A Discussion of Theoretical Advances in Accelerating Samplers for Diffusion Models

Score-based diffusion models can be formulated using either SDEs or their deterministic counterparts, known as probability flow ODEs [36]. While SDE-based samplers generate samples through stochastic simulation, ODE-based samplers provide a deterministic alternative. Theoretical advancements in accelerating these samplers have emerged only recently. A significant step toward designing provably accelerated, training-free methods were made by [24], who propose and analyze acceleration for both ODE- and SDE-based samplers. Their accelerated SDE sampler leverages higher-order expansions of the conditional density to enhance efficiency. This was followed by the work of [25], which provided convergence guarantees for probability flow ODEs. Furthermore, [19] studies the convergence properties of deterministic samplers based on probability flow ODEs, using the Runge-Kutta integrator; [41] propose and analyze a training-free acceleration algorithm for SDE-based samplers, based on the stochastic Runge-Kutta method. [27] proposes a novel accelerated SDE-based sampler when Hessian information is available. Another line of work involves the midpoint randomized method. In particular, [15] explore ODE acceleration by incorporating a randomized midpoint method, leveraging its advantages in parallel computation. A more recent work by [23] improved upon the ODE sampler proposed by [15], achieving the state-of-the-art convergence rate.

We note that all of these works provide convergence analysis in terms of either KL divergence or TV distance. Among these, [27] accelerates the stochastic DDPM sampler by leveraging precise score and Hessian estimations of the log density, even for possibly non-smooth target distributions. This is achieved through a novel Bayesian approach based on tilting factor representation and Tweedie's formula. [19] accelerates the ODE sampler by utilizing p-th ($p \geqslant 1$) order information of the score function, with the target distribution supported on a compact set and employing early stopping. These two works are the most similar to our proposed accelerated sampler in that they all rely on the Hessian information of the log density. However, their settings differ from ours, and their convergence analyses are neither directly applicable to our framework nor precisely expressed in terms of Wasserstein distance[4].

B Proof of Section 3

We define several stochastic processes associated with the backward process $X_t^{\leftarrow}$ and the sample path ϑ_n. First, recall that $X_t^{\leftarrow}$ is described by the following SDE:

$$\mathrm{d}X_t^{\leftarrow} = \left(\frac{1}{2}X_t^{\leftarrow} + \nabla \log p_t(X_t^{\leftarrow})\right) \mathrm{d}t + \mathrm{d}W_t, \quad X_0^{\leftarrow} \sim p_T$$

and $\{\vartheta_n^{\alpha}, 0 \leqslant n \leqslant N\}$ satisfies the iterative law:

$$\vartheta_{n+1}^{\alpha} = \mathcal{G}_h^{\alpha}(\vartheta_n^{\alpha}, \{W_t\}_{nh \leqslant t \leqslant (n+1)h}),$$

where $\alpha \in \{\mathsf{EM}, \mathsf{EI}, \mathsf{REM}, \mathsf{REI}, \mathsf{SO}\}$.

Based on $X_t^{\leftarrow}$, we define the following two processes, $\{Y_t, 0 \leqslant t \leqslant T\}$ and $\{\tilde{Y}_t, 0 \leqslant t \leqslant h\}$. Y_t satisfies the SDE

$$\mathrm{d}Y_t = \left(\frac{1}{2}Y_t + \nabla \log p_{T-t}(Y_t)\right) \mathrm{d}t + \mathrm{d}W_t, \quad Y_0 \sim \hat{p}_T.$$

$\tilde{Y}_t$ actually relies on $X_t^{\leftarrow}$ on the time interval $[nh, (n+1)h]$ for each n. However, we only need this notation in the proof of one-step discretization error, then we allow for some slight abuse of notation by omitting n, since it will not lead to any confusion. Therefore, $\{\tilde{Y}_t, 0 \leqslant t \leqslant h\}$ satisfies

$$\mathrm{d}\tilde{Y}_t = \left(\frac{1}{2}\tilde{Y}_t + \nabla \log p_{T-t}(\tilde{Y}_t)\right) \mathrm{d}t + \mathrm{d}W_t, \quad \tilde{Y}_0 = \vartheta_n. \tag{12}$$

[4] When the target distribution is compactly supported, Pinsker's inequality allows translating TV or KL divergence into Wasserstein distance. However, this often yields loose bounds, especially in high dimensions, where the actual Wasserstein distance may be much smaller.

