\tilde{s}_{\theta^\star}(T - t_n, Y_{t_n})\big) + \sum_{j=0}^{i-1}\sqrt{\tfrac{2h}{k}}Z_{t_{n,j}},\\
Y_{t_{n,i+1}} &= Y_{t_{n,i}} + \tfrac{h}{k}\big(-Y_{t_n} + 2\tilde{s}_{\theta^\star}(T - t_n, Y_{t_n})\big) + hB_{t_{n,i}}\Big[\big(-\tilde{Y}_{t_{n,i}}\\
&\qquad + 2\tilde{s}_{\theta^\star}(T - (t_n + \tfrac{ih}{k}), \tilde{Y}_{t_{n,i}})\big) - \big(-Y_{t_n} + 2\tilde{s}_{\theta^\star}(T - t_n, Y_{t_n})\big)\Big] + \sqrt{\tfrac{2h}{k}}Z_{t_{n,i}},\\
Y_{t_{n+1}} &= Y_{t_{n+1},0} = Y_{t_{n,k}}, \quad Y_0 \sim \gamma^d,
\end{aligned} \tag{9}$$

  where $Z_{t_{n,i}}$ denote independent standard Gaussians and $B_{t_{n,i}}$ denote independent Bernoulli random variables with parameter $1/k$.

For simplicity, we denote $t_{n,i} := t_n + \frac{ih}{k}$ for any $n \in [N-1]_0$, $i \in [k]_0$ in the sequel. The Poisson midpoint method was first introduced in [23] as a discrete variant of the randomized midpoint method proposed by [41]. Noting that $\mathbb{E}[\sum_{j=0}^{k-1} B_{t_{n,j}}] = 1$, we observe that an expected number of 2 gradient score function evaluations per batch is required, which matches the randomized midpoint method in [41].

3 Assumptions and main results

We begin with the following definition of weak convexity, quoted from [12, 42].

Definition 3.1 (Weak convexity)**.** *The weak convexity profile of a function $g \in \mathcal{C}^1(\mathbb{R}^d, \mathbb{R})$ is defined as*

$$\kappa_g(r) := \inf_{x,y \in \mathbb{R}^d : \|x-y\|=r} \left\{ \frac{\langle \nabla g(x) - \nabla g(y), x - y \rangle}{\|x-y\|^2} \right\}, \quad r > 0. \tag{10}$$

We say that g is (α, M)-weakly convex if

$$\kappa_g(r) \geq \alpha - \frac{1}{r} f_M(r) \quad \forall\, r > 0 \tag{11}$$

for some constants $\alpha, M > 0$ and

$$f_M(r) := 2\sqrt{M} \tanh\left(\tfrac{1}{2} \sqrt{M} r \right). \tag{12}$$

Moreover, we say that g is (α, M)-weakly concave if $-g$ is (α, M)-weakly convex.

Next we make the following weak convexity assumption of the potential function V.

Assumption 3.2 (Weak convexity)**.** *The target distribution μ in the form $\mu(\mathrm{d}x) \propto \exp(-V(x))\mathrm{d}x$ is absolutely continuous with respect to Lebesgue measure, where $V : \mathbb{R}^d \to \mathbb{R}$ is (α, M)-weakly convex for some constants $\alpha, M > 0$.*

Assumption 3.2 relaxes the strong log-concavity condition ($\alpha - M \geq 0$), commonly imposed in prior works [17, 52, 5, 47]. Indeed, some Gaussian mixture models satisfy this weak convexity condition, instead of the strong log-concavity condition (see [42, Appendix A]). A standard sufficient condition to imply Assumption 3.2 is the contractivity at infinity condition (see [12, Assumption 1.1, H2'] or [35, Assumption 1.1]); see [12, Proposition 5.1] for details.

Assumption 3.3 (One-sided Lipschitz)**.** *The function $\nabla V : \mathbb{R}^d \to \mathbb{R}^d$ is one-sided Lipschitz continuous. Namely, there exists a constant $L_V \geq 0$, independent of d, such that, for any $x, y \in \mathbb{R}^d$,*

$$\langle \nabla V(x) - \nabla V(y), x - y \rangle \leq L_V \|x - y\|^2. \tag{13}$$

All the results rely on the following L^2-accuracy of the score estimator. This assumption quantifies the ability of the neural network architecture to approximate the true score function and serves to control the score approximation error.

Assumption 3.4 (Score estimation error)**.** *There exist $\epsilon_0 \geq 0$ and $\theta^\star \in \Theta$ such that for any $n \in [N-1]_0$, $i \in [k-1]_0$,*

$$\max_{U \in \{Y_{t_{n,i}}, \tilde{Y}_{t_{n,i}}\}} \left\| \nabla \log \tilde{p}_{T-t_{n,i}}(U) - \tilde{s}_{\theta^\star}(T - t_{n,i}, U) \right\|_{L^2(\Omega, \mathbb{R}^d)} \leq \epsilon_0, \tag{14}$$

where $Y_{t_{n,i}}$ is the numerical approximations produced by the EM method (8) *or the PM method* (9)*, and $\tilde{Y}_{t_{n,i}}$ is the predictor of the PM method* (9)*.*

Indeed, Assumption 3.4 can be replaced by the standard score-matching assumption along the forward noising distribution commonly used in the literature (e.g.,[10, 8]) and the additional assumptions that $\tilde{s}_{\theta^\star}(T - t_{n,i}, \cdot)$, $\nabla \log \tilde{p}_{T-t_{n,i}}(\cdot)$ are Lipschitz in space. With this substitution, one can still arrive at the same error bounds after very minor modifications to the proof of the main theorem. In this situation, the required spatial Lipschitz continuity of $\nabla \log \tilde{p}_{T-t_{n,i}}(\cdot)$ can be guaranteed by Lemma A.3 and it only remains to assume the spatial Lipschitz of the learned score $\tilde{s}_{\theta^\star}(T - t_{n,i}, \cdot)$, as also used in [9]. Here we adopt the score error along the trajectory of the approximate process in Assumption 3.4 because it leads to the most direct and tractable treatment of the score approximation error within our proof framework. Such assumptions were also employed in existing literature, e.g., [17, 25, 52].

3.1 Main result for the Euler-Maruyama method

In the following theorem, we quantify the $\mathcal{W}_2$-distance between the distribution of EM method (8) and the target distribution μ.

Theorem 3.5. *[Main result: error bounds for EM] Suppose that Assumptions 3.2, 3.3 and 3.4 hold. Let $\{Y_{t_n}\}_{n\in[N]_0}$ be produced by the EM method* (8) *with the uniform step-size $h = T/N$. For any $\lambda \in (0,1)$, assume that*

$$0 < h < \tfrac{2-\lambda}{2(2L+1)^2},$$

where $L > 0$ is the Lipschitz constant of $x \mapsto \nabla \log \tilde{p}_t(x)$ from Lemma A.3. Then there exists a constant $C > 0$, independent of d and h, such that

$$\mathcal{W}_2\left(\mu, \mathcal{L}\left(Y_{t_N}\right)\right) \le C\mathrm{e}^{C\eta(\alpha,M,(2L+1)^2h+\lambda/2)}\Big(\underbrace{\sqrt{Td}h}_{\text{Discretization error}} + \underbrace{\mathrm{e}^{-T}\mathcal{W}_2(\mu,\gamma^d)}_{\text{Initialization error}} + \underbrace{\epsilon_0 T}_{\text{Score-matching error}} \Big), \tag{15}$$

where ϵ_0 comes from Assumption 3.4 and $\eta\big(\alpha, M, (2L+1)^2h + \lambda/2\big)$ is defined in (30).

A brief proof overview is provided in the next section, while the complete proof is deferred to Appendix. As a direct consequence, we obtain the following scaling of the iteration number, terminal time T, and score approximation error required to achieve a given $\mathcal{W}_2$ accuracy. The proof is straightforward and omitted here.

Corollary 3.6. *Let Assumptions 3.2, 3.3 and 3.4 hold. To achieve a given precision level $\epsilon > 0$ in $\mathcal{W}_2$-distance, the required number of iterations of the EM method* (8) *is of order $\tilde{\mathcal{O}}\big(\frac{\sqrt{d}}{\epsilon}\big)$ for the discretization error to be small, the time scale T is of order $\mathcal{O}\big(\log(\frac{\sqrt{d}}{\epsilon})\big)$ for the initialization error to be small, and the score-matching error ϵ_0 is of order $\mathcal{O}\Big(\frac{\epsilon}{\log(\epsilon^{-1}\sqrt{d})}\Big)$ for the propagated score-matching error to be small.*

In the case that the score error in Assumption 3.4 polynomially depends on T, e.g., $\mathrm{Poly}(T)\cdot\epsilon_0$, the third error term in (15) would have higher-order polynomial dependence on T, which affects only the logarithmic factors in the iteration complexity.

As validated in [42, Proposition 4.1], in the case of Gaussian mixture distributions with equal variances,

$$p_0(x) := \sum_{i=1}^{m} \beta_i \frac{1}{(2\pi\sigma^2)^{d/2}} \exp\left(-\frac{\|x-\mu_i\|^2}{2\sigma^2}\right), \qquad x \in \mathbb{R}^d, \tag{16}$$

where $\sigma > 0$, $\mu_i \in \mathbb{R}^d$, and $\beta_i \in [0,1]$ for $i \in [m]$, such that $\sum_{i=1}^m \beta_i = 1$, $-\log p_0$ is weakly convex and $\nabla \log p_0$ is Lipschitz. Moreover, the corresponding weak convexity parameters $\alpha_{p_0} = 1/\sigma^2$, $\sqrt{M_{p_0}} := 2m\sum_{i=1}^m \|\mu_i\|/\sigma^2$, and the Lipschitz constant $L_{p_0} := \alpha_{p_0} + M_{p_0}$ are independent of d, whenever $\max_{1\le i\le m}\|\mu_i\|$ is bounded independently of d (see [42, Appendix A] for details). Hence, for Gaussian mixture distributions with equal variances, the stated complexity bound $\tilde{\mathcal{O}}(\sqrt{d}/\epsilon)$ has no hidden dependence on d. In a more general setting, where the modes of the Gaussian mixture have different covariance matrices, the required weak convexity and Lipschitz condition might be violated, with the associated constants possibly depending on d. As a consequence, our analysis does not guarantee that all hidden constants in the stated complexity bound are independent of the dimension d for general Gaussian mixture distributions with unequal covariance matrices.

Finally, we mention that, in the non-strongly-log-concave setting (i.e., $\alpha - M \le 0$), the constant $\eta(\alpha, M, \rho)$ in (30) is well-defined, as $\alpha > 0$ and ρ can stay uniformly bounded away from 1 for sufficiently small step-size. Improving the constant $\eta(\alpha, M, \rho)$ is left for future work.

3.2 Main result for the Poisson midpoint method

Unlike the error analysis for EM, the error analysis of PM requires an additional Lipschitz-in-time regularity condition on the score function for the subsequent discretization error analysis. Such a regularity condition has also been previously used in the existing literature [25, 52, 17].

Assumption 3.7 (Lipschitz in time)**.** *There exists a constant $\tilde{L} > 0$ such that for any $n \in [N-1]_0$, it holds that*

$$\sup_{t_n \le t, s \le t_{n+1}} \left\| \nabla \log p_{T-t}(x) - \nabla \log p_{T-s}(x) \right\| \le \tilde{L} h (1 + \|x\|), \quad \forall x \in \mathbb{R}^d. \tag{17}$$

Next, we present examples satisfying the above assumption, whose proof is deferred to Appendix E.

Proposition 3.8. *Let p_0 be a Gaussian mixture in $\mathbb{R}^d$ having density law*

$$p_0(x) = \sum_{i=1}^{m} \beta_i \frac{1}{(2\pi\sigma^2)^{d/2}} \exp\left(-\frac{\|x - \mu_i\|^2}{2\sigma^2} \right),$$

where $\sigma > 0$, $\mu_i \in \mathbb{R}^d$ and $\beta_i \in [0,1]$, for $i \in [m]$, such that $\sum_{i=1}^m \beta_i = 1$. Then there exists a constant $\tilde{L} > 0$, independent of d, such that

$$\|\partial_r \nabla \log p_r(x)\| \le \tilde{L}(1 + \|x\|), \qquad r \ge 0,\ x \in \mathbb{R}^d.$$

Consequently, for any $n \in [N-1]_0$,

$$\sup_{t,s \in [t_n, t_{n+1}]} \|\nabla \log p_{T-t}(x) - \nabla \log p_{T-s}(x)\| \le \tilde{L} h (1 + \|x\|), \quad x \in \mathbb{R}^d.$$

Thanks to the above assumptions, we are able to derive a non-asymptotic error bound in $\mathcal{W}_2$-distance between the distribution of PM method (9) and the target distribution μ. Its sketch of the complete proof is provided in Section 4.

Theorem 3.9. *[Main result: error bounds for PM] Suppose that Assumptions 3.2, 3.3, 3.4 and 3.7 hold. Let $\{Y_{t_n}\}_{n \in [N]_0}$ be produced by the PM method (9) using the uniform step-size. For any $\lambda \in (0, 1/4)$, let the step-size satisfy*

$$h < \tfrac{2(1-\lambda)}{C(L,\tilde{L})} \wedge \tfrac{1}{C(L,\alpha)},$$

where $C(L,\tilde{L})$ and $C(L,\alpha)$ are given by (89) and (155), respectively. Then there exists a constant C, independent of d and h, such that,

$$\begin{aligned} &\mathcal{W}_2\left(\mu, \mathcal{L}\left(Y_{t_N}\right)\right) \\ &\le C \mathrm{e}^{C\tilde{\eta}(\alpha, M, \rho_h)} \Big(\underbrace{T^3\big(\tfrac{\sqrt{d}h}{k} + kdh^2 + dh^2 + \sqrt{d}h^{3/2}\big)}_{\text{Discretization error}} + \underbrace{\mathrm{e}^{-T}\mathcal{W}_2(\mu, \gamma^d)}_{\text{Initialization error}} + \underbrace{\epsilon_0 T}_{\text{Score matching error}} \Big). \end{aligned} \tag{18}$$

with $\tilde{\eta}(\alpha, M, \rho_h) := \eta(\alpha, M, C(L,\tilde{L})h/2 + \lambda) + \eta(\alpha, M, (3(2L+1)^2 h/2 + \lambda) \vee C(L,\alpha)h)$, and $\eta\big(\alpha, M, \rho\big)$ as defined in (30), where $L > 0$ is the Lipschitz constant of $x \mapsto \nabla \log \tilde{p}_t(x)$ from Lemma A.3 and ϵ_0 comes from Assumption 3.4.

As a direct consequence, to achieve a given precision level $\epsilon > 0$, meaning that $\mathcal{W}_2\left(\mu, \mathcal{L}\left(Y_{t_N}\right)\right) \le \epsilon$, we obtain the following result regarding the required number of iterations, the time scale T and the order of the score-matching error, whose proof is postponed to Appendix F.

Corollary 3.10. *Let Assumptions 3.2, 3.3, 3.4 and 3.7 hold. To achieve a given precision level $\epsilon > 0$ in $\mathcal{W}_2$-distance, the required number of iterations of the PM method (9) is of order $\tilde{\mathcal{O}}\big(\frac{\sqrt{d}}{\epsilon^{2/3}}\big)$, the batch size k is of order $\tilde{\mathcal{O}}\big(\frac{\sqrt{d}h}{\epsilon}\big)$ for the discretization error to be small, the time scale T is of order $\mathcal{O}\big(\log(\frac{\sqrt{d}}{\epsilon})\big)$ for the initialization error to be small, and the score-matching error ϵ_0 is of order $\mathcal{O}\Big(\frac{\epsilon}{\log(\epsilon^{-1}\sqrt{d})}\Big)$ for the propagated score-matching error to be small.*

4 Technical Overview

4.1 Regime shift property

For simplicity, in the following we denote $b_t(x) := -x + 2\nabla \log \tilde{p}_{T-t}(x)$. Then the backward SDE (5) can be rewritten as follows:

$$\mathrm{d}\overleftarrow{X}_t = b_t(\overleftarrow{X}_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t, \quad \overleftarrow{X}_0 \sim \mathcal{L}(\overrightarrow{X}_T), \qquad 0 \le t \le T. \tag{19}$$

In this subsection, we revisit the regime shift results of p_t and $b_t(x)$ established in [42], as they play a central role in proving the convergence guarantees of the two SGMs.

Regime shift of p_t.

- p_t is only weak log-concave for $t \in \big[0, \tau(\alpha, M)\big)$.
- p_t is log-concave for $t \in [\tau(\alpha, M), T]$.

Regime shift of $b_t(x)$.

- $b_t(x)$ is contractive for $t \in \big[0, \hat{T}(\alpha, M, 0)\big]$.
- $b_t(x)$ is not (necessarily) contractive for $t \in [\hat{T}(\alpha, M, 0), T]$ (i.e., it is contractive in $[\hat{T}(\alpha, M, 0), T^\star]$, while not contractive in $(T^\star, T]$).

Appendix A provides further details and the explicit formulas for $\tau(\alpha, M)$, $\hat{T}(\alpha, M, 0)$ and $T^\star$. We mentioned that the introduction of $\hat{T}(\alpha, M, 0)$ as a lower bound on $T^\star$ is not a limitation of our proof technique but a direct consequence of the fact that $T^\star$ is not explicitly computable.

4.2 Overview of proofs of Theorem 3.5 and Theorem 3.9

We sketch the proofs of Theorem 3.5 (EM) and Theorem 3.9 (PM). Before it, we introduce the following two types of auxiliary process as used in [42].

Auxiliary processes. For any $n \in [N-1]_0$, $i \in [k]_0$, we introduce the following auxiliary processes:

- $\{X^N_{t_{n,i}}\}$: discretization of $\{\overleftarrow{X}_t\}_{0 \le t \le T}$ with $X^N_0 \sim \mathcal{L}(\overleftarrow{X}_0)$;
- $\{X^\infty_{t_{n,i}}\}$: same discretization as X^N but with $X^\infty_0 \sim \gamma^d$.

For EM, both auxiliary processes follow

$$X^\star_{t_{n+1}} = X^\star_{t_n} + h b_{t_n}(X^\star_{t_n}) + \sqrt{2}\Delta W_{t_n}, \qquad \star \in \{N, \infty\}. \tag{20}$$

For PM, with refined sub-grid $t_{n,i} := t_n + ih/k$,

$$\begin{aligned} \tilde{X}^\star_{t_{n,i}} &= X^\star_{t_n} + \tfrac{ih}{k} b_{t_n}(X^\star_{t_n}) + \sqrt{\tfrac{2h}{k}} \sum_{j=0}^{i-1} Z_{t_{n,j}}, \\ X^\star_{t_{n,i+1}} &= X^\star_{t_{n,i}} + \tfrac{h}{k} b_{t_n}(X^\star_{t_n}) + h B_{t_{n,i}} \Big(b_{t_{n,i}}(\tilde{X}^\star_{t_{n,i}}) - b_{t_n}(X^\star_{t_n}) \Big) + \sqrt{\tfrac{2h}{k}}\, Z_{t_{n,i}}, \\ X^\star_{t_{n+1}} &= X^\star_{t_{n+1},0} = X^\star_{t_{n,k}}, \quad \star \in \{N, \infty\}, \end{aligned} \tag{21}$$

where $\{Z_{t_{n,i}}\}$ are i.i.d. γ^d, $\{B_{t_{n,i}}\}$ are i.i.d. $\mathrm{Bernoulli}(1/k)$, and the two families are mutually independent.

Error decomposition. By the triangle inequality, with $\mu = \mathcal{L}(\overleftarrow{X}_T)$,

$$\mathcal{W}_2\big(\mu, \mathcal{L}(Y_{t_N})\big) \le \underbrace{\mathcal{W}_2\big(\mu, \mathcal{L}(X^N_{t_N})\big)}_{\mathcal{E}_{\mathrm{disc}}} + \underbrace{\mathcal{W}_2\big(\mathcal{L}(X^N_{t_N}), \mathcal{L}(X^\infty_{t_N})\big)}_{\mathcal{E}_{\mathrm{init}}} + \underbrace{\mathcal{W}_2\big(\mathcal{L}(X^\infty_{t_N}), \mathcal{L}(Y_{t_N})\big)}_{\mathcal{E}_{\mathrm{score}}}. \tag{22}$$

The initialization and score-approximation error, i.e., $\mathcal{E}_{\mathrm{init}}$ and $\mathcal{E}_{\mathrm{score}}$, are treated in the same way for EM and PM methods, whereas the discretization term $\mathcal{E}_{\mathrm{disc}}$ is treated separately. Compared with the error analysis of EM in [42], the main novelty of our analysis lies in the refined treatment of $\mathcal{E}_{\mathrm{disc}}$.

Step 1: Discretization error $\mathcal{E}_{\mathrm{disc}}$. Combining with a general one-step recursion (Proposition B.1 for PM, and a parallel statement for EM), and iterated with the regime-shift drift bound (Proposition A.4) gives,

$$\mathcal{E}_{\mathrm{disc}} = \begin{cases} \mathcal{O}\big(\sqrt{Td}\,h\big), & \text{EM}, \\ \mathcal{O}\left(T^3\left(\frac{\sqrt{d}\,h}{k} + \sqrt{k}dh^2 + \sqrt{d}\,h^{3/2}\right)\right), & \text{PM}. \end{cases}$$

Compared with [42], our analysis for EM improves the discretization error to first order in h, while retaining the same dimension dependence. This is achieved through a refined estimate of the cross terms in the L^2 one-step error recursion, which isolates the martingale component of the local residual. Since this component has zero conditional expectation by the martingale property of the Itô integral, it makes no contribution to the cross term. For PM, we further exploit the rank-one structure of the centered Bernoulli correction to obtain a sharper mean-square bound (Lemma B.5), which yields the significant speed-up over EM. Further algorithmic and analytical intuition behind the improved bound for PM, together with the corresponding technical coupling argument, is provided in Appendix B.1.

Step 2: Initialization error $\mathcal{E}_{\mathrm{init}}$. The processes X^N and X^∞ share the same numerical randomness and differ only at t_0. Choosing (X_0^N, X_0^∞) as the $\mathcal{W}_2$-optimal coupling of $\mathcal{L}(\overleftarrow{X}_0)$ and γ^d,

$$\mathcal{E}_{\mathrm{init}}^2 \le \mathbb{E}\big[\|X_{t_N}^N - X_{t_N}^\infty\|^2\big] \le Ce^{-T}\,\mathcal{W}_2^2\big(\mu, \gamma^d\big),$$

where the exponential factor comes from the contractivity of the forward OU process (2).

Step 3: Score-approximation error $\mathcal{E}_{\mathrm{score}}$. Since X^∞ and Y differ only in the score used inside the drift, standard arguments yield

$$\mathcal{E}_{\mathrm{score}} \le C\,T\epsilon_0,$$

where ϵ_0 is the score-approximation error in Assumption 3.4.

Thus, for a target $\mathcal{W}_2$-accuracy ϵ, the above convergence rates give the improved iteration complexity $\tilde{\mathcal{O}}(\sqrt{d}/\epsilon)$ for EM (Corollary 3.6) and the sharper iteration complexity $\tilde{\mathcal{O}}(\sqrt{d}/\epsilon^{2/3})$ for PM (Corollary 3.10).

5 Conclusion and future work

In this work, we provide enhanced iteration complexity in $\mathcal{W}_2$ distance for two discretization methods applied to SGMs. First, by carrying out more delicate error estimates, we derive an $\tilde{\mathcal{O}}(\sqrt{d}/\epsilon)$ iteration complexity for the EM-based SGM beyond log-concavity, significantly improving the existing order $\tilde{\mathcal{O}}(d/\epsilon^2)$ obtained in [42]. Second, we establish an $\tilde{\mathcal{O}}(\sqrt{d}/\epsilon^{2/3})$ iteration complexity for the PM-based SGM in $\mathcal{W}_2$-distance, filling a theoretical guarantee left in [23] and improving upon existing state-of-the-art Wasserstein convergence guarantees for SDE-based diffusion samplers.

Several interesting avenues remain to be explored in our future works. First, we seek to develop new techniques that improve the dependence on α, M and L_V. The dependence we have established is exponential, whereas obtaining a polynomial dependence constitutes our future work. Second, we intend to build on this idea by developing more efficient sampling methods based on critically-damped Langevin diffusion, under weaker assumptions than weak log-concavity, and also for other metrics. Third, we plan to extend our analysis to sampling on low-dimensional manifolds, seeking convergence bounds that depend on the intrinsic rather than the ambient dimension.

References

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

[2] Vahan Arsenyan, Elen Vardanyan, and Arnak S. Dalalyan. Assessing the quality of denoising diffusion models in Wasserstein distance: noisy score and optimal bounds. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

[3] Valentin De Bortoli. Convergence of denoising diffusion models under the manifold hypothesis. *Transactions on Machine Learning Research*, 2022. Expert Certification.

[4] Stefano Bruno and Sotirios Sabanis. Wasserstein Convergence of Score-based Generative Models under Semiconvexity and Discontinuous Gradients. *Transactions on Machine Learning Research*, 2025.

[5] 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.

[6] Gabriel Cardoso, Sylvain Le Corff, Eric Moulines, et al. Monte Carlo guided denoising diffusion models for Bayesian linear inverse problems. In *The Twelfth International Conference on Learning Representations*, 2024.

[7] Patrick Cattiaux, Giovanni Conforti, Ivan Gentil, and Christian Léonard. Time reversal of diffusion processes under a finite entropy condition. *Annales de l'Institut Henri Poincaré Probabilités et Statistiques*, 59(4):1844–1881, 2023.

[8] 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*, pages 4735–4763. PMLR, 2023.

[9] 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, 2023.

[10] 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*, 2023.

[11] Hyungjin Chung, Jeongsol Kim, Michael Thompson Mccann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *The Eleventh International Conference on Learning Representations*, 2023.

[12] Giovanni Conforti. Weak semiconvexity estimates for Schrödinger potentials and logarithmic Sobolev inequality for Schrödinger bridges. *Probability Theory and Related Fields*, 189(3-4):1045–1071, 2024.

[13] 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.

[14] Giovanni Conforti, Daniel Lacker, and Soumik Pal. Projected Langevin dynamics and a gradient flow for entropic optimal transport. *Journal of the European Mathematical Society*, 2025.

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

[16] Alessio Figalli and Federico Glaudo. *An invitation to optimal transport, Wasserstein distances, and gradient flows*, volume 1. EMS press Berlin, 2021.

[17] 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 (JMLR)*, 26:Paper No. [43], 54, 2025.

[18] Xuefeng Gao and Lingjiong Zhu. Convergence analysis for general probability flow ODEs of diffusion models in Wasserstein distances. In *The 28th International Conference on Artificial Intelligence and Statistics*, 2025.

[19] Ian J 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.

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

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

[22] Yuchen Jiao, Yuchen Zhou, and Gen Li. Optimal convergence analysis of DDPM for general distributions. *arXiv preprint arXiv:2510.27562*, 2025.

[23] Saravanan Kandasamy and Dheeraj Nagaraj. The poisson midpoint method for Langevin dynamics:provably efficient discretization for diffusion models. *Advances in Neural Information Processing Systems*, 37:65972–66024, 2024.

[24] Diederik P Kingma and Max Welling. Auto-encoding variational Bayes. In *The Eleventh International Conference on Learning Representations*, 2014.

[25] Gitte Kremling, Francesco Iafrate, Mahsa Taheri, and Johannes Lederer. Non-asymptotic error bounds for probability flow ODEs under weak log-concavity. *arXiv preprint arXiv:2510.17608*, 2025.

[26] Dohyun Kwon, Ying Fan, and Kangwook Lee. Score-based generative modeling secretly minimizes the Wasserstein distance. *Advances in Neural Information Processing Systems*, 35:20205–20217, 2022.

[27] 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.

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

[29] Gen Li, Yu Huang, Timofey Efimov, Yuting Wei, Yuejie Chi, and Yuxin Chen. Accelerating convergence of score-based diffusion models, provably. In *Forty-first International Conference on Machine Learning*, 2024.

[30] Gen Li and Yuchen Jiao. Improved convergence rate for diffusion probabilistic models. In *The Thirteenth International Conference on Learning Representations*, 2025.

[31] 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*, 2024.

[32] Gen Li and Yuling Yan. $O(d/T)$ convergence theory for diffusion probabilistic models under minimal assumptions. *Journal of Machine Learning Research (JMLR)*, 26:Paper No. [292], 55, 2025.

[33] Xiang Li, John Thickstun, Ishaan Gulrajani, Percy S Liang, and Tatsunori B Hashimoto. Diffusion-lm improves controllable text generation. *Advances in Neural Information Processing Systems*, 35:4328–4343, 2022.

[34] Xingchao Liu, Lemeng Wu, Mao Ye, and qiang liu. Let us build bridges: understanding and extending diffusion generative models. In *NeurIPS 2022 Workshop on Score-Based Methods*, 2022.

[35] Mateusz B Majka, Aleksandar Mijatović, and Łukasz Szpruch. Nonasymptotic bounds for sampling algorithms without log-concavity. *The Annals of Applied Probability*, 30(4):1534–1581, 2020.

[36] Xuerong Mao. *Stochastic differential equations and applications*. Elsevier, 2007.

[37] David Nualart. *The Malliavin calculus and related topics*. Springer, 2006.

[38] Vadim Popov, Ivan Vovk, Vladimir Gogoryan, Tasnima Sadekova, and Mikhail Kudinov. Grad-tts: a diffusion probabilistic model for text-to-speech. In *International Conference on Machine Learning*, pages 8599–8608. PMLR, 2021.

[39] Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International Conference on Machine Learning*, pages 1530–1538. PMLR, 2015.

[40] Adrien Saumard and Jon A. Wellner. Log-concavity and strong log-concavity: a review. *Statistics Surveys*, 8:45–114, 2014.

[41] Ruoqi Shen and Yin Tat Lee. The randomized midpoint method for log-concave sampling. *Advances in Neural Information Processing Systems*, 32, 2019.

[42] Marta Gentiloni Silveri and Antonio Ocello. Beyond log-concavity and score regularity: improved convergence bounds for score-based generative models in $\mathcal{W}_2$-distance. In *Forty-second International Conference on Machine Learning*, 2025.

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

[44] Yang Song, Sahaj Garg, Jiaxin Shi, and Stefano Ermon. Sliced score matching: a scalable approach to density and score estimation. In *Uncertainty in artificial intelligence*, pages 574–584. PMLR, 2020.

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

[46] Rishikesh Srinivasan and Dheeraj Nagaraj. Poisson midpoint method for log concave sampling: beyond the strong error lower bounds. In *International Conference on Learning Representations*, 2026.

[47] Stanislas Strasman, Antonio Ocello, Claire Boyer, Sylvain Le Corff, and Vincent Lemaire. An analysis of the noise schedule for score-based generative models. *Transactions on Machine Learning Research Journal*, 2025.

[48] Stanislas Strasman, Sobihan Surendran, Claire Boyer, Sylvain Le Corff, Vincent Lemaire, and Antonio Ocello. Wasserstein convergence of critically damped Langevin diffusions. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

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

[50] Xixian Wang and Zhongjian Wang. Wasserstein Bounds for generative diffusion models with Gaussian tail targets. *Transactions on Machine Learning Research*, 2026.

[51] Yuchen Wu, Yuxin Chen, and Yuting Wei. Stochastic Runge-Kutta methods: provable acceleration of diffusion models. *arXiv preprint arXiv:2410.04760*, 2024.

[52] Yifeng Yu and Lu Yu. Advancing Wasserstein convergence analysis of score-based models: insights from discretization and second-order acceleration. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

[53] Matthew S Zhang, Stephen Huan, Jerry Huang, Nicholas M Boffi, Sitan Chen, and Sinho Chewi. Sublinear iterations can suffice even for DDPMs. *arXiv preprint arXiv:2511.04844*, 2025.

A Propagation of the regularity assumption

The aim of this section is to revisit some regime shift properties established in [42]. The following proposition shows how the constant of weak log-concavity propagates. The proof follows the similar argument of [14, Lemma 5.9] and is omitted.

Proposition A.1 (Propagation of weak log-concavity in time)**.** *Let Assumption 3.2 hold. Then, for any $t \in [0,T]$, the function $-\log \tilde{p}_{T-t}(x)$ is weakly convex with weak convexity profile $\kappa_{-\log \tilde{p}_{T-t}}$ satisfying*

$$\kappa_{-\log \tilde{p}_{T-t}}(r) \geq \frac{\alpha}{\alpha + (1-\alpha)e^{-2(T-t)}} - 1 - \frac{e^{-(T-t)}}{\alpha + (1-\alpha)e^{-2(T-t)}} \frac{1}{r} f_M \left(\frac{e^{-(T-t)}}{\alpha + (1-\alpha)e^{-2(T-t)}} r \right), \tag{23}$$

where α, M come from Assumption 3.2. Moreover, we have, for any $x, y \in \mathbb{R}^d$,

$$\left\langle \nabla \log \tilde{p}_{T-t}(x) - \nabla \log \tilde{p}_{T-t}(y), x - y \right\rangle \leq -C_{T-t} \|x - y\|^2, \tag{24}$$

where C_t is defined as

$$C_t := \frac{\alpha}{\alpha + (1-\alpha)e^{-2t}} - \frac{e^{-2t} M}{(\alpha + (1-\alpha)e^{-2t})^2} - 1. \tag{25}$$

As a direct consequence of $\nabla \log \tilde{p}_t(x) = x + \nabla \log p_t(x)$ for any $x \in \mathbb{R}^d$ and $t \in [0,T]$, the weak log-concavity constant of the density p_t is $C_t + 1$. The estimate for $(C_t + 1)$ is coherent with the intuition we have on the SGMs. Such a constant starts from $C_0 + 1 = \alpha - M$, corresponding to the weak log-concavity constant of μ, and converges to the

$$C_t + 1 = \frac{\alpha}{\alpha + (1-\alpha)e^{-2t}} - \frac{e^{-2t} M}{(\alpha + (1-\alpha)e^{-2t})^2} \longrightarrow 1 \quad \text{as } t \to \infty,$$

which is the log-concavity constant of the standard Gaussian law γ^d.

In what follows, we turn to the regime-shift property of p_t, which underlies the contraction estimates used in the proof of Theorem 3.5 and Theorem 3.9

Proposition A.2 (Regime shift of p_t)**.** *Let Assumption 3.2 hold. For any $t \in [0,T]$, it holds that*

$$\begin{cases} p_t \text{ is weakly log-concave,} & t \in \big[0, \tau(\alpha, M)\big] \\ p_t \text{ is strongly log-concave,} & t \in \big(\tau(\alpha, M), T\big], \end{cases} \tag{26}$$

with

$$\tau(\alpha, M) := \begin{cases} \log\left(\sqrt{\frac{\alpha^2+M-\alpha}{\alpha^2}}\right) \wedge T, & \text{if} \quad \alpha - M < 0, \\ 0, & \text{otherwise.} \end{cases} \tag{27}$$

An interesting property of Proposition A.2 is that the law with density p_t becomes strongly log-concave after a finite amount of time, even if p_0 is only weakly log-concave. Indeed, when $\alpha - M < 0$, we have that $\frac{\alpha^2+M-\alpha}{\alpha^2} > 1$. Therefore, $\tau(\alpha, M) > 0$ and two distinct regimes are present. Intuitively, it states that, if $\alpha - M \geq 0$, i.e. p_0 is strongly log-concave, then p_t is guaranteed to remain strongly log-concave in the whole interval $[0,T]$, which matches the property of [40, Theorem 3.7].

We introduce the time thresholds used to separate the contractive and non-contractive parts of the reverse dynamics. Denote by $T^\star$ the following time

$$T^\star := \inf\big\{t \in [0,T] : 2\kappa_{-\log \tilde{p}_{T-t}} + 1 \leq 0\big\}, \tag{28}$$

defining $\inf \varnothing := T$. Also, denote by

$$\widehat{T}(\alpha, M, \rho) := \sup\left\{t \in [0,T] : 2C_{T-t} + 1 \geq \rho\right\}, \tag{29}$$

for $\rho \in [0,1)$, with $\inf \varnothing := T$. Since the weak profile of $-\log \tilde{p}_t$ satisfies $\kappa_{-\log \tilde{p}_t} \geq C_t$, with C_t being defined as in (25), we have that $\widehat{T}(\alpha, M, \rho) \leq T^\star$, for any $\rho \in [0,1)$. Next, we define

$$N_h := \sup\big\{m \in \{0, \ldots, Nk\} : mh/k \leq \widehat{T}(\alpha, M, \rho) = T - \eta(\alpha, M, \rho)\big\},$$

with

$$\eta(\alpha, M, \rho) := \begin{cases} \frac{1}{2}\log\left(\frac{M+\rho\alpha(1-\alpha)}{(1-\rho)\alpha^2} + \sqrt{\frac{(1+\rho)(1-\alpha)^2}{(1-\rho)\alpha^2} + \left(\frac{M+\rho\alpha(1-\alpha)}{(1-\rho)\alpha^2}\right)^2}\right) \wedge T, & \text{if } 2\alpha - 2M - 1 < 0 \\ 0, & \text{otherwise.} \end{cases} \tag{30}$$

Note that the condition

$$\frac{M+\rho\alpha(1-\alpha)}{(1-\rho)\alpha^2} + \sqrt{\frac{(1+\rho)(1-\alpha)^2}{(1-\rho)\alpha^2} + \left(\frac{M+\rho\alpha(1-\alpha)}{(1-\rho)\alpha^2}\right)^2} > 1$$

is equivalent to the condition $2\alpha - 2M - 1 < 0$. This means that $\widehat{T}(\alpha, M, \rho)$ is well-defined and $\widehat{T}(\alpha, M, \rho) \in [0,T]$. For clarity, we sketch the above regime shift of the drift b_t of backward SDE (5) in the following Figure 1 with the help of above preliminaries.

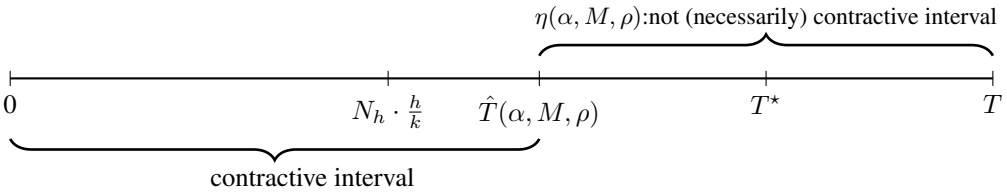

Figure 1: Regime shift of the drift $b_t(x)$.

Lemma A.3 (Lipschitz condition for the modified score)**.** *Let Assumptions 3.2 and 3.3 hold. Then, for any $t \in [0,T)$, it holds that*

$$\sup_{x \in \mathbb{R}^d} \|\nabla^2 \log \tilde{p}_{T-t}(x)\| \leq L_t \leq L, \tag{31}$$

where

$$L_t := \left(\left(\tfrac{1}{1-e^{-2(T-t)}} \wedge e^{2(T-t)} L_V\right) \vee \big(-(C_t + 1)\big)\right) + 1, \quad L := (1/\alpha \vee 1)(2 + L_V).$$

In particular, the modified score function $x \mapsto \nabla \log \tilde{p}_t(x)$ is L-Lipschitz, i.e., for any $x, y \in \mathbb{R}^d$,

$$\big\|\nabla \log \tilde{p}_t(x) - \nabla \log \tilde{p}_t(y)\big\| \leq L\|x - y\|. \tag{32}$$

See [42, Proposition B.4] for a similar proof of this lemma. The following proposition establishes a regime shift result of the drift $b_t(x)$ of the time-reversal process.

Proposition A.4 (Regime shift of $b_t(x)$)**.** *Let Assumptions 3.2 and 3.3 hold. It holds that, for any $x, y \in \mathbb{R}^d$, and $t \in [0, T]$,*

$$\big\langle b_t(x) - b_t(y), x - y \big\rangle \le G_t \|x - y\|^2, \tag{33}$$

with

$$G_t := \begin{cases} -(2C_{T-t} + 1), & t \in \big[0, \widehat{T}(\alpha, M, \rho)\big], \\ 2L + 1, & t \in \big(\widehat{T}(\alpha, M, \rho), T\big]. \end{cases}$$

B Bound on discretization error

B.1 Discretization error of PM method

First we prove the following proposition, describing the error propagation between the PM auxiliary process (21) and the backward process SDE (5).

Proposition B.1. *Suppose that Assumptions 3.2 and 3.3 hold. Let $\{\overleftarrow{X}_t\}_{t\in[0,T]}$ be the solution of SDE (5). For $n \in [N-1]_0$ and $i \in [k-1]_0$, denote*

$$\Delta W_{t_{n,i}} := W(t_{n,i+1}) - W(t_{n,i}), \qquad \xi_{t_{n,i}} := \sqrt{k/h}\Delta W_{t_{n,i}}. \tag{34}$$

Define the local drift residual

$$R_{t_{n,i}} := \int_{t_{n,i}}^{t_{n,i+1}} \big[b_s(\overleftarrow{X}_s) - b_{t_{n,i}}(\overleftarrow{X}_{t_{n,i}})\big]\,\mathrm{d}s, \tag{35}$$

and the random perturbation

$$\begin{aligned} M_{t_{n,i}} := -\sqrt{\tfrac{2h}{k}}\Big[& Z_{t_{n,i}} - \xi_{t_{n,i}} + \sqrt{\tfrac{hk}{2}}\big(B_{t_{n,i}} - 1/k\big)\Big(b_{t_{n,i}}(\tilde{X}^N_{t_{n,i}}) - b_{t_n}(X^N_{t_n})\Big) \\ & + \sqrt{\tfrac{h}{2k}}\Big(b_{t_{n,i}}(\tilde{X}^N_{t_{n,i}}) - b_{t_{n,i}}(X^N_{t_{n,i}})\Big)\Big]. \end{aligned} \tag{36}$$

Then, for any $\lambda \in (0, 1/4)$, we have

$$\begin{aligned} \mathbb{E}\Big[\big\|\overleftarrow{X}_{t_{n,i+1}} - X^N_{t_{n,i+1}}\big\|^2\Big] \le \gamma_{t_{n,i}}\,\mathbb{E}\Big[\big\|\overleftarrow{X}_{t_{n,i}} - X^N_{t_{n,i}}\big\|^2\Big] + \tfrac{k}{\lambda h}\mathbb{E}\Big[\big\|\mathbb{E}\big[R_{t_{n,i}} \mid \mathcal{F}_{t_{n,i}}\big]\big\|^2\Big] \\ + 3\,\mathbb{E}\big[\|R_{t_{n,i}}\|^2\big] + 3\,\mathbb{E}\big[\|M_{t_{n,i}}\|^2\big] + \tfrac{(2L+1)^2 h}{\lambda k}\mathbb{E}\Big[\big\|X^N_{t_{n,i}} - \tilde{X}^N_{t_{n,i}}\big\|^2\Big], \end{aligned} \tag{37}$$

where L comes from Lemma A.3 and

$$\gamma_{t_{n,i}} := \begin{cases} 1 + \Big(\frac{3(2L+1)^2 h}{k} - 2(2C_{T-t_{n,i}} + 1) + 2\lambda\Big)\frac{h}{k}, & \text{for } nk + i \le N_h \\ 1 + \big(\frac{3(2L+1)^2 h}{k} + 2(2L+1) + 2\lambda\big)\frac{h}{k}, & \text{for } nk + i > N_h. \end{cases}$$

Proof. For any $n \in [N-1]$, $i \in [k-1]_0$, it follows from the definition (5) that

$$\overleftarrow{X}_{t_{n,i+1}} = \overleftarrow{X}_{t_{n,i}} + \tfrac{h}{k} b_{t_{n,i}}(\overleftarrow{X}_{t_{n,i}}) + \sqrt{2}\Delta W_{t_{n,i}} + R_{t_{n,i}}, \tag{38}$$

where $\Delta W_{t_{n,i}} := W(t_{n,i+1}) - W(t_{n,i})$ and $R_{t_{n,i}}$ is defined as (35). Next, we rewrite (21) as follows:

$$\begin{aligned} X^N_{t_{n,i+1}} = X^N_{t_{n,i}} + \tfrac{h}{k} b_{t_{n,i}}(X^N_{t_{n,i}}) + \sqrt{2}\Delta W_{t_{n,i}} \\ + \sqrt{\tfrac{2h}{k}}\Big[Z_{t_{n,i}} - \xi_{t_{n,i}} + \sqrt{\tfrac{hk}{2}}\big(B_{t_{n,i}} - 1/k\big)\Big(b_{t_{n,i}}(\tilde{X}^N_{t_{n,i}}) - b_{t_n}(X^N_{t_n})\Big) \\ + \sqrt{\tfrac{h}{2k}}\Big(b_{t_{n,i}}(\tilde{X}^N_{t_{n,i}}) - b_{t_{n,i}}(X^N_{t_{n,i}})\Big)\Big]. \end{aligned} \tag{39}$$

Here, we denote $\sqrt{\frac{2h}{k}}\xi_{t_{n,i}} := \sqrt{2}\Delta W_{t_{n,i}}$ for sake of notation. Subtracting (39) from (38) yields

$$\begin{aligned}\overleftarrow{X}_{t_{n,i+1}} - X^N_{t_{n,i+1}} &= \overleftarrow{X}_{t_{n,i}} - X^N_{t_{n,i}} + \tfrac{h}{k}\Big(b_{t_{n,i}}(\overleftarrow{X}_{t_{n,i}}) - b_{t_{n,i}}(X^N_{t_{n,i}})\Big) + R_{t_{n,i}} \\ &\quad - \sqrt{\tfrac{2h}{k}}\Big[Z_{t_{n,i}} - \xi_{t_{n,i}} + \sqrt{\tfrac{hk}{2}}\big(B_{t_{n,i}} - 1/k\big)\Big(b_{t_{n,i}}(\tilde{X}^N_{t_{n,i}}) - b_{t_n}(X^N_{t_n})\Big) \\ &\quad + \sqrt{\tfrac{h}{2k}}\Big(b_{t_{n,i}}(\tilde{X}^N_{t_{n,i}}) - b_{t_{n,i}}(X^N_{t_{n,i}})\Big)\Big].\end{aligned} \tag{40}$$