Recall that we have defined two processes $\vartheta^{\mathsf{REM}}_{n+u}$ and $\vartheta^{\mathsf{REI}}_{n+u}$ in Section 3.2, that is, for any $u \in [0,1]$ and $n = 0, \cdots, N$,

$$\vartheta^{\mathsf{REM}}_{n+u} := (1 + \frac{uh}{2})\vartheta^{\mathsf{REM}}_{n} + uhs_*(T - nh, \vartheta^{\mathsf{REM}}_{n}) + \sqrt{uh}\xi_n \,,$$
$$\vartheta^{\mathsf{REI}}_{n+u} := e^{uh/2}\vartheta^{\mathsf{REI}}_{n} + 2(e^{uh/2} - 1)s_*(T - nh, \vartheta^{\mathsf{REI}}_{n}) + \sqrt{e^{uh} - 1}\xi'_n \,.$$

This section is devoted to proving the convergence rate of the diffusion model under various discretization schemes. To this end, we need the following auxiliary lemmas.

Lemma 6 (Lemma 9 in [13]). *Suppose that Assumption 1 holds. Then, $\nabla \log p_t(x)$ is $L(t)$-Lipschitz, where $L(t)$ is given by*

$$L(t) = \min\{(1 - e^{-t})^{-1}, e^t L_0\} = \begin{cases} e^t L_0 & \text{if } t \leqslant \log(1 + \frac{1}{L_0}) \\ (1 - e^{-t})^{-1} & \text{if } t > \log(1 + \frac{1}{L_0}) \end{cases} .$$

Therefore,

$$L(t) \leqslant L_0 + 1 \,.$$

Lemma 7 (Proposition 7 in [13]). *Suppose that Assumption 1 holds. Then, $\nabla \log p_t(x)$ is $m(t)$-strongly log-concave, where $m(t)$ is given by*

$$m(t) = \frac{1}{e^{-t}/m_0 + (1 - e^{-t})} \,.$$

Therefore,

$$m(t) \geqslant \min\{1, m_0\} \,.$$

Combining these two lemmas, we conclude that the Hessian matrix of $\log p_t$ satisfies the following condition

$$-L(t)I_d \preccurlyeq \nabla^2 \log p_t(\cdot) \preccurlyeq -m(t)I_d.$$

We will frequently use Grönwall's inequality in the proof. Below, we present a specialized form tailored to the relevant processes.

Lemma 8. *Suppose the Assumption 1 holds, consider two stochastic processes H_t and G_t defined on the time interval $[t_1, t_2]$, if they satisfy the same SDE, especially motivated by the same Brownian motion, which means that*

$$\mathrm{d}H_t = \Big(\frac{1}{2}H_t + \nabla \log p_{T-t}(H_t)\Big)\,\mathrm{d}t + \mathrm{d}W_t,$$
$$\mathrm{d}G_t = \Big(\frac{1}{2}G_t + \nabla \log p_{T-t}(G_t)\Big)\,\mathrm{d}t + \mathrm{d}W_t.$$

then for each $t \in [t_1, t_2]$,

$$\|H_t - G_t\|_{\mathbb{L}_2} \leqslant e^{-\int_{t_1}^{t}(m(T-s)-\frac{1}{2})\,\mathrm{d}s}\,\|H_{t_1} - G_{t_1}\|_{\mathbb{L}_2} \,.$$

Applying Lemma 8 to different processes and time intervals, we derive the following inequalities essential for our proof.

$$\Big\|Y_{nh+t} - \tilde{Y}_t\Big\|_{\mathbb{L}_2} \leqslant e^{-\int_{nh}^{nh+t}(m(T-s)-\frac{1}{2})\,\mathrm{d}s}\,\Big\|Y_{nh} - \tilde{Y}_0\Big\|_{\mathbb{L}_2}, \quad \forall t \in [0,h]. \tag{13}$$

$$\|Y_t - X^{\leftarrow}_t\|_{\mathbb{L}_2} \leqslant e^{-\int_0^t(m(T-s)-\frac{1}{2})\,\mathrm{d}s}\,\|Y_0 - X^{\leftarrow}_0\|_{\mathbb{L}_2}, \quad \forall t \in [0,T]. \tag{14}$$

which follow from the fact that $\{Y_t, 0 \leqslant t \leqslant T\}$, $\{\tilde{Y}_t, 0 \leqslant t \leqslant h\}$, $\{X^{\leftarrow}_t, 0 \leqslant t \leqslant T\}$ all satisfy the same SDE as in Lemma 8, by applying a time-shifting operator to $\tilde{Y}_t$.