For the sake of convenience in notation, we denote, for any $n \in [N-1]_0$, $i \in [k]_0$,

$$\varepsilon_{t_{n,i}} := \overleftarrow{X}_{t_{n,i}} - X^N_{t_{n,i}}, \tag{41}$$

and recall $M_{t_{n,i}}$ from (36). Squaring both sides of (40) and taking expectations shows,

$$\begin{aligned}\mathbb{E}\big[\|\varepsilon_{t_{n,i+1}}\|^2\big] &= \mathbb{E}\big[\|\varepsilon_{t_{n,i}}\|^2\big] + \tfrac{h^2}{k^2}\,\mathbb{E}\Big[\big\|b_{t_{n,i}}(\overleftarrow{X}_{t_{n,i}}) - b_{t_{n,i}}(X^N_{t_{n,i}})\big\|^2\Big] + \mathbb{E}\Big[\big\|R_{t_{n,i}}\big\|^2\Big] \\ &\quad + \mathbb{E}\Big[\big\|M_{t_{n,i}}\big\|^2\Big] + \tfrac{2h}{k}\,\mathbb{E}\Big[\big\langle \varepsilon_{t_{n,i}}, b_{t_{n,i}}(\overleftarrow{X}_{t_{n,i}}) - b_{t_{n,i}}(X^N_{t_{n,i}})\big\rangle\Big] + 2\,\mathbb{E}\Big[\big\langle \varepsilon_{t_{n,i}}, R_{t_{n,i}}\big\rangle\Big] \\ &\quad + 2\,\mathbb{E}\Big[\big\langle \varepsilon_{t_{n,i}}, M_{t_{n,i}}\big\rangle\Big] + \tfrac{2h}{k}\,\mathbb{E}\Big[\big\langle b_{t_{n,i}}(\overleftarrow{X}_{t_{n,i}}) - b_{t_{n,i}}(X^N_{t_{n,i}}), R_{t_{n,i}}\big\rangle\Big] \\ &\quad + \tfrac{2h}{k}\,\mathbb{E}\Big[\big\langle b_{t_{n,i}}(\overleftarrow{X}_{t_{n,i}}) - b_{t_{n,i}}(X^N_{t_{n,i}}), M_{t_{n,i}}\big\rangle\Big] + 2\,\mathbb{E}\Big[\big\langle R_{t_{n,i}}, M_{t_{n,i}}\big\rangle\Big].\end{aligned} \tag{42}$$

For the estimate of the fifth term of (42), using Proposition A.4 yields,

$$\tfrac{2h}{k}\,\mathbb{E}\Big[\big\langle \varepsilon_{t_{n,i}}, b_{t_{n,i}}(\overleftarrow{X}_{t_{n,i}}) - b_{t_{n,i}}(X^N_{t_{n,i}})\big\rangle\Big] \le \begin{cases} \frac{-2(2C_{T-t_{n,i}}+1)h}{k}\,\mathbb{E}\big[\|\varepsilon_{t_{n,i}}\|^2\big], & \text{for} \quad nk+i \le N_h, \\ \frac{2(2L+1)h}{k}\,\mathbb{E}\big[\|\varepsilon_{t_{n,i}}\|^2\big], & \text{for} \quad nk+i > N_h. \end{cases} \tag{43}$$

Now, we deal with $2\,\mathbb{E}\big[\langle \varepsilon_{t_{n,i}}, R_{t_{n,i}}\rangle\big]$. In particular, if we let $\mathcal{F}_{t_{n,i}}$ denote the σ-algebra generated by $\{\overleftarrow{X}_{t_{n,i}}\}$, $\{\tilde{X}^N_{t_{n,i}}\}$, and $\{X^N_{t_{n,i}}\}$ for any $n \in [N]_0$ and $i \in [k-1]_0$. Since $\varepsilon_{t_{n,i}}$ is $\mathcal{F}_{t_{n,i}}$-measurable, we apply the Young inequality with $0 < \lambda < 1/4$ to obtain that

$$\begin{aligned}2\,\mathbb{E}\Big[\big\langle \varepsilon_{t_{n,i}}, R_{t_{n,i}}\big\rangle\Big] &= 2\,\mathbb{E}\Big[\Big\langle \varepsilon_{t_{n,i}}, \mathbb{E}\big[R_{t_{n,i}}\big|\mathcal{F}_{t_{n,i}}\big]\Big\rangle\Big] \\ &\le \tfrac{\lambda h}{k}\,\mathbb{E}\Big[\big\|\varepsilon_{t_{n,i}}\big\|^2\Big] + \tfrac{k}{\lambda h}\mathbb{E}\Big[\big\|\mathbb{E}\big[R_{t_{n,i}} \mid \mathcal{F}_{t_{n,i}}\big]\big\|^2\Big].\end{aligned} \tag{44}$$

For the term involving $M_{t_{n,i}}$, conditioning on $\mathcal{F}_{t_{n,i}}$ gives

$$\mathbb{E}\Big[Z_{t_{n,i}} - \xi_{t_{n,i}} + \sqrt{\tfrac{hk}{2}}\big(B_{t_{n,i}} - 1/k\big)\Big(b_{t_{n,i}}(\tilde{X}^N_{t_{n,i}}) - b_{t_n}(X^N_{t_n})\Big)\Big|\mathcal{F}_{t_{n,i}}\Big] = 0, \tag{45}$$

where we use that $Z_{t_{n,i}}$, $\xi_{t_{n,i}}$, and $B_{t_{n,i}}$ are independent of $\mathcal{F}_{t_{n,i}}$. Hence, for $2\,\mathbb{E}\big[\langle \varepsilon_{t_{n,i}}, M_{t_{n,i}}\rangle\big]$, the Young inequality and the Lipschitz condition of $b_t(\cdot)$ imply

$$\begin{aligned}&2\,\mathbb{E}\Big[\big\langle \varepsilon_{t_{n,i}}, M_{t_{n,i}}\big\rangle\Big] = 2\,\mathbb{E}\Big[\Big\langle \varepsilon_{t_{n,i}}, \mathbb{E}\big[M_{t_{n,i}}\big|\mathcal{F}_{t_{n,i}}\big]\Big\rangle\Big] \\ &= -2\sqrt{\tfrac{2h}{k}}\,\mathbb{E}\Big[\Big\langle \varepsilon_{t_{n,i}}, \mathbb{E}\Big[Z_{t_{n,i}} - \xi_{t_{n,i}} + \sqrt{\tfrac{hk}{2}}\big(B_{t_{n,i}} - 1/k\big)\Big(b_{t_{n,i}}(\tilde{X}^N_{t_{n,i}}) - b_{t_n}(X^N_{t_n})\Big)\Big]\Big|\mathcal{F}_{t_{n,i}}\Big]\Big\rangle\Big] \\ &\quad + \tfrac{2h}{k}\,\mathbb{E}\Big[\big\langle \varepsilon_{t_{n,i}}, b_{t_{n,i}}(X^N_{t_{n,i}}) - b_{t_{n,i}}(\tilde{X}^N_{t_{n,i}})\big\rangle\Big] \\ &\le \tfrac{\lambda h}{k}\,\mathbb{E}\big[\|\varepsilon_{t_{n,i}}\|^2\big] + \tfrac{(2L+1)^2 h}{\lambda k}\mathbb{E}\Big[\big\|X^N_{t_{n,i}} - \tilde{X}^N_{t_{n,i}}\big\|^2\Big].\end{aligned} \tag{46}$$

Regarding the last three terms of (42), using the Young inequality and Lipschitz condition of $b_t(\cdot)$ leads to

$$\begin{aligned}\tfrac{2h}{k}\,\mathbb{E}\Big[\big\langle b_{t_{n,i}}(\overleftarrow{X}_{t_{n,i}}) - b_{t_{n,i}}(X^N_{t_{n,i}}), R_{t_{n,i}}\big\rangle\Big] &\le \tfrac{(2L+1)^2h^2}{k^2}\,\mathbb{E}\Big[\big\|\varepsilon_{t_{n,i}}\big\|^2\Big] + \mathbb{E}\Big[\big\|R_{t_{n,i}}\big\|^2\Big], \\ \tfrac{2h}{k}\,\mathbb{E}\Big[\big\langle b_{t_{n,i}}(\overleftarrow{X}_{t_{n,i}}) - b_{t_{n,i}}(X^N_{t_{n,i}}), M_{t_{n,i}}\big\rangle\Big] &\le \tfrac{(2L+1)^2h^2}{k^2}\,\mathbb{E}\Big[\big\|\varepsilon_{t_{n,i}}\big\|^2\Big] + \mathbb{E}\Big[\big\|M_{t_{n,i}}\big\|^2\Big],\end{aligned} \tag{47}$$

and

$$2\,\mathbb{E}\Big[\big\langle R_{t_{n,i}}, M_{t_{n,i}}\big\rangle\Big] \le \mathbb{E}\Big[\big\|R_{t_{n,i}}\big\|^2\Big] + \mathbb{E}\Big[\big\|M_{t_{n,i}}\big\|^2\Big]. \tag{48}$$

Plugging the above estimates into (42) yields

$$\begin{aligned}\mathbb{E}\big[\|\varepsilon_{t_{n,i+1}}\|^2\big] \le \gamma_{t_{n,i}}\,\mathbb{E}\big[\|\varepsilon_{t_{n,i}}\|^2\big] + \tfrac{k}{\lambda h}\mathbb{E}\Big[\big\|\mathbb{E}\big[R_{t_{n,i}} \mid \mathcal{F}_{t_{n,i}}\big]\big\|^2\Big] + 3\,\mathbb{E}\Big[\big\|R_{t_{n,i}}\big\|^2\Big] \\ + 3\,\mathbb{E}\Big[\big\|M_{t_{n,i}}\big\|^2\Big] + \tfrac{(2L+1)^2 h}{\lambda k}\mathbb{E}\Big[\big\|X^N_{t_{n,i}} - \tilde{X}^N_{t_{n,i}}\big\|^2\Big],\end{aligned} \tag{49}$$

where

$$\gamma_{t_{n,i}} := \begin{cases} 1 + \big(\frac{3(2L+1)^2 h}{k} - 2(2C_{T-t_{n,i}} + 1) + 2\lambda\big)\frac{h}{k}, & \text{for } nk + i \le N_h \\ 1 + \big(\frac{3(2L+1)^2 h}{k} + 2(2L+1) + 2\lambda\big)\frac{h}{k}, & \text{for } nk + i > N_h. \end{cases}$$

The proof is completed. □

The remainder of this subsection collects the auxiliary estimates needed to iterate Proposition B.1, starting with moment bounds of the processes $\{\overleftarrow{X}_t\}_{t\in[0,T]}$ and $\{X^N_{t_{n,i}}\}_{n\in[N]_0, i\in[k-1]_0}$ defined by (5) and (21).

Lemma B.2. *Let Assumption 3.2 hold. Then μ admits a finite $2p$-th order moment for any $p \in [1,2]$. That is, there exists a constant $\mathcal{M}$, independent of d and t, satisfying*

$$\mathbb{E}\big[\|x_0\|^{2p}\big] \le \mathcal{M} d^p. \tag{50}$$

Moreover, the process $\big\{\overleftarrow{X}_t\big\}_{t\in[0,T]}$ defined by (5) *satisfies, for any $t \in [0,T]$,*

$$\mathbb{E}\Big[\big\|\overleftarrow{X}_t\big\|^{2p}\Big] \le \mathcal{M}_1 d^p, \tag{51}$$

where $\mathcal{M}_1$ is independent of d and t.

The case $p = 1$ is established in [42, Proposition B.1 and Lemma D.1]; the extension to $p \in [1,2]$ follows from Assumption 3.2 by an analogous argument and is therefore omitted. We mentioned that $\gamma_{t_{n,i}}$ and $\Gamma_{t_{n,i}}$ are generic finite constants whose values may vary across different propositions, but will not depend upon the dimension d.

Lemma B.3. *Let Assumptions 3.2, 3.3 and 3.7 hold. The process $\{X^N_{t_n}\}_{n\in[N]_0}$ generated by* (21) *using the uniform step-size. For any $\lambda \in (0, 1/4)$, let the step-size satisfy*

$$h < \tfrac{2(1-\lambda)}{C(L,\tilde{L})}, \tag{52}$$

where $C(L,\tilde{L})$ is given by (89)*. Then, there exists a constant $\mathcal{M}_2$, satisfying, for any $n \in [N]_0$,*

$$\mathbb{E}\Big[\big\|X^N_{t_n}\big\|^4\Big] \le \mathcal{M}_2 T^5 d^2, \tag{53}$$

where $\mathcal{M}_2 := C\mathrm{e}^{C\eta(\alpha,M,C(L,\tilde{L})h/2+\lambda)}$, and $\eta(\alpha, M, C(L,\tilde{L})h/2+\lambda)$ as defined in (30)*, and C is a constant independent of d and h.*

Proof. Recalling (21), we have, for any $n \in [N-1]_0$,

$$X^N_{t_{n+1}} = X^N_{t_n} + h b_{t_n}(X^N_{t_n}) + h\sum_{j=0}^{k-1} B_{t_{n,j}}\Big(b_{t_{n,j}}(\tilde{X}^N_{t_{n,j}}) - b_{t_n}(X^N_{t_n})\Big) + \sqrt{\tfrac{2h}{k}}\sum_{j=0}^{k-1} Z_{t_{n,j}}. \tag{54}$$

Taking square on both sides shows

$$
\begin{aligned}
&\big\|X^N_{t_{n+1}}\big\|^2\\
&= \big\|X^N_{t_n}\big\|^2 + h^2\big\|b_{t_n}(X^N_{t_n})\big\|^2 + h^2\Big\|\sum_{j=0}^{k-1} B_{t_{n,j}}\Big(b_{t_{n,j}}(\tilde{X}^N_{t_{n,j}}) - b_{t_n}(X^N_{t_n})\Big)\Big\|^2 + \tfrac{2h}{k}\Big\|\sum_{j=0}^{k-1} Z_{t_{n,j}}\Big\|^2\\
&\quad + 2h\Big\langle X^N_{t_n}, b_{t_n}(X^N_{t_n})\Big\rangle + 2h\Big\langle X^N_{t_n}, \sum_{j=0}^{k-1} B_{t_{n,j}}\Big(b_{t_{n,j}}(\tilde{X}^N_{t_{n,j}}) - b_{t_n}(X^N_{t_n})\Big)\Big\rangle\\
&\quad + 2\sqrt{\tfrac{2h}{k}}\Big\langle X^N_{t_n}, \sum_{j=0}^{k-1} Z_{t_{n,j}}\Big\rangle + 2h^2\Big\langle b_{t_n}(X^N_{t_n}), \sum_{j=0}^{k-1} B_{t_{n,j}}\Big(b_{t_{n,j}}(\tilde{X}^N_{t_{n,j}}) - b_{t_n}(X^N_{t_n})\Big)\Big\rangle\\
&\quad + 2h\sqrt{\tfrac{2h}{k}}\Big\langle b_{t_n}(X^N_{t_n}), \sum_{j=0}^{k-1} Z_{t_{n,j}}\Big\rangle + 2h\sqrt{\tfrac{2h}{k}}\Big\langle \sum_{j=0}^{k-1} B_{t_{n,j}}\Big(b_{t_{n,j}}(\tilde{X}^N_{t_{n,j}}) - b_{t_n}(X^N_{t_n})\Big), \sum_{j=0}^{k-1} Z_{t_{n,j}}\Big\rangle.
\end{aligned}
\tag{55}
$$

Next using the Young inequality gives

$$
\begin{aligned}
&\big\|X^N_{t_{n+1}}\big\|^2\\
&\le \big\|X^N_{t_n}\big\|^2 + 3h^2\big\|b_{t_n}(X^N_{t_n})\big\|^2 + 3h^2\Big\|\sum_{j=0}^{k-1} B_{t_{n,j}}\Big(b_{t_{n,j}}(\tilde{X}^N_{t_{n,j}}) - b_{t_n}(X^N_{t_n})\Big)\Big\|^2 + \tfrac{6h}{k}\Big\|\sum_{j=0}^{k-1} Z_{t_{n,j}}\Big\|^2\\
&\quad + 2h\Big\langle X^N_{t_n}, b_{t_n}(X^N_{t_n})\Big\rangle + 2h\Big\langle X^N_{t_n}, \sum_{j=0}^{k-1} B_{t_{n,j}}\Big(b_{t_{n,j}}(\tilde{X}^N_{t_{n,j}}) - b_{t_n}(X^N_{t_n})\Big)\Big\rangle\\
&\quad + 2\sqrt{\tfrac{2h}{k}}\Big\langle X^N_{t_n}, \sum_{j=0}^{k-1} Z_{t_{n,j}}\Big\rangle.
\end{aligned}
\tag{56}
$$

Before proceeding further, by the definition of $\tilde{X}^N_{t_{n,j}}$ in (21), Assumption 3.7 and the triangle inequality, we have, for any $j \in [k-1]_0$,

$$
\begin{aligned}
&\big\|b_{t_{n,j}}(\tilde{X}^N_{t_{n,j}}) - b_{t_n}(X^N_{t_n})\big\|\\
&\le (2L+1)\big\|\tilde{X}^N_{t_{n,j}} - X^N_{t_n}\big\| + \tilde{L}h\Big(\big\|X^N_{t_n}\big\| + 1\Big)\\
&\le (2L+1)\Big\|\tfrac{jh}{k} b_{t_n}(X^N_{t_n}) + \sum_{l=0}^{j-1}\sqrt{\tfrac{2h}{k}} Z_{t_{n,l}}\Big\| + \tilde{L}h\Big(\big\|X^N_{t_n}\big\| + 1\Big)\\
&\le \big((2L+1)^2 + \tilde{L}\big)h\big\|X^N_{t_n}\big\| + (2L+1)\sqrt{\tfrac{2h}{k}}\Big\|\sum_{l=0}^{j-1} Z_{t_{n,l}}\Big\| + (2L+1)h\big\|b_{t_n}(0)\big\| + \tilde{L}h,
\end{aligned}
\tag{57}
$$

where for short we denote

$$
\mathcal{I}_{t_{n,j}} := (2L+1)\sqrt{\tfrac{2h}{k}}\Big\|\sum_{l=0}^{j-1} Z_{t_{n,l}}\Big\| + (2L+1)h\big\|b_{t_n}(0)\big\| + \tilde{L}h. \tag{58}
$$

Thanks to the above estimate, the short hand notation (58), the Cauchy-Schwarz inequality and the Young inequality, one can derive

$$
\begin{aligned}
&2h\Big\langle X^N_{t_n}, \sum_{j=0}^{k-1} B_{t_{n,j}}\Big(b_{t_{n,j}}(\tilde{X}^N_{t_{n,j}}) - b_{t_n}(X^N_{t_n})\Big)\Big\rangle\\
&\le 2h\sum_{j=0}^{k-1} B_{t_{n,j}}\big\|X^N_{t_n}\big\| \cdot \Big\|b_{t_{n,j}}(\tilde{X}^N_{t_{n,j}}) - b_{t_n}(X^N_{t_n})\Big\|\\
&\le \big(2\big((2L+1)^2 + \tilde{L}\big) + 1\big)h^2\sum_{j=0}^{k-1} B_{t_{n,j}}\big\|X^N_{t_n}\big\|^2 + \sum_{j=0}^{k-1} B_{t_{n,j}}\mathcal{I}^2_{t_{n,j}}.
\end{aligned}
\tag{59}
$$

Regarding the third term of (56), in view of (57), one can obtain

$$\begin{aligned}
&\Big\|\sum_{j=0}^{k-1} B_{t_{n,j}}\Big(b_{t_{n,j}}(\tilde{X}^N_{t_{n,j}}) - b_{t_n}(X^N_{t_n})\Big)\Big\|^2 \\
&\le \Big(\sum_{j=0}^{k-1} B_{t_{n,j}}\big\|b_{t_{n,j}}(\tilde{X}^N_{t_{n,j}}) - b_{t_n}(X^N_{t_n})\big\|\Big)^2 \\
&\le \Big(\big((2L+1)^2+\tilde{L}\big)h\sum_{j=0}^{k-1} B_{t_{n,j}}\big\|X^N_{t_n}\big\| + \sum_{j=0}^{k-1} B_{t_{n,j}}\mathcal{I}_{t_{n,j}}\Big)^2 \\
&\le 2\big((2L+1)^2+\tilde{L}\big)^2h^2\Big(\sum_{j=0}^{k-1} B_{t_{n,j}}\Big)^2\big\|X^N_{t_n}\big\|^2 + 2\Big(\sum_{j=0}^{k-1} B_{t_{n,j}}\mathcal{I}_{t_{n,j}}\Big)^2.
\end{aligned} \tag{60}$$

Moreover, we apply the fact that $\|b_t(x)\| \le (2L+1)\|x\| + \|b_t(0)\|$ to arrive at

$$\big\|b_{t_n}(X^N_{t_n})\big\|^2 \le 2(2L+1)^2\big\|X^N_{t_n}\big\|^2 + 2\big\|b_{t_n}(0)\big\|^2. \tag{61}$$

Now, we estimate $\big\langle X^N_{t_n}, b_{t_n}(X^N_{t_n})\big\rangle$ for two cases: $nk \le N_h$ and $nk > N_h$. For $nk \le N_h$, Proposition A.4, in conjunction with the Young inequality with $\lambda \in (0, 1/4)$, directly yields

$$\begin{aligned}
2h\Big\langle X^N_{t_n}, b_{t_n}(X^N_{t_n})\Big\rangle &\le 2h\Big\langle X^N_{t_n}, b_{t_n}(X^N_{t_n}) - b_{t_n}(0)\Big\rangle + 2h\Big\langle X^N_{t_n}, b_{t_n}(0)\Big\rangle \\
&\le -2h(2C_{T-t_n} + 1 - \lambda/2)\big\|X^N_{t_n}\big\|^2 + \tfrac{h}{\lambda}\big\|b_{t_n}(0)\big\|^2.
\end{aligned} \tag{62}$$

In the case $nk > N_h$, we apply Proposition A.4 and the Young inequality to get

$$2h\Big\langle X^N_{t_n}, b_{t_n}(X^N_{t_n})\Big\rangle \le (4L+3)h\big\|X^N_{t_n}\big\|^2 + h\big\|b_{t_n}(0)\big\|^2. \tag{63}$$

Due to (52), $B^2_{t_{n,j}} = B_{t_{n,j}}$ and $\mathcal{I}_{t_{n,j}} \ge 0$, we have

$$6h^2\Big(\sum_{j=0}^{k-1} B_{t_{n,j}}\mathcal{I}_{t_{n,j}}\Big)^2 + \sum_{j=0}^{k-1} B_{t_{n,j}}\mathcal{I}^2_{t_{n,j}} \le (1+6h^2)\Big(\sum_{j=0}^{k-1} B_{t_{n,j}}\mathcal{I}_{t_{n,j}}\Big)^2 \le 7\Big(\sum_{j=0}^{k-1} B_{t_{n,j}}\mathcal{I}_{t_{n,j}}\Big)^2.$$

Equipped with the above estimates, and using the fact that $\sum_{j=0}^{k-1} B_{t_{n,j}} \le \big(\sum_{j=0}^{k-1} B_{t_{n,j}}\big)^2$, one can derive from (56) that

$$\begin{aligned}
&1 + \big\|X^N_{t_{n+1}}\big\|^2 \\
&\le \gamma_{t_n}\Big(1 + \big\|X^N_{t_n}\big\|^2\Big) + 8\Big(2\big((2L+1)^2+\tilde{L}\big)^2 + 1\Big)\Big(\sum_{j=0}^{k-1} B_{t_{n,j}}\Big)^2h^2\big\|X^N_{t_n}\big\|^2 \\
&\quad + \big(6 + C(\lambda)\big)h\big\|b_{t_n}(0)\big\|^2 + 7\Big(\sum_{j=0}^{k-1} B_{t_{n,j}}\mathcal{I}_{t_{n,j}}\Big)^2 + \tfrac{6h}{k}\Big\|\sum_{j=0}^{k-1} Z_{t_{n,j}}\Big\|^2 + 2\sqrt{\tfrac{2h}{k}}\Big\langle X^N_{t_n}, \sum_{j=0}^{k-1} Z_{t_{n,j}}\Big\rangle \\
&=: \gamma_{t_n}\Big(1 + \big\|X^N_{t_n}\big\|^2\Big) + 8\Big(2\big((2L+1)^2+\tilde{L}\big)^2 + 1\Big)\Big(\sum_{j=0}^{k-1} B_{t_{n,j}}\Big)^2h^2\big\|X^N_{t_n}\big\|^2 + \Xi_{t_{n+1}},
\end{aligned} \tag{64}$$

where

$$\begin{aligned}
\Xi_{t_{n+1}} := \big(6 + C(\lambda)\big)h\big\|b_{t_n}(0)\big\|^2 &+ 7\Big(\sum_{j=0}^{k-1} B_{t_{n,j}}\mathcal{I}_{t_{n,j}}\Big)^2 + \tfrac{6h}{k}\Big\|\sum_{j=0}^{k-1} Z_{t_{n,j}}\Big\|^2 \\
&+ 2\sqrt{\tfrac{2h}{k}}\Big\langle X^N_{t_n}, \sum_{j=0}^{k-1} Z_{t_{n,j}}\Big\rangle
\end{aligned} \tag{65}$$

and

$$\gamma_{t_n} := \begin{cases} 1 + \big(6(2L+1)^2h - 2(2C_{T-t_n}+1) + \lambda\big)h, & \text{for } nk \le N_h \\ 1 + \big(6(2L+1)^2 + 4L + 3\big)h, & \text{for } nk > N_h. \end{cases}$$

Furthermore, we have

$$
\begin{aligned}
&1+\big\|X^N_{t_{n+1}}\big\|^2\\
&=\gamma_{t_n}\Big(1+\big\|X^N_{t_n}\big\|^2\Big)+8\Big(2\big((2L+1)^2+\tilde{L}\big)^2+1\Big)\Big(\sum_{j=0}^{k-1}B_{t_{n,j}}\Big)^2h^2\big\|X^N_{t_n}\big\|^2+\Xi_{t_{n+1}}\\
&\le\gamma_{t_n}\Big(1+\big\|X^N_{t_n}\big\|^2\Big)\Big(1+8\Big(2\big((2L+1)^2+\tilde{L}\big)^2+1\Big)\Big(\sum_{j=0}^{k-1}B_{t_{n,j}}\Big)^2h^2/\gamma_{t_n}+\frac{\Xi_{t_{n+1}}}{\gamma_{t_n}\left(1+\|X^N_{t_n}\|^2\right)}\Big)\\
&=:\gamma_{t_n}\Big(1+\big\|X^N_{t_n}\big\|^2\Big)\big(1+\mathbb{I}_1+\mathbb{I}_2\big).
\end{aligned}
\tag{66}
$$

Taking square and then expectations of (66) to deduce

$$
\mathbb{E}\bigg[\Big(1+\big\|X^N_{t_{n+1}}\big\|^2\Big)^2\bigg]\le\gamma^2_{t_n}\mathbb{E}\bigg[\Big(1+\big\|X^N_{t_n}\big\|^2\Big)^2\big(1+\mathbb{I}_1+\mathbb{I}_2\big)^2\bigg].
\tag{67}
$$

With regard to the right term of above equality, we first notice that $X^N_{t_n}$ is $\mathcal{F}_{t_n}$-measurable. By further taking conditional expectation with respect to $\mathcal{F}_{t_n}$, one can see that

$$
\begin{aligned}
\mathbb{E}\bigg[\Big(1+\big\|X^N_{t_n}\big\|^2\Big)^2\big(1+\mathbb{I}_1+\mathbb{I}_2\big)^2\bigg]&=\mathbb{E}\bigg[\Big(1+\big\|X^N_{t_n}\big\|^2\Big)^2\big(1+\mathbb{I}_1+\mathbb{I}_2\big)^2\big|\mathcal{F}_{t_n}\bigg]\bigg]\\
&=\mathbb{E}\bigg[\Big(1+\big\|X^N_{t_n}\big\|^2\Big)^2\mathbb{E}\Big[\big(1+\mathbb{I}_1+\mathbb{I}_2\big)^2\big|\mathcal{F}_{t_n}\Big]\bigg].
\end{aligned}
\tag{68}
$$

Next, expanding the square and applying the Young inequality to the cross term $2\mathbb{I}_1\mathbb{I}_2$ show

$$
\mathbb{E}\Big[\big(1+\mathbb{I}_1+\mathbb{I}_2\big)^2\big|\mathcal{F}_{t_n}\Big]\le1+2\,\mathbb{E}\Big[\mathbb{I}_1^2\big|\mathcal{F}_{t_n}\Big]+2\,\mathbb{E}\Big[\mathbb{I}_2^2\big|\mathcal{F}_{t_n}\Big]+2\,\mathbb{E}\Big[\mathbb{I}_1\big|\mathcal{F}_{t_n}\Big]+2\,\mathbb{E}\Big[\mathbb{I}_2\big|\mathcal{F}_{t_n}\Big].
\tag{69}
$$

Since $\{B_{t_{n,j}}\}_{j\in[k-1]_0}$ are independent of $\mathcal{F}_{t_n}$, and given that $\mathbb{E}\big[\big(\sum_{j=0}^{k-1}B_{t_{n,j}}\big)^2\big]\le2$, $\mathbb{E}\big[\big(\sum_{j=0}^{k-1}B_{t_{n,j}}\big)^4\big]\le15$, one can derive that

$$
\begin{aligned}
\mathbb{E}\Big[\mathbb{I}_1\big|\mathcal{F}_{t_n}\Big]&\le16\Big(2\big((2L+1)^2+\tilde{L}\big)^2+1\Big)h^2/\gamma_{t_n},\\
\mathbb{E}\Big[\mathbb{I}_1^2\big|\mathcal{F}_{t_n}\Big]&\le960\Big(2\big((2L+1)^2+\tilde{L}\big)^2+1\Big)^2h^4/\gamma^2_{t_n}.
\end{aligned}
\tag{70}
$$

Next, we recall some properties of the standard Gaussians: for any $n\in[N-1]_0$, $j\in[k]_0$,

$$
\mathbb{E}\bigg[\Big\|\sum_{l=0}^{j-1}Z_{t_{n,l}}\Big\|^2\big|\mathcal{F}_{t_n}\bigg]=\sum_{l=0}^{j-1}\mathbb{E}\Big[\big\|Z_{t_{n,l}}\big\|^2\big|\mathcal{F}_{t_n}\Big]\le kd,\quad\mathbb{E}\bigg[\Big\|\sum_{l=0}^{j-1}Z_{t_{n,l}}\Big\|^4\big|\mathcal{F}_{t_n}\bigg]\le3k^2d^2,
\tag{71}
$$

and

$$
\mathbb{E}\Big[\sum_{l=0}^{j-1}Z_{t_{n,l}}\big|\mathcal{F}_{t_n}\Big]=0.
\tag{72}
$$

As a result,

$$
\begin{aligned}
\mathbb{E}\big[\mathcal{I}^2_{t_{n,j}}\big|\mathcal{F}_{t_n}\big]&\le\tfrac{6(2L+1)^2h}{k}\,\mathbb{E}\bigg[\Big\|\sum_{l=0}^{j-1}Z_{t_{n,l}}\Big\|^2\big|\mathcal{F}_{t_n}\bigg]+3(2L+1)h^2\big\|b_{t_n}(0)\big\|^2+3\tilde{L}^2h^2\\
&\le6(2L+1)^2dh+3(2L+1)^2h^2\big\|b_{t_n}(0)\big\|^2+3\tilde{L}^2h^2.
\end{aligned}
\tag{73}
$$

Using the same technique, we have

$$
\mathbb{E}\big[\mathcal{I}^4_{t_{n,j}}\big|\mathcal{F}_{t_n}\big]\le324(2L+1)^4d^2h^2+27(2L+1)^4h^4\big\|b_{t_n}(0)\big\|^4+27\tilde{L}^4h^4.
\tag{74}
$$

Similarly, since

$$
\Big(\sum_{j=0}^{k-1}B_{t_{n,j}}\mathcal{I}_{t_{n,j}}\Big)^2\le\Big(\sum_{j=0}^{k-1}B_{t_{n,j}}\Big)\sum_{j=0}^{k-1}B_{t_{n,j}}\mathcal{I}^2_{t_{n,j}},
$$

we have

$$\mathbb{E}\Big[\Big(\sum_{j=0}^{k-1} B_{t_{n,j}}\mathcal{I}_{t_{n,j}}\Big)^2 \mid \mathcal{F}_{t_n}\Big] \le C(L,\tilde{L})\Big(dh + h^2\|b_{t_n}(0)\|^2 + h^2\Big).$$

Therefore, we have

$$\begin{aligned}\mathbb{E}\big[\Xi_{t_{n+1}}\big|\mathcal{F}_{t_n}\big] \le & C\,\mathbb{E}\Big[\Big(\sum_{j=0}^{k-1} B_{t_{n,j}}\mathcal{I}_{t_{n,j}}\Big)^2\big|\mathcal{F}_{t_n}\Big] + \tfrac{6h}{k}\,\mathbb{E}\Big[\Big\|\sum_{j=0}^{k-1} Z_{t_{n,j}}\Big\|^2\big|\mathcal{F}_{t_n}\Big] \\ & + 2\sqrt{\tfrac{2h}{k}}\mathbb{E}\Big[\Big\langle X_{t_n}^N, \sum_{j=0}^{k-1} Z_{t_{n,j}}\Big\rangle\big|\mathcal{F}_{t_n}\Big] + \big(6+C(\lambda)\big)h\big\|b_{t_n}(0)\big\|^2 \\ & \le C(L,\tilde{L},\lambda)\big(d+\|b_{t_n}(0)\|^2\big)h.\end{aligned} \tag{75}$$

Based on the above estimate, we have

$$\begin{aligned}\mathbb{E}\Big[\mathbb{I}_2\big|\mathcal{F}_{t_n}\Big] &= \tfrac{1}{\gamma_{t_n}\left(1+\|X_{t_n}^N\|^2\right)}\mathbb{E}\Big[\Xi_{t_{n+1}}\big|\mathcal{F}_{t_n}\Big] \\ &\le \tfrac{C(L,\tilde{L},\lambda)(d+\|b_{t_n}(0)\|^2)h}{\gamma_{t_n}\left(1+\|X_{t_n}^N\|^2\right)}.\end{aligned} \tag{76}$$

For the estimate of $\mathbb{E}\big[\mathbb{I}_2^2\big|\mathcal{F}_{t_n}\big]$, using a fundamental inequality shows

$$\begin{aligned}\big(\Xi_{t_{n+1}}\big)^2 \le & 4\Big(\big(6+C(\lambda)\big)^2h^2\big\|b_{t_n}(0)\big\|^4 + 49\Big(\sum_{j=0}^{k-1} B_{t_{n,j}}\mathcal{I}_{t_{n,j}}\Big)^4 + \tfrac{36h^2}{k^2}\Big\|\sum_{j=0}^{k-1} Z_{t_{n,j}}\Big\|^4 \\ & + \tfrac{8h}{k}\Big\langle X_{t_n}^N, \sum_{j=0}^{k-1} Z_{t_{n,j}}\Big\rangle^2\Big) \\ \le & 4\Big(\big(6+C(\lambda)\big)^2h^2\big\|b_{t_n}(0)\big\|^4 + 49\Big(\sum_{j=0}^{k-1} B_{t_{n,j}}\mathcal{I}_{t_{n,j}}\Big)^4 + \tfrac{36h^2}{k^2}\Big\|\sum_{j=0}^{k-1} Z_{t_{n,j}}\Big\|^4 \\ & + \tfrac{8h}{k}\big\|X_{t_n}^N\big\|^2\Big\|\sum_{j=0}^{k-1} Z_{t_{n,j}}\Big\|^2\Big).\end{aligned} \tag{77}$$

Moreover, for the nonnegative Bernoulli variables $B_{t_{n,j}}$, we use

$$\Big(\sum_{j=0}^{k-1} B_{t_{n,j}}\mathcal{I}_{t_{n,j}}\Big)^4 \le \Big(\sum_{j=0}^{k-1} B_{t_{n,j}}\Big)^3\sum_{j=0}^{k-1} B_{t_{n,j}}\mathcal{I}_{t_{n,j}}^4.$$

Since the Bernoulli variables are independent of the Gaussian variables, the following estimate hold true,

$$\begin{aligned}\mathbb{E}\Big[\Big(\sum_{j=0}^{k-1} B_{t_{n,j}}\mathcal{I}_{t_{n,j}}\Big)^4 \mid \mathcal{F}_{t_n}\Big] &\le \sum_{j=0}^{k-1}\mathbb{E}\Big[B_{t_{n,j}}\Big(\sum_{\ell=0}^{k-1} B_{t_{n,\ell}}\Big)^3\Big]\,\mathbb{E}\big[\mathcal{I}_{t_{n,j}}^4 \mid \mathcal{F}_{t_n}\big] \\ &\le \mathbb{E}\Big[\Big(\sum_{\ell=0}^{k-1} B_{t_{n,\ell}}\Big)^4\Big]\sup_{0\le j\le k-1}\mathbb{E}\big[\mathcal{I}_{t_{n,j}}^4 \mid \mathcal{F}_{t_n}\big] \\ &\le C(L,\tilde{L})\Big(d^2h^2 + h^4\|b_{t_n}(0)\|^4 + h^4\Big).\end{aligned} \tag{78}$$

Moreover,

$$\mathbb{E}\Big[\Big\|\sum_{j=0}^{k-1} Z_{t_{n,j}}\Big\|^2 \mid \mathcal{F}_{t_n}\Big] = kd, \qquad \mathbb{E}\Big[\Big\|\sum_{j=0}^{k-1} Z_{t_{n,j}}\Big\|^4 \mid \mathcal{F}_{t_n}\Big] = k^2(d^2+2d) \le 3k^2d^2. \tag{79}$$

Keeping this in mind and using the Young inequality leads to

$$\mathbb{E}\Big[\big(\Xi_{t_{n+1}}\big)^2\big|\mathcal{F}_{t_n}\Big] \le 32\,dh\big\|X_{t_n}^N\big\|^2 + C(L,\tilde{L},\lambda)\big(d^2+\|b_{t_n}(0)\|^4\big)h^2. \tag{80}$$

Substituting (78) and (79) into (77) yields

$$\begin{aligned}\mathbb{E}\Big[\big(\Xi_{t_{n+1}}\big)^2\big|\mathcal{F}_{t_n}\Big] &\le 196\,\mathbb{E}\Big[\mathcal{I}_{t_n}^4\,\mathbb{E}\Big[\Big(\sum_{j=0}^{k-1}B_{t_{n,j}}\Big)^4\big|\mathcal{F}_{t_n}\Big]\Big]+\tfrac{144h^2}{k^2}\,\mathbb{E}\Big[\Big\|\sum_{j=0}^{k-1}Z_{t_{n,j}}\Big\|^4\big|\mathcal{F}_{t_n}\Big]\\ &\quad+\tfrac{32h}{k}\big\|X_{t_n}^N\big\|^2\mathbb{E}\Big[\Big\|\sum_{j=0}^{k-1}Z_{t_{n,j}}\Big\|^2\big|\mathcal{F}_{t_n}\Big]+4\big(6+C(\lambda)\big)^2h^2\big\|b_{t_n}(0)\big\|^4\\ &\le 32dh\big\|X_{t_n}^N\big\|^2+C(L,\tilde{L},\lambda)\big(d^2+\|b_{t_n}(0)\|^4\big)h^2.\end{aligned}\tag{81}$$

Therefore, one can obtain that

$$\begin{aligned}\mathbb{E}\Big[\mathbb{I}_2^2\big|\mathcal{F}_{t_n}\Big]&=\tfrac{1}{\gamma_{t_n}^2\left(1+\|X_{t_n}^N\|^2\right)^2}\,\mathbb{E}\Big[\big(\Xi_{t_{n+1}}\big)^2\big|\mathcal{F}_{t_n}\Big]\\ &\le\tfrac{32dh}{\gamma_{t_n}^2\left(1+\|X_{t_n}^N\|^2\right)}+\tfrac{C(L,\tilde{L},\lambda)\left(d^2+\|b_{t_n}(0)\|^4\right)h^2}{\gamma_{t_n}^2\left(1+\|X_{t_n}^N\|^2\right)^2}.\end{aligned}\tag{82}$$

Equipped with estimates, one can derive that

$$\begin{aligned}&\mathbb{E}\Big[\big(1+\mathbb{I}_1+\mathbb{I}_2\big)^2\big|\mathcal{F}_{t_n}\Big]\\ &\le 1+1952\Big(2\big((2L+1)^2+\tilde{L}\big)^2+1\Big)^2h^2/\gamma_{t_n}^2+\tfrac{C(L,\tilde{L},\lambda)\left(d+\|b_{t_n}(0)\|^2\right)h}{\gamma_{t_n}\left(1+\|X_{t_n}^N\|^2\right)}\\ &\quad+\tfrac{64dh}{\gamma_{t_n}^2\left(1+\|X_{t_n}^N\|^2\right)}+\tfrac{C(L,\tilde{L},\lambda)\left(d^2+\|b_{t_n}(0)\|^4\right)h^2}{\gamma_{t_n}^2\left(1+\|X_{t_n}^N\|^2\right)^2}.\end{aligned}\tag{83}$$

Thus, we get

$$\begin{aligned}\mathbb{E}\Big[\Big(1+\big\|X_{t_{n+1}}^N\big\|^2\Big)^2\Big]\le&\gamma_{t_n}^2\,\mathbb{E}\Big[\Big(1+\big\|X_{t_n}^N\big\|^2\Big)^2\Big]\\ &+1952\Big(2\big((2L+1)^2+\tilde{L}\big)^2+1\Big)^2h^2\,\mathbb{E}\Big[\Big(1+\big\|X_{t_n}^N\big\|^2\Big)^2\Big]\\ &+\gamma_{t_n}\mathbb{E}\Big[\Big(1+\big\|X_{t_n}^N\big\|^2\Big)C(L,\tilde{L},\lambda)\big(d+\|b_{t_n}(0)\|^2\big)h\Big]\\ &+64dh\,\mathbb{E}\Big[\Big(1+\big\|X_{t_n}^N\big\|^2\Big)\Big]+C(L,\tilde{L},\lambda)\big(d^2+\|b_{t_n}(0)\|^4\big)h^2.\end{aligned}\tag{84}$$

Based on the above estimate, using the Young inequality gives

$$\begin{aligned}&\gamma_{t_n}\mathbb{E}\Big[\Big(1+\big\|X_{t_n}^N\big\|^2\Big)C(L,\tilde{L},\lambda)\big(d+\|b_{t_n}(0)\|^2\big)h\Big]\\ &\le\lambda h\gamma_{t_n}^2/2\,\mathbb{E}\Big[\Big(1+\big\|X_{t_n}^N\big\|^2\Big)^2\Big]+C(L,\tilde{L},\lambda)\big(d^2+\|b_{t_n}(0)\|^4\big)h,\end{aligned}\tag{85}$$

and

$$64dh\,\mathbb{E}\Big[\Big(1+\big\|X_{t_n}^N\big\|^2\Big)\Big]\le\lambda h/2\,\mathbb{E}\Big[\Big(1+\big\|X_{t_n}^N\big\|^2\Big)^2\Big]+C(\lambda)d^2h.\tag{86}$$

Combining (85) with (84), one can obtain

$$\begin{aligned}
&\mathbb{E}\Big[\Big(1+\big\|X^N_{t_{n+1}}\big\|^2\Big)^2\Big]\\
\le&\Big(\gamma_{t_n}^2+1952\Big(2\big((2L+1)^2+\tilde{L}\big)^2+1\Big)^2h^2+\lambda h(\gamma_{t_n}^2+1)/2\Big)\mathbb{E}\Big[\Big(1+\big\|X^N_{t_n}\big\|^2\Big)^2\Big]\\
&+C(L,\tilde{L},\lambda)\big(d^2+\|b_{t_n}(0)\|^4\big)h\\
\le&\Big(\gamma_{t_n}^2+1952\Big(2\big((2L+1)^2+\tilde{L}\big)^2+1\Big)^2h^2+\lambda h(\gamma_{t_n}^2+1)/2\Big)\mathbb{E}\Big[\Big(1+\big\|X^N_{t_n}\big\|^2\Big)^2\Big]\\
&+C(L,\tilde{L},L',\lambda)\big(d^2+T^4\big)h,
\end{aligned}\tag{87}$$

where, the last inequality is an application of Lemma A.3, Assumption 3.7, and the fact $L'd^{1/2}:=\|\nabla V(0)\|$, for any $t\in[t_n,t_{n+1}]$,

$$\begin{aligned}
\big\|b_t(0)\big\|&=2\big\|\nabla\log\tilde{p}_{T-t}(0)\big\|=2\big\|\nabla\log p_{T-t}(0)\big\|\\
&\le C\bigg(\big\|\nabla\log p_{T-t}(0)-\nabla\log p_{T-t_n}(0)\big\|\\
&\quad+\sum_{j=n}^{N-1}\big\|\nabla\log p_{T-t_{j+1}}(0)-\nabla\log p_{T-t_j}(0)\big\|+\big\|\nabla\log p_0(0)\big\|\bigg)\\
&\le C\big(\tilde{L}T+L'd^{1/2}\big).
\end{aligned}\tag{88}$$

Case I. $nk\le N_h$. By noting that if h satisfies (52):

$$h<\tfrac{2(1-\lambda)}{C(L,\tilde{L})},$$

where

$$C(L,\tilde{L}):=6(2L+1)^2+1952\Big(2\big((2L+1)^2+\tilde{L}\big)^2+1\Big)^2,\tag{89}$$

then we can set $\rho_h^{(1)}:=C(L,\tilde{L})h/2+\lambda$, such that $\rho_h^{(1)}\in(\lambda,1)$. Then we have:

$$\gamma_{t_n}^2\le\gamma_{t_n}\le1$$

and

$$\gamma_{t_n}+1952\Big(2\big((2L+1)^2+\tilde{L}\big)^2+1\Big)^2h^2+\lambda h\le1.$$

One can then apply these inequalities iteratively, together with the facts that $X_0^N=\overleftarrow{X}_T$ and Lemma B.2 to obtain

$$\begin{aligned}
&\mathbb{E}\Big[\Big(1+\big\|X^N_{t_{n+1}}\big\|^2\Big)^2\Big]\\
&\le\Big(\gamma_{t_n}+1952\Big(2\big((2L+1)^2+\tilde{L}\big)^2+1\Big)^2h^2+\lambda h\Big)\mathbb{E}\Big[\Big(1+\big\|X^N_{t_n}\big\|^2\Big)^2\Big]\\
&\quad+C(L,\tilde{L},L',\lambda)\big(d^2+T^4\big)h\\
&:=\Gamma_{t_n}\mathbb{E}\Big[\Big(1+\big\|X^N_{t_n}\big\|^2\Big)^2\Big]+C(L,\tilde{L},L',\lambda)\big(d^2+T^4\big)h\\
&=\prod_{\ell=0}^{n}\Gamma_{t_\ell}\mathbb{E}\Big[\Big(1+\big\|X^N_0\big\|^2\Big)^2\Big]+C(L,\tilde{L},L',\lambda)\big(d^2+T^4\big)h\Big(\sum_{m=1}^{n}\prod_{\ell=m}^{n}\Gamma_{t_\ell}+1\Big)\\
&\le Cd^2+C(L,\tilde{L},L',\lambda)\big(d^2+T^4\big)T,
\end{aligned}\tag{90}$$

where

$$\Gamma_{t_n}:=\gamma_{t_n}+1952\Big(2\big((2L+1)^2+\tilde{L}\big)^2+1\Big)^2h^2+\lambda h,\tag{91}$$

and the last inequality follows from

$$h\Big(\sum_{m=1}^{n}\prod_{\ell=m}^{n}\big(\Gamma_{t_\ell}\big)+1\Big)\le h\big(\lfloor N_h/k\rfloor+1\big)\le \widehat{T}(\alpha,M,0)+h\le CT.$$

Case II. $nk > N_h$. By iteration, we acquire

$$\begin{aligned}
&\mathbb{E}\Big[\Big(1+\big\|X^N_{t_{n+1}}\big\|^2\Big)^2\Big]\\
&\le\Big(\gamma^2_{t_n}+1952\gamma_{t_n}\Big(2\big((2L+1)^2+\tilde{L}\big)^2+1\Big)^2h^2+\lambda h(\gamma^2_{t_n}+1)/2\Big)\mathbb{E}\Big[\Big(1+\big\|X^N_{t_n}\big\|^2\Big)^2\Big]\\
&\quad+C(L,\tilde{L},\lambda)\big(d^2+T^4\big)h\\
&:=\tilde{\Gamma}_{t_n}\mathbb{E}\Big[\Big(1+\big\|X^N_{t_n}\big\|^2\Big)^2\Big]+C(L,\tilde{L},L',\lambda)\big(d^2+T^4\big)h\\
&\le\prod_{\ell=0}^{n}\tilde{\Gamma}_{t_\ell}\mathbb{E}\Big[\Big(1+\big\|X^N_0\big\|^2\Big)^2\Big]+C(L,\tilde{L},L',\lambda)\big(d^2+T^4\big)h\Big(\sum_{m=1}^{n}\prod_{\ell=m}^{n}\tilde{\Gamma}_{t_\ell}+1\Big)\\
&\le\Big(1+\mathcal{M}_1d^2+C(L,\tilde{L},L',\lambda)\big(d^2+T^4\big)T\Big)\mathrm{e}^{\tilde{C}(L,\tilde{L})\eta(\alpha,M,\rho^{(1)}_h)},
\end{aligned}\tag{92}$$

where the last inequality follows by an argument similar to those used for (157) and (119), and $\tilde{C}(L,\tilde{L})$ is independent of d and h. The proof is completed. □

To obtain a fourth moment bound of the numerical process, which is required in the error analysis of PM, we arrive at estimates with a polynomial dependence on T (see the factor T^5 in Lemma B.3). Since T is eventually chosen logarithmically in d and ϵ^{-1}, this factor affects only the logarithmic factors in the iteration complexity. Based on the above preliminaries, we next present some regularity estimates of the process $\{X^N_{t_{n,i}}\}_{n\in[N-1]_0,i\in[k]_0}$ defined by (21).

Lemma B.4. *Let Assumptions 3.2, 3.3 and 3.7 hold. Let* $\big\{\tilde{X}_{t_{n,i}}\big\}$ *and* $\big\{X^N_{t_{n,i}}\big\}$*, for any* $n\in[N-1]_0, i\in[k]_0$ *be defined as* (21)*. For any* $\lambda\in(0,1/4)$*, let the step-size satisfy*

$$h<\tfrac{2(1-\lambda)}{C(L,\tilde{L})},$$

where $C(L,\tilde{L})$ *is given by* (89)*. Then, there exists a constant* $C(\mathcal{M}_2)$*, satisfying*

$$\mathbb{E}\Big[\big\|b_{t_{n,i}}(\tilde{X}^N_{t_{n,i}})-b_{t_n}(X^N_{t_n})\big\|^4\Big]\le C(\mathcal{M}_2)T^5d^2h^2,\tag{93}$$

and

$$\mathbb{E}\Big[\big\|\tilde{X}^N_{t_{n,i}}-X^N_{t_{n,i}}\big\|^2\Big]\le C(\mathcal{M}_2)T^{5/2}dh^3,\tag{94}$$

where $\mathcal{M}_2$ *comes from Lemma B.3.*

Proof. First, one can do the following decomposition:

$$\begin{aligned}
\mathbb{E}\Big[\big\|b_{t_{n,i}}(\tilde{X}^N_{t_{n,i}})-b_{t_n}(X^N_{t_n})\big\|^4\Big]&\le 8\,\mathbb{E}\Big[\big\|b_{t_{n,i}}(\tilde{X}^N_{t_{n,i}})-b_{t_{n,i}}(X^N_{t_n})\big\|^4\Big]\\
&\quad+8\,\mathbb{E}\Big[\big\|b_{t_{n,i}}(X^N_{t_n})-b_{t_n}(X^N_{t_n})\big\|^4\Big].
\end{aligned}\tag{95}$$

Regarding the first term, we use the Lipschitz condition of $b_t(\cdot)$, Lemma B.3 and the estimate (88), to obtain

$$\begin{aligned}
\mathbb{E}\Big[\big\|b_{t_{n,i}}(\tilde{X}^N_{t_{n,i}})-b_{t_{n,i}}(X^N_{t_n})\big\|^4\Big]&\le(2L+1)^4\,\mathbb{E}\Big[\big\|\tilde{X}^N_{t_{n,i}}-X^N_{t_n}\big\|^4\Big]\\
&=(2L+1)^4\,\mathbb{E}\Big[\big\|\tfrac{ih}{k}b_{t_n}(X^N_{t_n})+\sum_{j=0}^{i-1}\sqrt{\tfrac{2h}{k}}Z_{t_{n,j}}\big\|^4\Big]\\
&\le C(\mathcal{M}_2)T^5d^2h^2.
\end{aligned}\tag{96}$$

With regard to the second term, applying Assumption 3.7 and Lemma B.3, we have

$$\mathbb{E}\Big[\big\|b_{t_{n,i}}(X^N_{t_n}) - b_{t_n}(X^N_{t_n})\big\|^4\Big] \le C(\mathcal{M}_2)T^5 d^2 h^4. \tag{97}$$

The first desired assertion (93) is thus validated. Concerning the second assertion, we recall (21) to get, for any $n \in [N-1]_0$, $i \in [k]_0$,

$$\tilde{X}^N_{t_{n,i}} - X^N_{t_{n,i}} = -h\sum_{j=0}^{i-1} B_{t_{n,j}}\Big(b_{t_{n,j}}(\tilde{X}^N_{t_{n,j}}) - b_{t_n}(X^N_{t_n})\Big). \tag{98}$$

In view of the above equality, using (93) and the identities $\mathbb{E}[B_{t_{n,r}}B_{t_{n,\ell}}] = 1/k^2$ for $r \neq \ell$ and $\mathbb{E}[B^2_{t_{n,j}}] = 1/k$, we obtain

$$\begin{aligned}
\mathbb{E}\Big[\big\|\tilde{X}^N_{t_{n,i}} - X^N_{t_{n,i}}\big\|^2\Big] &= h^2\,\mathbb{E}\Big[\Big\|\sum_{j=0}^{i-1} B_{t_{n,j}}\Big(b_{t_{n,j}}(\tilde{X}^N_{t_{n,j}}) - b_{t_n}(X^N_{t_n})\Big)\Big\|^2\Big] \\
&\le \underbrace{\tfrac{2h^2}{k^2}\sum_{0\le r<\ell\le i-1}\mathbb{E}\Big[\big\langle b_{t_{n,r}}(\tilde{X}^N_{t_{n,r}}) - b_{t_n}(X^N_{t_n}), b_{t_{n,\ell}}(\tilde{X}^N_{t_{n,\ell}}) - b_{t_n}(X^N_{t_n})\big\rangle\Big]}_{:=J_1} \\
&\quad + \underbrace{\tfrac{h^2}{k}\sum_{j=0}^{i-1}\mathbb{E}\Big[\big\|b_{t_{n,j}}(\tilde{X}^N_{t_{n,j}}) - b_{t_n}(X^N_{t_n})\big\|^2\Big]}_{=:J_2}.
\end{aligned} \tag{99}$$

In the following we deal with the above two items separately. Using the Cauchy-Schwarz inequality and (93), one can derive that

$$\begin{aligned}
J_1 &\le \tfrac{2h^2}{k^2}\sum_{0\le r<\ell\le i-1}\mathbb{E}\Big[\big\|b_{t_{n,r}}(\tilde{X}^N_{t_{n,r}}) - b_{t_n}(X^N_{t_n})\big\|\cdot\big\|b_{t_{n,\ell}}(\tilde{X}^N_{t_{n,\ell}}) - b_{t_n}(X^N_{t_n})\big\|\Big] \\
&\le h^2 \sup_{0\le r\le i-1}\mathbb{E}\Big[\big\|b_{t_{n,r}}(\tilde{X}^N_{t_{n,r}}) - b_{t_n}(X^N_{t_n})\big\|^2\Big] \\
&\le CT^{5/2}dh^3.
\end{aligned} \tag{100}$$

Regarding J_2, we use (93) to obtain

$$J_2 \le C(\mathcal{M}_2)T^{5/2}dh^3. \tag{101}$$

The proof is thus completed. □

Intuition behind the Poisson midpoint method.

Algorithmically, the intuition behind PM is to reduce the bias of a midpoint correction while avoiding the cost of evaluating the score at all k intermediate points. The key idea is to activate each interior correction $\Delta_{n,i} = b_{t_{n,i}}(x_i) - b_{t_n}(y)$ independently with probability $1/k$. Writing $\mathbb{E}_B$ for expectation with respect to the Bernoulli variables alone, and using $\mathbb{E}_B[B_{t_{n,i}}] = 1/k$, we obtain, for any $x_i, y \in \mathbb{R}^d$,

$$\mathbb{E}_B\Big[h\,b_{t_n}(y) + h\sum_{i=0}^{k-1} B_{t_{n,i}}\Delta_{n,i}\Big] = \tfrac{h}{k}\sum_{i=0}^{k-1} b_{t_{n,i}}(x_i).$$

The PM update therefore has the same conditional mean as the full k-point midpoint update, while $\mathbb{E}_B[\sum_i B_{t_{n,i}}] = 1$ means only one interior score is evaluated on average. Including the left-endpoint score, PM uses two expected score evaluations per macro-step, and hence a constant-order score-evaluation cost comparable to that of EM.

Analytically, the randomized correction admits the decomposition

$$hB_{t_{n,i}}\Delta_{n,i} = \tfrac{h}{k}\Delta_{n,i} + R_{n,i}, \qquad R_{n,i} := h\left(B_{t_n,i} - \tfrac{1}{k}\right)\Delta_{n,i},$$

where

$$\mathbb{E}_B[R_{n,i}] = 0.$$

Consequently, each refined PM sub-step can be viewed as an EM step plus a centered random perturbation. The key observation is that this perturbation is one-dimensional since the only additional randomness comes from the scalar Bernoulli variable. A coupling adapted to this structure makes its Wasserstein contribution enter at a higher order than under a direct synchronous coupling. This is the key reason that the randomized thinning preserves the midpoint acceleration.

Rank-one perturbation and optimal coupling. Each refined PM sub-step produces, conditionally on $\mathcal{F}_{t_{n,j}}$, a Gaussian increment together with a mean-zero, rank-one perturbation

$$V(x, y, B_{t_{n,i}}) := \sqrt{\tfrac{hk}{2}}\Big(B_{t_{n,i}} - \tfrac{1}{k}\Big)\Big(b_{t_{n,j}}(x) - b_{t_n}(y)\Big),$$

supported on the line spanned by the drift discrepancy $b_{t_{n,j}}(x) - b_{t_n}(y)$. The following Wasserstein bound exploits this rank-one structure [46, Lemma 1] or [23, Lemma 7].

Lemma B.5 (Wasserstein bound for one-dimensional bounded perturbations)**.** *Let V be a random vector in $\mathbb{R}^d$ such that*

(i) $\|V\| \le \beta$ almost surely; $\mathbb{E}[V] = 0$; $\mathbb{E}[VV^\top] = \Sigma$;

(ii) there exists a deterministic unit vector $u \in \mathbb{S}^{d-1}$ with $V \in \operatorname{span}(u)$ almost surely.

Let $Z \sim \gamma^d$ be independent of V, and set $\nu := \operatorname{Tr}(\Sigma) = \mathbb{E}\big[\|V\|^2\big]$. Then

$$\mathcal{W}_2^2\big(\mathcal{L}(Z+V), \gamma^d\big) \le \tfrac{11}{2}\nu^2 + \mathbf{1}_{\{5\beta^2>1\}} \cdot 2\nu.$$

A direct synchronous coupling of $Z + V$ with Z would only give

$$\mathcal{W}_2^2(\mathcal{L}(Z+V), \gamma^d) \le \mathbb{E}\big[\|V\|^2\big] = \nu.$$

Lemma B.5 sharpens the linear dependence on ν to a quadratic one ν^2, which is the main source of the improved PM discretization bound. Since the error analysis is based on mean-square estimates, we need a coupling that realizes this Wasserstein distance as an L^2-error. The following lemma provides such a coupling via Brenier's optimal transport theorem.

Lemma B.6. *Fix $n \in [N]_0$. Let V_n satisfy the assumptions of Lemma B.5 with parameters β_n, Σ_n, and $\nu_n := \operatorname{Tr}(\Sigma_n)$, and let $Z_n \sim \gamma^d$ be independent of V_n. Then there exists a random vector $\xi_n \sim \gamma^d$ such that*

$$\mathbb{E}\big[\|Z_n + V_n - \xi_n\|^2\big] = \mathcal{W}_2^2\big(\mathcal{L}(Z_n + V_n), \gamma^d\big) \le \tfrac{11}{2}\nu_n^2 + \mathbf{1}_{\{5\beta_n^2>1\}} \cdot 2\nu_n. \tag{102}$$

Proof. Since V_n is independent of $Z_n \sim \gamma^d$, the convolution $\mathcal{L}(Z_n + V_n)$ admits a density with respect to Lebesgue measure. Brenier's theorem [16, Theorem 2.5.10] therefore yields an optimal transport map $\mathcal{T}_n = \nabla\phi_n$ from $\mathcal{L}(Z_n + V_n)$ to γ^d with $\phi_n : \mathbb{R}^d \to \mathbb{R}$ convex. Setting $\xi_n := \mathcal{T}_n(Z_n + V_n)$ gives $\xi_n \sim \gamma^d$ and

$$\mathbb{E}\big[\|Z_n + V_n - \xi_n\|^2\big] = \mathcal{W}_2^2\big(\mathcal{L}(Z_n + V_n), \gamma^d\big),$$

to which Lemma B.5 applies and gives (102). □

Embedding into a Wiener process. Since the numerical variables $\{Z_n\}_{n\in[N]_0}$ are independent across refined sub-steps and the optimal transport maps are measurable, applying Lemma B.6 successively over the refined grid yields an i.i.d. family $\{\xi_{t_{n,j}}\}_{n,j}$ with law γ^d. By the standard isonormal Gaussian process construction [37, Section 1.1], these variables can be embedded into a Wiener process W such that

$$\Delta W_{t_{n,j}} = \sqrt{\tfrac{h}{k}}\,\xi_{t_{n,j}},$$

which we take as the driving process of the reverse SDE (5) throughout the proof of Theorem 3.9.

Up to this stage, we have established sufficient machinery to obtain the discretization error estimate between the PM auxiliary process (21) and the backward process (5) as below.

Proposition B.7 (Discretization error of PM method)**.** *Suppose that Assumptions 3.2, 3.3 and 3.7 hold. For any $n \in [N-1]_0$, $j \in [k]_0$, consider the process $\{X^N_{t_{n,j}}\}$ generated by* (21) *using the uniform step-size. For any $\lambda \in (0, 1/4)$, let the step-size satisfy*

$$h < \tfrac{2(1-\lambda)}{3(2L+1)^2} \wedge \tfrac{2(1-\lambda)}{C(L,\tilde{L})}, \tag{103}$$

where $C(L, \tilde{L})$ is given by (89)*. Then there exists a constant C, independent of d and h, such that,*

$$\mathcal{W}_2\big(\mu, \mathcal{L}(X^N_T)\big) \le C\mathrm{e}^{C\hat{\eta}(\alpha,M,\rho_h)}T^3\big(\tfrac{\sqrt{d}h}{k} + kdh^2 + dh^2 + \sqrt{d}h^{3/2}\big). \tag{104}$$

with $\hat{\eta}(\alpha, M, \rho_h) := \eta(\alpha, M, C(L, \tilde{L})h/2 + \lambda) + \eta\big(\alpha, M, 3(2L+1)^2h/2 + \lambda\big)$, and $\eta(\alpha, M, \rho_h)$ as defined in (30)*, where L comes from Lemma A.3.*

Proof. By Proposition B.1, it remains to estimate the four residual terms in the one-step recursion.

Step I: estimate of $R_{t_{n,i}}$. Fix $0 \le \epsilon_2 \le h/k$, define T to denote $T - \epsilon_2$ and N to denote $(T-\epsilon_2)/h$. This transformation allows us to apply Lemma A.3 and [13, Proposition 2]. Using the Itô formula [13, Proposition 2] yields

$$\begin{aligned}
&\mathrm{d}b_t(\overleftarrow{X}_t)\\
&= -\mathrm{d}\overleftarrow{X}_t + \mathrm{d}\big(2\nabla \log \tilde{p}_{T-t}(\overleftarrow{X}_t)\big)\\
&= \Big\{\overleftarrow{X}_t - 2\nabla \log \tilde{p}_{T-t}(\overleftarrow{X}_t) + 2\nabla \log \tilde{p}_{T-t}(\overleftarrow{X}_t)\Big\}\,\mathrm{d}t + \sqrt{2}\big(-I_d + 2\nabla^2 \log \tilde{p}_{T-t}(\overleftarrow{X}_t)\big)\,\mathrm{d}W_t\\
&= \overleftarrow{X}_t\,\mathrm{d}t + \sqrt{2}\big(-I_d + 2\nabla^2 \log \tilde{p}_{T-t}(\overleftarrow{X}_t)\big)\,\mathrm{d}W_t.
\end{aligned} \tag{105}$$

Therefore, the integral form can be written as, for any $n \in [N-1]_0$, $i \in [k-1]_0$,

$$b_s(\overleftarrow{X}_s) - b_{t_{n,i}}(\overleftarrow{X}_{t_{n,i}}) = \int_{t_{n,i}}^{s} \overleftarrow{X}_r\,\mathrm{d}r + \sqrt{2}\int_{t_{n,i}}^{s} \big(-I_d + 2\nabla^2 \log \tilde{p}_{T-r}(\overleftarrow{X}_r)\big)\,\mathrm{d}W_r. \tag{106}$$

Substituting the above equation into $\mathbb{E}\big[\|R_{t_{n,i}}\|^2\big]$ and using the Hölder inequality, the Itô isometry [36, Lemma 5.4] and Lemma B.2 gives

$$\begin{aligned}
&\mathbb{E}\Big[\big\|R_{t_{n,i}}\big\|^2\Big]\\
&= \mathbb{E}\bigg[\Big\|\int_{t_{n,i}}^{t_{n,i+1}} \big[b_s(\overleftarrow{X}_s) - b_{t_{n,i}}(\overleftarrow{X}_{t_{n,i}})\big]\,\mathrm{d}s\Big\|^2\bigg]\\
&= \mathbb{E}\bigg[\Big\|\int_{t_{n,i}}^{t_{n,i+1}}\int_{t_{n,i}}^{s} \overleftarrow{X}_r\,\mathrm{d}r + \sqrt{2}\int_{t_{n,i}}^{t_{n,i+1}}\int_{t_{n,i}}^{s} \big(-I_d + 2\nabla^2 \log \tilde{p}_{T-r}(\overleftarrow{X}_r)\big)\mathrm{d}W_r\,\mathrm{d}s\Big\|^2\bigg]\\
&\le \tfrac{2h^2}{k^2}\mathbb{E}\bigg[\int_{t_{n,i}}^{t_{n,i+1}}\int_{t_{n,i}}^{s} \big\|\overleftarrow{X}_r\big\|^2\mathrm{d}r\bigg] + \tfrac{4h}{k}\,\mathbb{E}\bigg[\int_{t_{n,i}}^{t_{n,i+1}}\int_{t_{n,i}}^{s} \big\|-I_d + 2\nabla^2 \log \tilde{p}_{T-r}(\overleftarrow{X}_r)\big)\big\|_{\mathrm{F}}^2\,\mathrm{d}r\,\mathrm{d}s\bigg]\\
&\le 2\big(\mathcal{M}_1 + 4(1+4L^2)\big)\,\tfrac{dh^3}{k^3},
\end{aligned} \tag{107}$$

where the last inequality follows from the direct result of Lemma A.3 that

$$\big\|\nabla^2 \log \tilde{p}_{T-t}(\overleftarrow{X}_t)\big\|_{\mathrm{F}} \le \sqrt{d}\,\big\|\nabla^2 \log \tilde{p}_{T-t}(\overleftarrow{X}_t)\big\| \le \sqrt{d}\,L, \quad \forall\, t \in [0, T). \tag{108}$$

Step II: estimate of $\mathbb{E}\big[\big\|\mathbb{E}\big[R_{t_{n,i}} \mid \mathcal{F}_{t_{n,i}}\big]\big\|^2\big]$. Noting that, for any $n \in [N-1]_0$, $i \in [k-1]_0$,

$$\mathbb{E}\bigg[\int_{t_{n,i}}^{s} \big(-I_d + 2\nabla^2 \log \tilde{p}_{T-r}(\overleftarrow{X}_r)\big)\mathrm{d}W_r\Big|\mathcal{F}_{t_{n,i}}\bigg] = 0, \tag{109}$$

one can utilize (106), the Jensen inequality, Lemma B.2 and the properties of the conditional expectation to show that,

$$\begin{aligned}
\mathbb{E}\Big[\big\|\mathbb{E}\big[R_{t_{n,i}} \mid \mathcal{F}_{t_{n,i}}\big]\big\|^2\Big] &= \mathbb{E}\Big[\Big\|\mathbb{E}\Big[\int_{t_{n,i}}^{t_{n,i+1}} \big[b_s(\overleftarrow{X}_s) - b_{t_{n,i}}(\overleftarrow{X}_{t_{n,i}})\big]\,\mathrm{d}s\Big|\mathcal{F}_{t_{n,i}}\Big]\Big\|^2\Big] \\
&= \mathbb{E}\Bigg[\Big\|\mathbb{E}\Big[\int_{t_{n,i}}^{t_{n,i+1}}\int_{t_{n,i}}^{s} \overleftarrow{X}_r\,\mathrm{d}r\,\mathrm{d}s\Big|\mathcal{F}_{t_{n,i}}\Big] \\
&\quad + \sqrt{2}\,\mathbb{E}\Big[\int_{t_{n,i}}^{t_{n,i+1}}\int_{t_{n,i}}^{s}\big(-I_d + 2\nabla^2\log\tilde{p}_{T-r}(\overleftarrow{X}_r)\big)\,\mathrm{d}W_r\,\mathrm{d}s\Big|\mathcal{F}_{t_{n,i}}\Big]\Big\|^2\Bigg] \\
&\le \mathbb{E}\Big[\mathbb{E}\Big[\Big\|\int_{t_{n,i}}^{t_{n,i+1}}\int_{t_{n,i}}^{s}\overleftarrow{X}_r\,\mathrm{d}r\,\mathrm{d}s\Big\|^2\Big|\mathcal{F}_{t_{n,i}}\Big]\Big] \\
&\le \tfrac{h^2}{k^2}\,\mathbb{E}\Bigg[\int_{t_{n,i}}^{t_{n,i+1}}\int_{t_{n,i}}^{s}\big\|\overleftarrow{X}_r\big\|^2\,\mathrm{d}r\,\mathrm{d}s\Bigg] \\
&\le \tfrac{\mathcal{M}_1 d h^4}{k^4}.
\end{aligned} \tag{110}$$

Step III: estimate of $\mathbb{E}\big[\|M_{t_{n,i}}\|^2\big]$. One can do the following decomposition:

$$\begin{aligned}
\mathbb{E}\Big[\big\|M_{t_{n,i}}\big\|^2\Big] \le\ & \underbrace{\tfrac{4h}{k}\,\mathbb{E}\Big[\Big\|Z_{t_{n,i}} - \xi_{t_{n,i}} + \sqrt{\tfrac{hk}{2}}\big(B_{t_{n,i}} - 1/k\big)\Big(b_{t_{n,i}}(\tilde{X}^N_{t_{n,i}}) - b_{t_n}(X^N_{t_n})\Big)\Big\|^2\Big]}_{=:\mathbb{J}_1} \\
&+ \underbrace{\tfrac{2h^2}{k^2}\mathbb{E}\Big[\big\|b_{t_{n,i}}(\tilde{X}^N_{t_{n,i}}) - b_{t_{n,i}}(X^N_{t_{n,i}})\big\|^2\Big]}_{=:\mathbb{J}_2}.
\end{aligned} \tag{111}$$

The above two items will be treated one by one. We estimate $\mathbb{J}_1$ first. By abuse of notation, we denote, for any $n \in [N-1]_0$, $i \in [k-1]_0$,

$$V(x, y, B_{t_{n,i}}) := \sqrt{\tfrac{hk}{2}}(B_{t_{n,i}} - 1/k)\Big(b_{t_{n,i}}(x) - b_{t_n}(y)\Big).$$

Conditionally on $\mathcal{F}_{t_{n,i}}$, the perturbation $V(x, y, B_{t_{n,i}})$ is centered and supported on the line spanned by $b_{t_{n,i}}(x) - b_{t_n}(y)$. Moreover, the corresponding quantities in Lemma B.5 satisfy

$$\nu_{t_{n,i}}(x,y) := \mathbb{E}\Big[\tfrac{hk}{2}(B_{t_{n,i}} - 1/k)^2\,\big\|b_{t_{n,i}}(x) - b_{t_n}(y)\big\|^2\Big], \qquad \beta^2_{t_{n,i}}(x,y) := \tfrac{hk}{2}\big\|b_{t_{n,i}}(x) - b_{t_n}(y)\big\|^2.$$

Using Lemma B.6, the Cauchy–Schwarz inequality, and Lemma B.4, we obtain

$$\begin{aligned}
\mathbb{J}_1 &= \mathbb{E}\Big[\mathbb{E}\Big[\big\|Z_{t_{n,i}} - \xi_{t_{n,i}} + V(x, y, B_{t_{n,i}})\big\|^2\big|\tilde{X}^N_{t_{n,i}} = x, X^N_{t_n} = y\Big]\Big] \\
&\le \mathbb{E}\Big[\mathbf{1}_{\{5\beta^2_{t_{n,i}} > 1\}}\cdot 2\nu_{t_{n,i}} + \tfrac{11}{2}\nu^2_{t_{n,i}}\Big] \\
&\le h\sqrt{\mathbb{P}(5\beta^2_{t_{n,i}} > 1)}\Big(\mathbb{E}\big[\big\|b_{t_{n,i}}(\tilde{X}^N_{t_{n,i}}) - b_{t_n}(X^N_{t_n})\big\|^4\big]\Big)^{1/2} \\
&\quad + \tfrac{11h^2}{8}\mathbb{E}\Big[\big\|b_{t_{n,i}}(\tilde{X}^N_{t_{n,i}}) - b_{t_n}(X^N_{t_n})\big\|^4\Big] \\
&\le C(\mathcal{M}_2)T^5\big(kd^2h^4 + d^2h^4\big),
\end{aligned} \tag{112}$$

where the last step follows from Markov's inequality:

$$\mathbb{P}(5\beta^2_{t_{n,i}} > 1) \le \mathbb{E}\big[(5\beta^2_{t_{n,i}})^2\big] \le C(\mathcal{M}_2)T^5k^2d^2h^4.$$

For $\mathbb{J}_2$, the Lipschitz bound on b_t and Lemma B.4 imply

$$\mathbb{J}_2 \le \tfrac{(2L+1)^2h^2}{k^2}\mathbb{E}\Big[\big\|\tilde{X}^N_{t_{n,i}} - X^N_{t_{n,i}}\big\|^2\Big] \le \tfrac{C(\mathcal{M}_2)T^{5/2}dh^5}{k^2} \le \tfrac{C(\mathcal{M}_2)T^{5/2}dh^4}{k}. \tag{113}$$

Combining the estimates of $\mathbb{J}_1$ and $\mathbb{J}_2$ yields

$$\mathbb{E}\Big[\big\|M_{t_{n,i}}\big\|^2\Big] \le C(\mathcal{M}_2)T^5\big(d^2h^5 + \tfrac{d^2h^5}{k} + \tfrac{dh^4}{k}\big). \tag{114}$$

With regard to the last term of (49), we use Lemma B.4 to directly deduce

$$\tfrac{(2L+1)^2h}{\lambda k}\mathbb{E}\Big[\big\|X^N_{t_{n,i}} - \tilde{X}^N_{t_{n,i}}\big\|^2\Big] \le \tfrac{C(\mathcal{M}_2)T^5dh^4}{\lambda k}. \tag{115}$$

Step IV: global recursion. By virtue of Proposition B.1 and the above estimates, iterating and invoking the fact that $\overleftarrow{X}_0 = X_0^N$, we obtain

$$\begin{aligned}
&\mathbb{E}\big[\|\varepsilon_T\|^2\big] \\
&\le \gamma_{(Nk-1)h/k}\,\mathbb{E}\big[\|\varepsilon_{(Nk-1)h/k}\|^2\big] + C(\mathcal{M}_2)T^5\Big(\tfrac{dh^3}{k^3} + d^2h^5 + \tfrac{d^2h^5}{k} + \tfrac{dh^4}{k}\Big) \\
&\le \prod_{\ell=0}^{Nk-1}\gamma_{\ell h/k}\,\mathbb{E}\big[\|\varepsilon_0\|^2\big] + C(\mathcal{M}_2)T^5\Big(\tfrac{dh^3}{k^3} + d^2h^5 + \tfrac{d^2h^5}{k} + \tfrac{dh^4}{k}\Big)\Big(\sum_{w=1}^{Nk-1}\prod_{\ell=w}^{Nk-1}\gamma_{\ell h/k} + 1\Big).
\end{aligned} \tag{116}$$

Now we are left with the estimation of $\big(\sum_{w=1}^{Nk-1}\prod_{\ell=w}^{Nk-1}\gamma_{\ell h/k} + 1\big)$. Note that if h satisfies (103), then we can take $\rho_h^{(2)} := 3(2L+1)^2h/2 + \lambda$, such that $\rho_h^{(2)} \in (\lambda, 1)$. Otherwise said, we have

$$\tfrac{h}{k} \le h = \frac{2\rho_h^{(2)} - 2\lambda}{3(2L+1)^2} \le \frac{2(2C_{T-t_{n,i}} + 1) - 2\lambda}{3(2L+1)^2}, \tag{117}$$

i.e., $\gamma_{nh+ih/k} \le 1$ for $nk + i \le N_h$. Whereas, for $nk + i > N_h$, Lemma A.3 yields

$$\gamma_{t_{n,i}} = 1 + \big(\tfrac{3(2L+1)^2h}{k} + 2(2L+1) + 2\lambda\big)\tfrac{h}{k} \le 1 + \big(5(2L+1)^2 + 2\lambda\big)\tfrac{h}{k}. \tag{118}$$

Combining the previous remarks, we have

$$\begin{aligned}
&\sum_{w=1}^{Nk-1}\prod_{\ell=w}^{Nk-1}\gamma_{\ell h/k} + 1 \\
&= \sum_{w=1}^{N_h-1}\prod_{\ell=w}^{N_h-1}\gamma_{\ell h/k} \times \prod_{\ell=N_h}^{Nk-1}\gamma_{\ell h/k} + \sum_{w=N_h}^{Nk-1}\prod_{\ell=w}^{Nk-1}\gamma_{\ell h/k} + 1 \\
&\le (N_h - 1)\prod_{\ell=N_h}^{Nk-1}\gamma_{\ell h/k} + \sum_{w=N_h}^{Nk-1}\prod_{\ell=w}^{Nk-1}\gamma_{\ell h/k} + 1 \\
&\le (N_h - 1)\Big(1 + \big(5(2L+1)^2 + 2\lambda\big)\tfrac{h}{k}\Big)^{Nk-N_h} + \sum_{w=N_h}^{Nk-1}\Big(1 + \big(5(2L+1)^2 + 2\lambda\big)\tfrac{h}{k}\Big)^{Nk-w} + 1 \\
&\le (N_h - 1)\Big(1 + \big(5(2L+1)^2 + 2\lambda\big)\tfrac{h}{k}\Big)^{Nk-N_h} + \sum_{w=1}^{Nk-N_h}\Big(1 + \big(5(2L+1)^2 + 2\lambda\big)\tfrac{h}{k}\Big)^{w} + 1 \\
&= (N_h - 1)\Big(1 + \big(5(2L+1)^2 + 2\lambda\big)\tfrac{h}{k}\Big)^{Nk-N_h} \\
&\quad + \frac{\big(1+(5(2L+1)^2+2\lambda)\frac{h}{k}\big)\big(1-\big(1+(5(2L+1)^2+2\lambda)\frac{h}{k}\big)^{Nk-N_h}\big)}{1-\big(1+(5(2L+1)^2+2\lambda)\frac{h}{k}\big)} + 1 \\
&\le \Big(N_h + \tfrac{k}{(5(2L+1)^2+2\lambda)h}\Big) \times \Big(1 + \big(5(2L+1)^2 + 2\lambda\big)\tfrac{h}{k}\Big)^{Nk-N_h+1} \\
&\le \Big(\tfrac{(T-\eta(\alpha,M,\rho_h))k}{h} + \tfrac{k}{(5(2L+1)^2+2\lambda)h}\Big)\,\mathrm{e}^{(5(2L+1)^2+2\lambda)\eta(\alpha,M,\rho_h)}.
\end{aligned} \tag{119}$$

Putting the above estimate into (116), we get

$$\begin{aligned}
\mathcal{W}_2^2\big(\mu, \mathcal{L}(X_T^N)\big) &\le \mathbb{E}\Big[\big\|\overleftarrow{X}_{t_N} - X_{t_N}^N\big\|^2\Big] \\
&\le C(\mathcal{M}_2)T^5\Big(\tfrac{dh^3}{k^3} + d^2h^5 + \tfrac{d^2h^5}{k} + \tfrac{dh^4}{k}\Big) \\
&\quad \cdot \Big(\tfrac{(T-\eta(\alpha,M,\rho_h^{(2)}))k}{h} + \tfrac{k}{(5(2L+1)^2+2\lambda)h}\Big)\,\mathrm{e}^{(5(2L+1)^2+2\lambda)\eta(\alpha,M,\rho_h^{(2)})} \\
&\le C(\mathcal{M}_2)T^6\big(\tfrac{dh^2}{k^2} + kd^2h^4 + d^2h^4 + dh^3\big)\mathrm{e}^{(5(2L+1)^2+2\lambda)\eta(\alpha,M,\rho_h^{(2)})},
\end{aligned} \tag{120}$$

where, in the last inequality, we have used the fact that

$$\hat{T}\big(\alpha, M, \rho_h^{(2)}\big) = T - \eta(\alpha, M, \rho_h^{(2)}) \le T - \eta\big(\alpha, M, 0\big) = \hat{T}(\alpha, M, 0) \le T. \tag{121}$$

Now, letting ϵ_2 tend to zero and applying Fatou's lemma, we complete the proof. □

B.2 Discretization error of EM method

In this subsection, we present an enhanced discretization error bound for the EM method. When $k = 1$, the midpoint correction term vanishes, and the PM scheme (9) reduces to the classical EM scheme (8). Therefore, it is straightforward to obtain the following proposition with the same technique as in Proposition B.7.

Proposition B.8 (Discretization error of EM method)**.** *Suppose that Assumptions 3.2 and 3.3 hold. For any $n \in [N]_0$, consider the process $\{X_{t_n}^N\}$ generated by* (20) *using the uniform step-size. For any $\lambda \in (0,1)$, let the step-size satisfy*

$$h < \tfrac{2-\lambda}{2(2L+1)^2}. \tag{122}$$

Then there exists a constant C, independent of d and h, such that,

$$\mathcal{W}_2\big(\mu, \mathcal{L}(X_{t_N}^N)\big) \le C\mathrm{e}^{C\eta(\alpha,M,(2L+1)^2h+\lambda/2)}\sqrt{Td}h, \tag{123}$$

with $\eta(\alpha, M, (2L+1)^2h + \lambda/2)$ as defined in (30)*, where L comes from Lemma A.3.*

Proof. Using the same arguments as used in Proposition B.1, we first recast the backward SDE (5) as

$$\overleftarrow{X}_{t_{n+1}} = \overleftarrow{X}_{t_n} + hb_{t_n}(\overleftarrow{X}_{t_n}) + \sqrt{2}\Delta W_{t_n} + \mathcal{R}_{t_n}, \tag{124}$$

where

$$\mathcal{R}_{t_n} = \int_{t_n}^{t_{n+1}} \big[b_s(\overleftarrow{X}_s) - b_{t_n}(\overleftarrow{X}_{t_n})\big]\,\mathrm{d}s. \tag{125}$$

Using the same notation as before yields, for any $n \in [N]_0$,

$$\varepsilon_{t_n} := \overleftarrow{X}_{t_n} - X_{t_n}^N. \tag{126}$$

Subtracting (20) from this and taking expectations on both sides lead to

$$\begin{aligned}\mathbb{E}\big[\|\varepsilon_{t_{n+1}}\|^2\big] &= \mathbb{E}\big[\|\varepsilon_{t_n}\|^2\big] + h^2\,\mathbb{E}\Big[\big\|b_{t_n}(\overleftarrow{X}_{t_n}) - b_{t_n}(X_{t_n}^N)\big\|^2\Big] + \mathbb{E}\Big[\big\|\mathcal{R}_{t_n}\big\|^2\Big]\\ &\quad + 2h\,\mathbb{E}\Big[\big\langle \varepsilon_{t_n}, b_{t_n}(\overleftarrow{X}_{t_n}) - b_{t_n}(X_{t_n}^N)\big\rangle\Big] + 2\,\mathbb{E}\Big[\big\langle \varepsilon_{t_n}, \mathcal{R}_{t_n}\big\rangle\Big]\\ &\quad + 2h\,\mathbb{E}\Big[\big\langle b_{t_n}(\overleftarrow{X}_{t_n}) - b_{t_n}(X_{t_n}^N), \mathcal{R}_{t_n}\big\rangle\Big].\end{aligned} \tag{127}$$

For the estimate of the fourth term of (127), using Proposition A.4 gets,

$$2h\,\mathbb{E}\Big[\big\langle \varepsilon_{t_n}, b_{t_n}(\overleftarrow{X}_{t_n}) - b_{t_n}(X_{t_n}^N)\big\rangle\Big] \le \begin{cases} -2(2C_{T-t_n}+1)h\,\mathbb{E}\big[\|\varepsilon_{t_n}\|^2\big], & \text{for} \quad n \le N_h,\\ 2(2L+1)h\,\mathbb{E}\big[\|\varepsilon_{t_n}\|^2\big], & \text{for} \quad n > N_h.\end{cases} \tag{128}$$

The remaining terms are handled following the same approach as in the proof of Proposition B.7, namely by employing the Young inequality and the Lipschitz condition of $b_t(\cdot)$. Plugging the previous estimates into (127) yields

$$\mathbb{E}\big[\|\varepsilon_{t_{n+1}}\|^2\big] \le \gamma_{t_n}\,\mathbb{E}\big[\|\varepsilon_{t_n}\|^2\big] + 2\,\mathbb{E}\big[\|\mathcal{R}_{t_n}\|^2\big] + \tfrac{1}{\lambda h}\mathbb{E}\Big[\big\|\mathbb{E}\big[\mathcal{R}_{t_n}\big|\mathcal{F}_{t_n}\big]\big\|\Big]^2, \tag{129}$$

where

$$\gamma_{t_n} := \begin{cases} 1 + \big(2(2L+1)^2h - 2(2C_{T-t_n}+1) + \lambda\big)h, & \text{for} \quad n \le N_h\\ 1 + \big(4(2L+1)^2 + \lambda\big)h, & \text{for} \quad n > N_h.\end{cases} \tag{130}$$

Following the same arguments as used in the estimate of (107) and (110) in Proposition B.7, one derives

$$2\,\mathbb{E}\big[\|\mathcal{R}_{t_n}\|^2\big] \le 2\big(\mathcal{M}_1 + 4(1+4L^2)\big)dh^3, \quad \tfrac{1}{\lambda h}\mathbb{E}\Big[\big\|\mathbb{E}[\mathcal{R}_{t_n} \mid \mathcal{F}_{t_n}]\big\|^2\Big] \le \tfrac{\mathcal{M}_1 dh^3}{\lambda}. \tag{131}$$

Since h satisfies (122), then we take $\rho_h^{(3)} := (2(2L+1)^2h + \lambda)/2$, such that $\rho_h^{(3)} \in (\lambda/2, 1)$. Then, we have

$$h = \frac{2\rho_h - \lambda}{2(2L+1)^2} \le \frac{2(2C_{T-t_n}+1) - \lambda}{2(2L+1)^2},$$

i.e., $\gamma_{t_n} \le 1$ for $n \le N_h$. Combining the above estimates with (129) yields

$$\begin{aligned}\mathcal{W}_2^2\big(\mu, \mathcal{L}(X^N_{t_N})\big) &\le \mathbb{E}\big[\|\varepsilon_{t_N}\|^2\big] \\ &\le \gamma_{t_N}\, \mathbb{E}\big[\|\varepsilon_{t_N}\|^2\big] + C_{\mathcal{M}_1,L,\lambda} dh^3 \\ &\le \prod_{\ell=0}^{N-1} \gamma_{t_\ell}\, \mathbb{E}\big[\|\varepsilon_0\|^2\big] + C_{\mathcal{M}_1,L,\lambda} dh^3 \Big(\sum_{m=1}^{N-1} \prod_{\ell=m}^{N-1} \gamma_{t_\ell} + 1\Big) \\ &\le C_{\mathcal{M}_1,L,\lambda} dh^2 \left(\hat{T}(\alpha, M, 0) + \tfrac{1}{C_{L,\lambda}}\right) \mathrm{e}^{(4(2L+1)^2+\lambda)\eta(\alpha,M,\rho_h^{(3)})},\end{aligned} \tag{132}$$

where we also used $\overleftarrow{X}_0 - X_0^N = 0$ and an argument similar to the proof of (119). The proof of Proposition B.8 is completed. □

C Bound on initialization error

Before discussing the initialization error, we mention that we consider the coupling between the continuous time interpolations $\{X^N_{t_{n,j}}\}_{n\in[N-1]_0,\, j\in[k]_0}$ and $\{X^\infty_{t_{n,j}}\}_{n\in[N-1]_0,\, j\in[k]_0}$ with initialization satisfying

$$\mathcal{W}_2^2\big(\mathcal{L}(\overrightarrow{X}_T), \gamma^d\big) = \mathbb{E}\Big[\big\|X_0^N - X_0^\infty\big\|^2\Big].$$

We are now ready to present an initialization error bound for two types of methods.

C.1 Initialization error of PM method

Proposition C.1 (Initialization error of PM method)**.** *Suppose that Assumptions 3.2 and 3.3 hold. For any $n \in [N-1]_0$, $j \in [k]_0$, let $\{X^N_{t_{n,j}}\}$ and $\{X^\infty_{t_{n,j}}\}$ be produced by* (21) *with the uniform step-size*

$$0 < h < \tfrac{1}{C(L,\alpha)}, \tag{133}$$

where $C(L,\alpha)$ comes from (155)*. Then there exists a constant C, independent of d and h, such that,*

$$\mathcal{W}_2\big(\mathcal{L}(X^N_{t_N}), \mathcal{L}(X^\infty_{t_N})\big) \le C\mathrm{e}^{C\eta(\alpha,M,C(L,\alpha)h)}\mathrm{e}^{-T}\mathcal{W}_2(\mu,\gamma^d), \tag{134}$$

with $\eta\big(\alpha, M, C(L,\alpha)h\big)$ as defined in (30)*, where L comes from Lemma A.3.*

Proof. First, according to (21), we have, for any $n \in [N-1]_0$, $i \in [k-1]_0$,

$$\begin{aligned}X^N_{t_{n,i+1}} &= X^N_{t_n} + \tfrac{(i+1)h}{k} b_{t_n}(X^N_{t_n}) + h\sum_{j=0}^{i} B_{t_{n,j}} \Big(b_{t_{n,j}}(\tilde{X}^N_{t_{n,j}}) - b_{t_n}(X^N_{t_n})\Big) + \sqrt{\tfrac{2h}{k}} \sum_{j=0}^{i} Z_{t_{n,j}} \\ &= X^N_{t_n} + \Big(\tfrac{(i+1)h}{k} - h\sum_{j=0}^{i} B_{t_{n,j}}\Big) b_{t_n}(X^N_{t_n}) + h\sum_{j=0}^{i} B_{t_{n,j}} b_{t_{n,j}}(\tilde{X}^N_{t_{n,j}}) + \sqrt{\tfrac{2h}{k}} \sum_{j=0}^{i} Z_{t_{n,j}},\end{aligned} \tag{135}$$

and

$$\begin{aligned}X^\infty_{t_{n,i+1}} &= X^\infty_{t_n} + \tfrac{(i+1)h}{k} b_{t_n}(X^\infty_{t_n}) + h\sum_{j=0}^{i} B_{t_{n,j}} \Big(b_{t_{n,j}}(\tilde{X}^\infty_{t_{n,j}}) - b_{t_n}(X^\infty_{t_n})\Big) + \sqrt{\tfrac{2h}{k}} \sum_{j=0}^{i} Z_{t_{n,j}} \\ &= X^\infty_{t_n} + \Big(\tfrac{(i+1)h}{k} - h\sum_{j=0}^{i} B_{t_{n,j}}\Big) b_{t_n}(X^\infty_{t_n}) + h\sum_{j=0}^{i} B_{t_{n,j}} b_{t_{n,j}}(\tilde{X}^\infty_{t_{n,j}}) + \sqrt{\tfrac{2h}{k}} \sum_{j=0}^{i} Z_{t_{n,j}}.\end{aligned} \tag{136}$$

Taking $i = k-1$ in (135) and (136) gives, for any $n \in [N-1]_0$,

$$\begin{aligned}X^N_{t_{n+1}} - X^\infty_{t_{n+1}} = X^N_{t_n} - X^\infty_{t_n} &+ h\Big(1 - \sum_{j=0}^{k-1} B_{t_{n,j}}\Big)\Big(b_{t_n}(X^N_{t_n}) - b_{t_n}(X^\infty_{t_n})\Big) \\ &+ h\sum_{j=0}^{k-1} B_{t_{n,j}} \Big(b_{t_{n,j}}(\tilde{X}^N_{t_{n,j}}) - b_{t_{n,j}}(\tilde{X}^\infty_{t_{n,j}})\Big).\end{aligned} \tag{137}$$

To simplify the notation, we denote, for any $n \in [N-1]_0$, $j \in [k-1]_0$,

$$e_{t_n} := X^N_{t_n} - X^\infty_{t_n}, \quad \mathcal{M}_{t_{n,j}} := B_{t_{n,j}}\Big(b_{t_{n,j}}(\tilde{X}^N_{t_{n,j}}) - b_{t_{n,j}}(\tilde{X}^\infty_{t_{n,j}})\Big).$$

Squaring both sides of the above equation, taking expectations and using the Cauchy-Schwarz inequality show,

$$\begin{aligned}
\mathbb{E}\big[\|e_{t_{n+1}}\|^2\big] &= \mathbb{E}\big[\|e_{t_n}\|^2\big] + h^2\,\mathbb{E}\Big[\big(1-\sum_{j=0}^{k-1} B_{t_{n,j}}\big)^2\Big]\,\mathbb{E}\Big[\big\|b_{t_n}(X^N_{t_n}) - b_{t_n}(X^\infty_{t_n})\big\|^2\Big] \\
&\quad + h^2\,\mathbb{E}\Big[\big\|\sum_{j=0}^{k-1}\mathcal{M}_{t_{n,j}}\big\|^2\Big] + 2h\,\mathbb{E}\Big[\big\langle e_{t_n}, \sum_{j=0}^{k-1}\mathcal{M}_{t_{n,j}}\big\rangle\Big] \\
&\quad + 2h\,\mathbb{E}\Big[\Big\langle e_{t_n}, \Big(1-\sum_{j=0}^{k-1} B_{t_{n,j}}\Big)\Big(b_{t_n}(X^N_{t_n}) - b_{t_n}(X^\infty_{t_n})\Big)\Big\rangle\Big] \\
&\quad + 2h^2\,\mathbb{E}\Big[\Big\langle \Big(1-\sum_{j=0}^{k-1} B_{t_{n,j}}\Big)\Big(b_{t_n}(X^N_{t_n}) - b_{t_n}(X^\infty_{t_n})\Big), \sum_{j=0}^{k-1}\mathcal{M}_{t_{n,j}}\Big\rangle\Big] \qquad (138) \\
&\le \mathbb{E}\big[\|e_{t_n}\|^2\big] + 2h^2\,\mathbb{E}\Big[\big(1-\sum_{j=0}^{k-1} B_{t_{n,j}}\big)^2\Big]\,\mathbb{E}\Big[\big\|b_{t_n}(X^N_{t_n}) - b_{t_n}(X^\infty_{t_n})\big\|^2\Big] \\
&\quad + 2h^2\,\mathbb{E}\Big[\big\|\sum_{j=0}^{k-1}\mathcal{M}_{t_{n,j}}\big\|^2\Big] + 2h\,\mathbb{E}\Big[\big\langle e_{t_n}, \sum_{j=0}^{k-1}\mathcal{M}_{t_{n,j}}\big\rangle\Big] \\
&\quad + 2h\,\mathbb{E}\Big[\Big\langle e_{t_n}, \Big(1-\sum_{j=0}^{k-1} B_{t_{n,j}}\Big)\Big(b_{t_n}(X^N_{t_n}) - b_{t_n}(X^\infty_{t_n})\Big)\Big\rangle\Big].
\end{aligned}$$

Noting that $\mathbb{E}[\sum_{j=0}^{k-1} B_{t_{n,j}}] = 1$ and $\mathbb{E}\big[\big(\sum_{j=0}^{k-1} B_{t_{n,j}}\big)^2\big] \le 2$, one directly derives that

$$\mathbb{E}\Big[\Big(1-\sum_{j=0}^{k-1} B_{t_{n,j}}\Big)^2\Big] = 1 - 2\,\mathbb{E}\Big[\sum_{j=0}^{k-1} B_{t_{n,j}}\Big] + \mathbb{E}\Big[\Big(\sum_{j=0}^{k-1} B_{t_{n,j}}\Big)^2\Big] \le 1, \qquad (139)$$

and

$$\begin{aligned}
&2h\,\mathbb{E}\Big[\Big\langle e_{t_n}, \Big(1-\sum_{j=0}^{k-1} B_{t_{n,j}}\Big)\big(b_{t_n}(X^N_{t_n}) - b_{t_n}(X^\infty_{t_n})\big)\Big\rangle\Big] \\
&= 2h\,\mathbb{E}\Big[\Big\langle e_{t_n}, \big(b_{t_n}(X^N_{t_n}) - b_{t_n}(X^\infty_{t_n})\big)\,\mathbb{E}\Big[\big(1-\sum_{j=0}^{k-1} B_{t_{n,j}}\big)\big|\sigma\big(X^\infty_{t_n}, X^N_{t_n}\big)\Big]\Big\rangle\Big] \qquad (140) \\
&= 0.
\end{aligned}$$

It remains to estimate $2h\ \mathbb{E}\Big[\big\langle e_{t_n}, \sum_{j=0}^{k-1}\mathcal{M}_{t_{n,j}}\big\rangle\Big]$, which can be divided into two cases depending on different ranges of n.

Case I: $(n+1)k \le N_h$. In the case $(n+1)k \le N_h$, we do the following decomposition:

$$\begin{aligned}
&2h\ \mathbb{E}\Big[\big\langle e_{t_n}, \sum_{j=0}^{k-1}\mathcal{M}_{t_{n,j}}\big\rangle\Big] \\
&= \underbrace{2h\sum_{j=0}^{k-1}\mathbb{E}\Big[\big\langle \tilde{X}^N_{t_{n,j}} - \tilde{X}^\infty_{t_{n,j}}, B_{t_{n,j}}\big(b_{t_{n,j}}(\tilde{X}^N_{t_{n,j}}) - b_{t_{n,j}}(\tilde{X}^\infty_{t_{n,j}})\big)\big\rangle\Big]}_{=:\mathbb{T}_1} \qquad (141) \\
&\quad \underbrace{+\,2h\sum_{j=0}^{k-1}\mathbb{E}\Big[\Big\langle X^N_{t_n} - X^\infty_{t_n} - \tilde{X}^N_{t_{n,j}} + \tilde{X}^\infty_{t_{n,j}}, B_{t_{n,j}}\big(b_{t_{n,j}}(\tilde{X}^N_{t_{n,j}}) - b_{t_{n,j}}(\tilde{X}^\infty_{t_{n,j}})\big)\Big\rangle\Big]}_{=:\mathbb{T}_2}.
\end{aligned}$$

The definitions in (21) give, for any $n \in [N-1]_0$ and $j \in [k-1]_0$,

$$\tilde{X}^N_{t_{n,j}} - \tilde{X}^\infty_{t_{n,j}} = X^N_{t_n} - X^\infty_{t_n} + \tfrac{jh}{k}\Big(b_{t_n}(X^N_{t_n}) - b_{t_n}(X^\infty_{t_n})\Big). \tag{142}$$

Regarding $\mathbb{T}_1$, using the identity $\mathbb{E}[B_{t_{n,j}}] = 1/k$, the inequality $(a+b)^2 \geq \frac{1}{2}a^2 - b^2$, and Proposition A.4 implies

$$\begin{aligned}
\mathbb{T}_1 &\leq -\tfrac{2h}{k}\sum_{j=0}^{k-1}(2C_{T-t_{n,j}}+1)\,\mathbb{E}\Big[\big\|\tilde{X}^N_{t_{n,j}} - \tilde{X}^\infty_{t_{n,j}}\big\|^2\Big] \\
&\leq -\tfrac{2h}{k}\min_{0\leq j\leq k-1}(2C_{T-t_{n,j}}+1)\sum_{j=0}^{k-1}\Big(\tfrac{1}{2}\,\mathbb{E}\big[\|e_{t_n}\|^2\big] - (\tfrac{jh}{k})^2\,\mathbb{E}\Big[\big\|b_{t_n}(X^N_{t_n}) - b_{t_n}(X^\infty_{t_n})\big\|^2\Big]\Big) \\
&\leq -h\min_{0\leq j\leq k-1}(2C_{T-t_{n,j}}+1)\,\mathbb{E}\big[\|e_{t_n}\|^2\big] + 2h^3(2L+1)^2(1\vee 2\alpha - 1)\,\mathbb{E}\big[\|e_{t_n}\|^2\big],
\end{aligned} \tag{143}$$

where the last inequality follows from the fact that, for any $t \in [0,T]$,

$$\begin{aligned}
2C_t + 1 &= \frac{2\alpha}{\alpha + (1-\alpha)e^{-2t}} - \frac{2e^{-2t}M}{(\alpha+(1-\alpha)e^{-2t})^2} - 1 \\
&\leq \frac{2\alpha}{\alpha + (1-\alpha)e^{-2t}} - 1 \\
&\leq 1 \vee (2\alpha - 1).
\end{aligned} \tag{144}$$

For the estimate of $\mathbb{T}_2$, using (142), the Young inequality and Lipschitz condition of $b_t(\cdot)$ leads to

$$\begin{aligned}
\mathbb{T}_2 &= 2h\sum_{j=0}^{k-1}\mathbb{E}\Big[\big\langle \tfrac{jh}{k}\big(b_{t_n}(X^\infty_{t_n}) - b_{t_n}(X^N_{t_n})\big), B_{t_{n,j}}\big(b_{t_{n,j}}(\tilde{X}^N_{t_{n,j}}) - b_{t_{n,j}}(\tilde{X}^\infty_{t_{n,j}})\big)\big\rangle\Big] \\
&= 2h\,\mathbb{E}\Big[\Big\langle \big(b_{t_n}(X^\infty_{t_n}) - b_{t_n}(X^N_{t_n})\big), \sum_{j=0}^{k-1}\tfrac{jh}{k}B_{t_{n,j}}\big(b_{t_{n,j}}(\tilde{X}^N_{t_{n,j}}) - b_{t_{n,j}}(\tilde{X}^\infty_{t_{n,j}})\big)\Big\rangle\Big] \\
&\leq h^2(2L+1)^2\,\mathbb{E}\big[\|e_{t_n}\|^2\big] + h^2\,\mathbb{E}\Big[\big\|\sum_{j=0}^{k-1}\tfrac{j}{k}B_{t_{n,j}}\big(b_{t_{n,j}}(\tilde{X}^N_{t_{n,j}}) - b_{t_{n,j}}(\tilde{X}^\infty_{t_{n,j}})\big)\big\|^2\Big] \\
&= h^2(2L+1)^2\,\mathbb{E}\big[\|e_{t_n}\|^2\big] + h^2\,\mathbb{E}\Big[\big\|\sum_{j=0}^{k-1}\tfrac{j}{k}\mathcal{M}_{t_{n,j}}\big\|^2\Big].
\end{aligned} \tag{145}$$

As a result,

$$\begin{aligned}
2h\;\mathbb{E}\Big[\big\langle e_{t_n}, \sum_{j=0}^{k-1}\mathcal{M}_{t_{n,j}}\big\rangle\Big] \leq\; & -h\min_{0\leq j\leq k-1}(2C_{T-t_{n,j}}+1)\,\mathbb{E}\big[\|e_{t_n}\|^2\big] \\
& + h^2(2L+1)^2\Big(1 + 2\big(1\vee(2\alpha-1)\big)\Big)\,\mathbb{E}\big[\|e_{t_n}\|^2\big] \\
& + h^2\,\mathbb{E}\Big[\big\|\sum_{j=0}^{k-1}\tfrac{j}{k}\mathcal{M}_{t_{n,j}}\big\|^2\Big].
\end{aligned} \tag{146}$$

Case II: $(n+1)k > N_h$. In the case $(n+1)k > N_h$, we use the Young inequality to obtain

$$2h\;\mathbb{E}\Big[\big\langle e_{t_n}, \sum_{j=0}^{k-1}\mathcal{M}_{t_{n,j}}\big\rangle\Big] \leq h\,\mathbb{E}\big[\|e_{t_n}\|^2\big] + h\,\mathbb{E}\Big[\big\|\sum_{j=0}^{k-1}\mathcal{M}_{t_{n,j}}\big\|^2\Big]. \tag{147}$$

Next, we cope with $\mathbb{E}\big[\|\sum_{j=0}^{k-1}\mathcal{M}_{t_{n,j}}\|^2\big]$. For any $j\in[k]_0$ and $n\in[N-1]_0$, $\mathbb{E}[B_{t_{n,j}}]=\mathbb{E}[B^2_{t_{n,j}}]=1/k$, one obtains

$$\begin{aligned}\mathbb{E}\Big[\|\sum_{j=0}^{k-1}\mathcal{M}_{t_{n,j}}\|^2\Big] &= \mathbb{E}\Big[\|\sum_{j=0}^{k-1}B_{t_{n,j}}\big(b_{t_{n,j}}(\tilde{X}^N_{t_{n,j}})-b_{t_{n,j}}(\tilde{X}^\infty_{t_{n,j}})\big)\|^2\Big]\\ &= \underbrace{\tfrac{2}{k^2}\sum_{0\le i<j\le k-1}\mathbb{E}\Big[\big\langle b_{t_{n,j}}(\tilde{X}^N_{t_{n,j}})-b_{t_{n,j}}(\tilde{X}^\infty_{t_{n,j}}),b_{t_{n,j}}(\tilde{X}^N_{t_{n,j}})-b_{t_{n,j}}(\tilde{X}^\infty_{t_{n,j}})\big\rangle\Big]}_{=:\mathbb{T}_3}\\ &\quad+\underbrace{\tfrac{1}{k}\sum_{j=0}^{k-1}\mathbb{E}\Big[\|b_{t_{n,j}}(\tilde{X}^N_{t_{n,j}})-b_{t_{n,j}}(\tilde{X}^\infty_{t_{n,j}})\|^2\Big]}_{=:\mathbb{T}_4}.\end{aligned}\tag{148}$$

In the following we deal with the above two items separately. Using the Cauchy-Schwarz inequality and the Lipschitz condition of $b_t(\cdot)$, one can derive that

$$\begin{aligned}\mathbb{T}_3 &\le \tfrac{2}{k^2}\sum_{0\le i<j\le k-1}\mathbb{E}\Big[\|b_{t_{n,i}}(\tilde{X}^N_{t_{n,i}})-b_{t_{n,i}}(\tilde{X}^\infty_{t_{n,i}})\|\cdot\|b_{t_{n,j}}(\tilde{X}^N_{t_{n,j}})-b_{t_{n,j}}(\tilde{X}^\infty_{t_{n,j}})\|\Big]\\ &\le \sup_{0\le i\le k-1}\mathbb{E}\Big[\|b_{t_{n,i}}(\tilde{X}^N_{t_{n,i}})-b_{t_{n,i}}(\tilde{X}^\infty_{t_{n,i}})\|^2\Big]\\ &\le (2L+1)^2\sup_{0\le i\le k-1}\mathbb{E}\Big[\|e_{t_n}+\tfrac{ih}{k}\big(b_{t_n}(X^N_{t_n})-b_{t_n}(X^\infty_{t_n})\big)\|^2\Big]\\ &\le 2(2L+1)^2\big(1+h^2(2L+1)^2\big)\,\mathbb{E}\big[\|e_{t_n}\|^2\big].\end{aligned}\tag{149}$$

Regarding $\mathbb{T}_4$, we use the Lipschitz condition of $b_t(\cdot)$ to obtain

$$\mathbb{T}_4\le 2(2L+1)^2\big(1+h^2(2L+1)^2\big)\,\mathbb{E}\big[\|e_{t_n}\|^2\big].\tag{150}$$

Equipped with estimates (149) and (150), one derives that

$$\mathbb{E}\Big[\|\sum_{j=0}^{k-1}\mathcal{M}_{t_{n,j}}\|^2\Big]\le 4\big((2L+1)^2+1\big)^2\,\mathbb{E}\big[\|e_{t_n}\|^2\big].\tag{151}$$

For the weighted sum, the same expansion and the bounds $0\le i/k,j/k\le 1$ give

$$\begin{aligned}&\mathbb{E}\Big[\|\sum_{j=0}^{k-1}\tfrac{j}{k}\mathcal{M}_{t_{n,j}}\|^2\Big]\\ &\le \tfrac{2}{k^2}\sum_{0\le i<j\le k-1}\tfrac{ij}{k^2}\,\mathbb{E}\Big[\|b_{t_{n,i}}(\tilde{X}^N_{t_{n,i}})-b_{t_{n,i}}(\tilde{X}^\infty_{t_{n,i}})\|\cdot\|b_{t_{n,j}}(\tilde{X}^N_{t_{n,j}})-b_{t_{n,j}}(\tilde{X}^\infty_{t_{n,j}})\|\Big]\\ &\quad+\tfrac{1}{k}\sum_{j=0}^{k-1}\big(\tfrac{j}{k}\big)^2\mathbb{E}\Big[\|b_{t_{n,j}}(\tilde{X}^N_{t_{n,j}})-b_{t_{n,j}}(\tilde{X}^\infty_{t_{n,j}})\|^2\Big]\\ &\le 4\big((2L+1)^2+1\big)^2\,\mathbb{E}\big[\|e_{t_n}\|^2\big].\end{aligned}\tag{152}$$

Based on the above estimates, we obtain

$$\mathbb{E}\big[\|e_{t_N}\|^2\big]\le\prod_{\ell=0}^{N-1}\Gamma_{t_\ell}\mathbb{E}\big[\|e_0\|^2\big],\tag{153}$$

where

$$\Gamma_{t_n}:=\begin{cases}1+\big(C(L,\alpha)h-\min_{0\le j\le k-1}(2C_{T-t_{n,j}}+1)\big)h, & \text{for } (n+1)k\le N_h\\ 1+\tilde{C}(L)h, & \text{for } (n+1)k>N_h\end{cases}\tag{154}$$

and

$$C(L,\alpha):=\big((2L+1)^2+1\big)^2\big(15+2(1\vee 2(\alpha-1))\big),\quad \tilde{C}(L):=1+4\big((2L+1)^2+1\big)^2.\tag{155}$$

With regard to $\mathbb{E}\big[\|e_0\|^2\big]$, it follows from [47, Proposition C.2] that,

$$\mathbb{E}\big[\|e_0\|^2\big] = \mathcal{W}_2^2\big(\mathcal{L}(\overrightarrow{X}_T), \gamma^d\big) \le \mathrm{e}^{-2T}\mathcal{W}_2^2(\mu, \gamma^d). \tag{156}$$

Since h satisfies (133), we take $\rho_h^{(4)} := C(L,\alpha)h$. It follows that $\Gamma_{t_n} \le 1$ whenever $(n+1)k \le N_h$. Using the definition of N_h with $\rho = \rho_h^{(4)}$, we also have

$$\big(N - \lfloor N_h/k \rfloor\big)h \le T - \widehat{T}(\alpha, M, \rho_h^{(4)}) + h = \eta(\alpha, M, \rho_h^{(4)}) + h.$$

Then

$$\prod_{\ell=0}^{N-1} \Gamma_{t_\ell} \le \prod_{\ell=N_h}^{N-1} \Gamma_{t_\ell} \le \big(1 + \tilde{C}(L)h\big)^{N-N_h} \le \mathrm{e}^{\tilde{C}(L)\eta(\alpha,M,\rho_h^{(4)})}. \tag{157}$$

All together, we have

$$\mathcal{W}_2\big(\mathcal{L}(X_{t_N}^N), \mathcal{L}(X_{t_N}^\infty)\big) \le C\mathrm{e}^{\tilde{C}(L)\eta(\alpha,M,\rho_h^{(4)})}\mathrm{e}^{-T}\mathcal{W}_2(\mu, \gamma^d). \tag{158}$$

The proof of Proposition C.1 is completed. □

C.2 Initialization error of EM method

The aim of this subsection is to prove the initialization error of EM method, by utilizing the same technique in Proposition C.1.

Proposition C.2 (Initialization error of EM method)**.** *Suppose that Assumptions 3.2 and 3.3 hold. For any $n \in [N]_0$ and let $\{X_{t_n}^N\}$ and $\{X_{t_n}^\infty\}$ be produced by* (20) *using the uniform step-size*

$$h < \tfrac{2}{(2L+1)^2}. \tag{159}$$

Then there exists a constant C, independent of d and h, such that,

$$\mathcal{W}_2\big(\mathcal{L}(X_{t_N}^N), \mathcal{L}(X_{t_N}^\infty)\big) \le C\mathrm{e}^{C\eta(\alpha,M,(2L+1)^2h/2)}\mathrm{e}^{-T}\mathcal{W}_2(\mu, \gamma^d), \tag{160}$$

with $\eta(\alpha, M, (2L+1)^2h/2)$ as defined in (30)*, where L comes from Lemma A.3.*

Proof. By following the proof of Proposition C.1, using the same notation and arguments as in the estimate of (157), and taking $\rho_h^{(5)} := (2L+1)^2h/2$, one will arrive at

$$\begin{aligned}\mathbb{E}\big[\|e_{t_N}\|^2\big] &\le \prod_{\ell=0}^{N-1} \Gamma_{t_\ell}\mathbb{E}\big[\|e_0\|^2\big] \\ &\le \mathrm{e}^{(2L+3)^2\eta(\alpha,M,\rho_h^{(5)})}\mathrm{e}^{-2T}\mathcal{W}_2^2(\mu, \gamma^d).\end{aligned} \tag{161}$$

where

$$\Gamma_{t_n} := \begin{cases} 1 + \big((2L+1)^2h - 2(2C_{T-t_n}+1)\big)h, & \text{for } n \le N_h \\ 1 + (2L+1)(2L+3)h, & \text{for } n > N_h. \end{cases} \tag{162}$$

The proof is completed. □

D Bound on score matching error

D.1 Score matching error of PM method

Proposition D.1 (Score matching error of PM method)**.** *Suppose that Assumptions 3.2, 3.3 and 3.4 hold. For any $n \in [N-1]_0$, $j \in [k]_0$ and let $\{X_{t_{n,j}}^\infty\}$ and $\{Y_{t_{n,j}}\}$ be produced by* (21) *and* (9) *using the uniform step-size*

$$h < \tfrac{1}{C(L,\alpha)},$$

where $C(L,\alpha)$ comes from (155)*. Then there exists a constant C, independent of d and h, such that,*

$$\mathcal{W}_2\big(\mathcal{L}(X_{t_N}^\infty), \mathcal{L}(Y_{t_N})\big) \le C\mathrm{e}^{C\eta(\alpha,M,C(L,\alpha)h)}T\epsilon_0, \tag{163}$$

with $\eta(\alpha, M, C(L,\alpha)h)$ as defined in (30)*, where L comes from Lemma A.3.*

Proof. First, we introduce a new auxiliary process as, for any $n \in [N]_0$, $i \in [k-1]_0$,

$$\widehat{Y}_{t_{n,i}} = Y_{t_n} + \tfrac{ih}{k} b_{t_n}(Y_{t_n}) + \sum_{j=0}^{i-1} \sqrt{\tfrac{2h}{k}} Z_{t_{n,j}}. \tag{164}$$

Next, we recast the second term of (9) as

$$\begin{aligned}
Y_{t_{n+1}} &= Y_{t_n} + h\big(1 - \sum_{j=0}^{k-1} B_{t_{n,j}}\big)\big(- Y_{t_n} + 2\tilde{s}_{\theta^\star}(T - t_n, Y_{t_n})\big) \\
&\quad + h\sum_{j=0}^{k-1} B_{t_{n,j}}\Big(- \tilde{Y}_{t_{n,j}} + 2\tilde{s}_{\theta^\star}(T - t_{n,j}, \tilde{Y}_{t_{n,j}})\Big) + \sum_{j=0}^{k-1} \sqrt{\tfrac{2h}{k}} Z_{t_{n,j}} \\
&= Y_{t_n} + h\big(1 - \sum_{j=0}^{k-1} B_{t_{n,j}}\big) b_{t_n}(Y_{t_n}) + h\sum_{j=0}^{k-1} B_{t_{n,j}} b_{t_{n,j}}(\widehat{Y}_{t_{n,j}}) + \sum_{j=0}^{k-1} \sqrt{\tfrac{2h}{k}} Z_{t_{n,j}} \\
&\quad + h\big(1 - \sum_{j=0}^{k-1} B_{t_{n,j}}\big)\Big(2\tilde{s}_{\theta^\star}(T - t_n, Y_{t_n}) - 2\nabla \log \tilde{p}_{T-t_n}(Y_{t_n})\Big) \\
&\quad + h\sum_{j=0}^{k-1} B_{t_{n,j}}\Big(- \tilde{Y}_{t_{n,j}} + 2\tilde{s}_{\theta^\star}(T - t_{n,j}, \tilde{Y}_{t_{n,j}}) - b_{t_{n,j}}(\widehat{Y}_{t_{n,j}})\Big).
\end{aligned} \tag{165}$$

For simplicity, we denote

$$\mathcal{E}_{t_n} := X^{\infty}_{t_n} - Y_{t_n}, \quad n \in [N]_0.$$

Subtracting (136) from (165), taking the norm of both sides of the inequality, and using the triangle inequality yield

$$\begin{aligned}
\big\|\mathcal{E}_{t_{n+1}}\big\|_{L^2(\Omega;\mathbb{R}^d)} &= \Big\|\mathcal{E}_{t_n} + h\big(1 - \sum_{j=0}^{k-1} B_{t_{n,j}}\big)\Big(b_{t_n}(X^{\infty}_{t_n}) - b_{t_n}(Y_{t_n})\Big) \\
&\quad + h\sum_{j=0}^{k-1} B_{t_{n,j}}\Big(b_{t_{n,j}}(\tilde{X}^{\infty}_{t_{n,j}}) - b_{t_{n,j}}(\widehat{Y}_{t_{n,j}})\Big)\Big\|_{L^2(\Omega;\mathbb{R}^d)} \\
&\quad + \Big\|h\big(1 - \sum_{j=0}^{k-1} B_{t_{n,j}}\big)\Big(2\tilde{s}_{\theta^\star}(T - t_n, Y_{t_n}) - 2\nabla \log \tilde{p}_{T-t_n}(Y_{t_n})\Big)\Big\|_{L^2(\Omega;\mathbb{R}^d)} \\
&\quad + \Big\|h\sum_{j=0}^{k-1} B_{t_{n,j}}\Big(- \tilde{Y}_{t_{n,j}} + 2\tilde{s}_{\theta^\star}(T - t_{n,j}, \tilde{Y}_{t_{n,j}}) - b_{t_{n,j}}(\widehat{Y}_{t_{n,j}})\Big)\Big\|_{L^2(\Omega;\mathbb{R}^d)} \\
&=: A_1 + A_2 + A_3.
\end{aligned} \tag{166}$$

For the estimate of A_2, using Assumption 3.4, one gets

$$A_2^2 \le 4h^2 \mathbb{E}\Big[\big(1 - \sum_{j=0}^{k-1} B_{t_{n,j}}\big)^2\Big]\epsilon_0^2 \le 4h^2\epsilon_0^2,$$

and hence $A_2 \le 2h\epsilon_0$. In the following, we cope with A_3. Applying the triangle inequality, the Lipschitz condition of $b_t(\cdot)$ and Assumption 3.4 yields

$$\begin{aligned}
A_3^2 \le\, & 2h^2\,\mathbb{E}\Big[\Big\|\sum_{j=0}^{k-1} B_{t_{n,j}}\big(-\tilde{Y}_{t_{n,j}} + 2\tilde{s}_{\theta^\star}(T-t_{n,j},\tilde{Y}_{t_{n,j}}) - b_{t_{n,j}}(\tilde{Y}_{t_{n,j}})\big)\Big\|^2\Big] \\
& + 2h^2\,\mathbb{E}\Big[\Big\|\sum_{j=0}^{k-1} B_{t_{n,j}}\big(b_{t_{n,j}}(\tilde{Y}_{t_{n,j}}) - b_{t_{n,j}}(\widehat{Y}_{t_{n,j}})\big)\Big\|^2\Big] \\
\le & 4h^2 \sup_{0\le j\le k-1} \mathbb{E}\Big[\big\|2\tilde{s}_{\theta^\star}(T-t_{n,j},\tilde{Y}_{t_{n,j}}) - 2\nabla\log\tilde{p}_{T-t_{n,j}}(\tilde{Y}_{t_{n,j}})\big\|^2\Big] \\
& + 4(2L+1)^2h^2 \sup_{0\le j\le k-1} \mathbb{E}\Big[\big\|\tilde{Y}_{t_{n,j}} - \widehat{Y}_{t_{n,j}}\big\|^2\Big] \\
\le & 16\big(1+(2L+1)^2\big)h^2\epsilon_0^2,
\end{aligned} \tag{167}$$

where we use the elementary estimate

$$\mathbb{E}\Big[\big\|\sum_{j=0}^{k-1} B_{t_{n,j}} V_j\big\|^2\Big] \le 2 \sup_{0\le j\le k-1} \mathbb{E}\big[\|V_j\|^2\big].$$

Following the same argument as in Proposition C.1, we get

$$\|\mathcal{E}_{t_{n+1}}\|_{L^2(\Omega;\mathbb{R}^d)} \le \zeta_{t_n}\|\mathcal{E}_{t_n}\|_{L^2(\Omega;\mathbb{R}^d)} + \big(2+4\sqrt{1+(2L+1)^2}\big)h\epsilon_0, \tag{168}$$

where

$$\zeta_{t_n} := \begin{cases} \Big(1+\big(C(L,\alpha)h - \min_{0\le j\le k-1}(2C_{T-t_{n,j}}+1)\big)h\Big)^{1/2}, & \text{for } (n+1)k \le N_h \\ 1+\tilde{C}(L)h, & \text{for } (n+1)k > N_h. \end{cases} \tag{169}$$

Recalling that we already know $\zeta_{t_n} \le 1$ for $(n+1)k \le N_h$, we obtain by iteration that

$$\big\|\mathcal{E}_{t_N}\big\|_{L^2(\Omega;\mathbb{R}^d)} \le \prod_{\ell=0}^{N-1}\zeta_{t_\ell}\big\|\mathcal{E}_0\big\|_{L^2(\Omega;\mathbb{R}^d)} + \Big(\sum_{m=1}^{N-1}\prod_{\ell=m}^{N-1}\zeta_{t_\ell}+1\Big)\Big(2+2\sqrt{2(2L+1)^2+1}\Big)h\epsilon_0, \tag{170}$$

Now, by virtue of (119) and the fact that $X_0^\infty - Y_0 = 0$, we get

$$\mathcal{W}_2\big(\mathcal{L}(X_{t_N}^\infty),\mathcal{L}(Y_{t_N})\big) \le C\left(\hat{T}(\alpha,M,0)+\tfrac{1}{\tilde{C}(L)}\right)\mathrm{e}^{\tilde{C}(L)\eta(\alpha,M,\rho_h^{(4)})}\epsilon_0. \tag{171}$$

The proof of Proposition D.1 is finished. □

D.2 Score matching error of EM method

Proposition D.2 (Score matching error of EM method)**.** *Suppose that Assumptions 3.2, 3.3 and 3.4 hold. For any $n \in [N]_0$ and let $\{X_{t_n}^\infty\}$ and $\{Y_{t_n}\}$ be produced by* (20) *and* (8) *using the uniform step-size*

$$h < \tfrac{2}{(2L+1)^2}.$$

Then there exists a constant C, independent of d and h, such that,

$$\mathcal{W}_2\big(\mathcal{L}(X_{t_N}^\infty),\mathcal{L}(Y_{t_N})\big) \le C\mathrm{e}^{C\eta(\alpha,M,(2L+1)^2h/2)}T\epsilon_0, \tag{172}$$

with $\eta(\alpha,M,(2L+1)^2h/2)$ as defined in (30)*, where L comes from Lemma A.3.*

Proof. By employing the same notation and arguments as those used in the proof of Proposition D.1, along with the triangle inequality, one derives

$$\begin{aligned}
\big\|\mathcal{E}_{t_{n+1}}\big\|_{L^2(\Omega;\mathbb{R}^d)} = & \underbrace{\big\|\mathcal{E}_{t_n} + h\big(b_{t_n}(X_{t_n}^\infty) - b_{t_n}(Y_{t_n})\big)\big\|_{L^2(\Omega;\mathbb{R}^d)}}_{=:\mathcal{A}_1} \\
& + \underbrace{\Big\|h\big(2\tilde{s}_{\theta^\star}(T-t_n,Y_{t_n}) - 2\nabla\log\tilde{p}_{T-t_n}(Y_{t_n})\big)\Big\|_{L^2(\Omega;\mathbb{R}^d)}}_{=:\mathcal{A}_2}.
\end{aligned} \tag{173}$$

By Assumption 3.4, we have

$$\mathcal{A}_2 \le 2h\epsilon_0.$$

With regard to $\mathcal{A}_1$, by revisiting the proof of Proposition C.2 and invoking the above estimation together with the fact that $X_0^\infty = Y_0 \sim \gamma^d$, we deduce

$$\begin{aligned}\left\|\mathcal{E}_{t_N}\right\|_{L^2(\Omega;\mathbb{R}^d)} &\le \prod_{\ell=0}^{N-1}\zeta_{t_\ell}\|\mathcal{E}_0\|_{L^2(\Omega;\mathbb{R}^d)} + \Big(\sum_{m=1}^{N-1}\prod_{\ell=m}^{N-1}\zeta_{t_\ell}+1\Big)2h\epsilon_0\\ &\le \Big(\sum_{m=1}^{N-1}\prod_{\ell=m}^{N-1}\zeta_{t_\ell}+1\Big)2h\epsilon_0\\ &\le C\left(\hat{T}(\alpha,M,0)+\tfrac{1}{(2L+1)}\right)\mathrm{e}^{(2L+1)\eta(\alpha,M,\rho_h^{(5)})}\epsilon_0,\end{aligned} \tag{174}$$

with

$$\zeta_{t_n} := \begin{cases}\big(1+\big((2L+1)^2h-2(2C_{T-t_n}+1)\big)h\big)^{1/2}, & \text{for } n\le N_h\\ 1+(2L+1)h, & \text{for } n>N_h.\end{cases} \tag{175}$$

The proof is completed. □

E Proof of Proposition 3.8

Proof. The explicit solution of the OU process (2) gives

$$\overrightarrow{X}_t \overset{\mathrm{d}}{=} \mathrm{e}^{-t}\overrightarrow{X}_0 + \sqrt{1-\mathrm{e}^{-2t}}\,Z, \quad Z\sim\gamma^d,$$

where Z is independent of $\overrightarrow{X}_0$. For the sake of notation, we denote

$$a_r := e^{-r}, \qquad v_r := 1+e^{-2r}(\sigma^2-1) \ge v := \min\{1,\sigma^2\} > 0.$$

Then, we have that the density distribution of this law is

$$p_r(x) = \sum_{i=1}^m \beta_i \frac{1}{(2\pi v_r)^{d/2}}\exp\left(-\frac{\|x-a_r\mu_i\|^2}{2v_r}\right). \tag{176}$$

Consequently, the score function of p_r is given by

$$\begin{aligned}\nabla\log p_r(x) &= -\frac{x}{v_r} + \frac{a_r}{v_r}\frac{\sum_{i=1}^m \beta_i\mu_i\exp\left(-\frac{\|x-a_r\mu_i\|^2}{2v_r}\right)}{\sum_{i=1}^m\beta_i\exp\left(-\frac{\|x-a_r\mu_i\|^2}{2v_r}\right)}\\ &= -\frac{x}{v_r} + \frac{a_r}{v_r}\frac{\sum_{i=1}^m \beta_i\mu_i\exp\left(-\frac{a_r^2\|\mu_i\|^2}{2v_r}+\frac{a_r\langle\mu_i,x\rangle}{v_r}\right)}{\sum_{i=1}^m\beta_i\exp\left(-\frac{a_r^2\|\mu_i\|^2}{2v_r}+\frac{a_r\langle\mu_i,x\rangle}{v_r}\right)}\\ &=: -\frac{x}{v_r} + \frac{a_r}{v_r}\frac{\sum_{i=1}^m\beta_i\mu_i\psi_i(r,x)}{\sum_{i=1}^m\beta_i\psi_i(r,x)},\end{aligned} \tag{177}$$

where

$$\psi_i(r,x) = \exp\left(\tfrac{a_r}{v_r}\langle\mu_i,x\rangle - \tfrac{a_r^2}{2v_r}\|\mu_i\|^2\right).$$

Define

$$\omega_i(r,x) := \frac{\beta_i\psi_i(r,x)}{\sum_{j=1}^m\beta_j\psi_j(r,x)}, \qquad \overline{\mu}_r(x) := \sum_{i=1}^m\omega_i(r,x)\mu_i.$$

As a consequence,

$$\nabla\log p_r(x) = -\frac{x}{v_r} + \frac{a_r}{v_r}\overline{\mu}_r(x).$$

It follows that

$$\partial_r\overline{\mu}_r(x) = \sum_{i=1}^m\mu_i\partial_r\omega_i(r,x) = \sum_{i=1}^m\omega_i(r,x)\big(\mu_i-\overline{\mu}_r(x)\big)\partial_r\ell_i(r,x), \tag{178}$$

where

$$\ell_i(r,x) := \log \psi_i(r,x) = \tfrac{a_r}{v_r}\langle \mu_i, x\rangle - \tfrac{a_r^2}{2v_r}\|\mu_i\|^2.$$

Before proceeding further, we give some estimates. First, we have

$$\left|\left(\tfrac{a_r}{v_r}\right)'\right| = \left|\tfrac{a_r'}{v_r} - \tfrac{a_r v_r'}{v_r^2}\right| \le \tfrac{1}{v} + \tfrac{2|\sigma^2-1|}{v^2} =: C_1,$$

and

$$\left|\left(\tfrac{a_r^2}{2v_r}\right)'\right| = \left|\tfrac{a_r a_r'}{v_r} - \tfrac{a_r^2 v_r'}{2v_r^2}\right| \le \tfrac{1}{v} + \tfrac{|\sigma^2-1|}{v^2} =: C_2.$$

Define $R := \max_{1\le i\le m}\|\mu_i\|$, which is independent of d. By using the Cauchy-Schwarz inequality and the above estimates, we can derive

$$\left|\partial_r \ell_i(r,x)\right| = \left|\left(\tfrac{a_r}{v_r}\right)'\langle \mu_i, x\rangle - \left(\tfrac{a_r^2}{2v_r}\right)'\|\mu_i\|^2\right| \le C_1 R\|x\| + C_2 R^2.$$

Since $\overline{\mu}_r(x)$ is a convex combination of the vectors μ_i, we have $\|\overline{\mu}_r(x)\| \le R$. Using $\|\mu_i - \overline{\mu}_r(x)\| \le \|\mu_i\| + \|\overline{\mu}_r(x)\| \le 2R$, we obtain

$$\begin{aligned}\|\partial_r \overline{\mu}_r(x)\| &\le \sum_{i=1}^m \omega_i(r,x)\big\|\mu_i - \overline{\mu}_r(x)\big\| \cdot |\partial_r \ell_i(r,x)| \\ &\le 2C_1 R^2\|x\| + 2C_2 R^3.\end{aligned}$$

All together, we obtain

$$\begin{aligned}\big\|\partial_r \nabla \log p_r(x)\big\| &\le \frac{\|v_r' x\|}{v_r^2} + \left|\left(\tfrac{a_r}{v_r}\right)'\right| \cdot \big\|\overline{\mu}_r(x)\big\| + \left|\tfrac{a_r}{v_r}\right| \cdot \big\|\partial_r \overline{\mu}_r(x)\big\| \\ &\le \tilde{L}(1+\|x\|),\end{aligned} \tag{179}$$

where $\tilde{L} := \max\left\{\frac{2|\sigma^2-1|}{v^2} + \frac{2C_1R^2}{v},\, C_1 R + \frac{2C_2R^3}{v}\right\}$, is independent of d. Therefore, for any $t_n \le t \le s \le t_{n+1}$,

$$\begin{aligned}\big\|\nabla \log p_{T-t}(x) - \nabla \log p_{T-s}(x)\big\| &\le \int_{T-s}^{T-t} \big\|\partial_r \nabla \log p_r(x)\big\|\,\mathrm{d}r \\ &\le \tilde{L}h(1+\|x\|).\end{aligned} \tag{180}$$

This concludes the proof. □

F Proof of Corollary 3.10

Proof. Given an error tolerance $\epsilon > 0$, Theorem 3.9 tells that, for T being large enough, h being small enough and ϵ_0 being small enough such that

$$T^3\big(\tfrac{\sqrt{d}h}{k} + kdh^2 + dh^2 + \sqrt{d}h^{3/2}\big) \le \tfrac{\epsilon}{3}, \quad \mathrm{e}^{-T}\mathcal{W}_2(\mu,\gamma^d) \le \tfrac{\epsilon}{3}, \quad \text{and} \quad \epsilon_0 T \le \tfrac{\epsilon}{3}, \tag{181}$$

one can arrive at

$$\mathcal{W}_2\left(\mu, \mathcal{L}\left(Y_{t_N}\right)\right) \le \epsilon.$$

According to Lemma B.2, we have $\mathcal{W}_2(\mu,\gamma^d) \le \sqrt{(\mathcal{M}+1)d}$. Then, the second and third part of inequality (181) requires

$$T \ge \log\big(\tfrac{3\sqrt{(\mathcal{M}+1)d}}{\epsilon}\big)\big), \quad \epsilon_0 \le \tfrac{\epsilon}{3T} \le \frac{\epsilon}{3\log\left(3\epsilon^{-1}\sqrt{(\mathcal{M}+1)d}\right)}.$$

Next, we deal with the first term. We first choose the batch size k to be of order $\tilde{\mathcal{O}}(d^{1/2}h/\epsilon)$ to ensure

$$Cd^{1/2}\,\tfrac{h}{k} \le \tfrac{\epsilon}{12}.$$

Then, based on the choice of k, one has

$$Cd^{3/2}h^3 \le \tfrac{\epsilon^2}{12}, \quad Cdh^2 \le \tfrac{\epsilon}{12}, \quad \text{and} \quad C\sqrt{d}h^{3/2} \le \tfrac{\epsilon}{12}, \tag{182}$$

which ensure that the first inequality in (181) holds. Rearranging (182) gives

$$h \le C\min\left\{\epsilon^{2/3}/d^{1/2}, \epsilon^{1/2}/d^{1/2}, \epsilon^{2/3}/d^{1/3}\right\}.$$

This completes the proof. □

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction clearly state the main contributions of the paper. Please see Subsection 1.2 for the contributions of the paper.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: In the last section (Section 5), we highlight several strong assumptions made in this work, along with potential directions for relaxing these assumptions and generalizing our current findings. A key limitation of our results is the exponential dependence on α, M and L_V. Obtaining a polynomial dependence on these parameters is an important direction for future work.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes],

Justification: The assumptions are clearly presented in the main paper. For each theoretical result, the paper explicitly states all necessary assumptions and provides complete and correct proofs in Appendix. We also give an overview of proof in Subsection 4.2.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [N/A]

Justification: This paper establishes enhanced convergence guarantees in $\mathcal{W}_2$-distance for two existing SGMs (Euler-Maruyama method and Poisson midpoint method) under conditions beyond log-concavity. The main contribution is through theoretical analyses and proofs. We refer the reader to [47, 23] for the corresponding experimental results.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [N/A]

   Justification: This paper establishes enhanced convergence guarantees in $\mathcal{W}_2$-distance for two existing SGMs (Euler-Maruyama method and Poisson midpoint method) under conditions beyond log-concavity. The main contribution is through theoretical analyses and proofs. We refer the reader to [47, 23] for the corresponding experimental results.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [N/A]

   Justification: This paper establishes enhanced convergence guarantees in $\mathcal{W}_2$-distance for two existing SGMs (Euler-Maruyama method and Poisson midpoint method) under conditions beyond log-concavity. The main contribution is through theoretical analyses and proofs. We refer the reader to [47, 23] for the corresponding experimental results.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [N/A]

Justification: This paper establishes enhanced convergence guarantees in $\mathcal{W}_2$-distance for two existing SGMs (Euler-Maruyama method and Poisson midpoint method) under conditions beyond log-concavity. The main contribution is through theoretical analyses and proofs. We refer the reader to [47, 23] for the corresponding experimental results.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [N/A]

   Justification: This paper establishes enhanced convergence guarantees in $\mathcal{W}_2$-distance for two existing SGMs (Euler-Maruyama method and Poisson midpoint method) under conditions beyond log-concavity. The main contribution is through theoretical analyses and proofs. We refer the reader to [47, 23] for the corresponding experimental results.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research fully complies with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.

- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: The paper focuses on theoretical findings of machine learning. There is no societal impact of the work performed.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper poses no such risks.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [N/A]

Justification: The paper does not use existing assets.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [N/A]

Justification: The paper does not release new assets.

Guidelines:

- The answer [N/A] means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.