B.1 Proof of Theorem 1: Part I

In this part, we provide the first part of the proof of Theorem 1, with respect to Euler-Maruyama method. To achieve this, we first prove the one-step discretization error in the following proposition.

Proposition 9. *Suppose that Assumption 1, 2 and 3 are satisfied. Then, the following two claims hold.*

(1) *Firstly, it holds that*

$$
\begin{aligned}
\left\|\tilde{Y}_h - \vartheta^{\mathsf{EM}}_{n+1}\right\|_{\mathbb{L}_2} \leqslant & h^2(C_1(n)^2 + M_1)\left\|Y_{nh} - \vartheta^{\mathsf{EM}}_n\right\|_{\mathbb{L}_2} \\
& + h^2\left[C_1(n)\left(C_1(n)C_2(n) + \frac{1}{2}C_4 + C_3(n)\right) + M_1\left(1 + C_2(n) + C_4\right)\right] \\
& + h^{3/2}\sqrt{d}C_1(n) \\
& + h\varepsilon_{sc},
\end{aligned}
$$

where

$$
\begin{aligned}
C_1(n) &= \frac{1}{2} + \frac{1}{h}\int_{nh}^{(n+1)h} L(T-t)\,\mathrm{d}t, \\
C_2(n) &= e^{-\int_0^{nh}(m(T-t)-\frac{1}{2})\,\mathrm{d}t}\left\|Y_0 - X_T\right\|_{\mathbb{L}_2}, \\
C_3(n) &= \frac{1}{h}\int_{nh}^{(n+1)h}(dL(T-t))^{1/2}\,\mathrm{d}t, \\
C_4 &= \sup_{0\leqslant t\leqslant T}\left\|X_t\right\|_{\mathbb{L}_2}.
\end{aligned}
$$

(2) *As a result,*

$$
\left\|Y_{(n+1)h} - \vartheta^{\mathsf{EM}}_{n+1}\right\|_{\mathbb{L}_2} \leqslant r^{\mathsf{EM}}_n\left\|Y_{nh} - \vartheta^{\mathsf{EM}}_n\right\|_{\mathbb{L}_2} + h^2 C^{\mathsf{EM}}_n + h^{3/2}\sqrt{d}C_1(n) + h\varepsilon_{sc},
$$

where

$$
\begin{aligned}
r^{\mathsf{EM}}_n &= e^{-\int_{nh}^{(n+1)h}(m(T-t)-\frac{1}{2})\,\mathrm{d}t} + h^2(C_1(n)^2 + M_1), \\
C^{\mathsf{EM}}_n &= C_1(n)\left(C_1(n)C_2(n) + \frac{1}{2}C_4 + C_3(n)\right) + M_1\left(1 + C_2(n) + C_4\right).
\end{aligned}
$$

Proof. We prove the two claims sequentially.
Proof of Claim (1). Rewrite display (12) in the integral form,

$$
\tilde{Y}_h = \tilde{Y}_0 + \int_0^h\left(\frac{1}{2}\tilde{Y}_t + \nabla\log p_{T-nh-t}(\tilde{Y}_t)\right)\mathrm{d}t + \int_{nh}^{(n+1)h}\mathrm{d}W_t.
$$

For Euler-Maruyama method, we can write $\vartheta^{\mathsf{EM}}_{n+1}$ in integral form as follows

$$
\vartheta^{\mathsf{EM}}_{n+1} = \vartheta^{\mathsf{EM}}_n + \int_{nh}^{(n+1)h}\left(\frac{1}{2}\vartheta^{\mathsf{EM}}_n + s_*(T-nh, \vartheta^{\mathsf{EM}}_n)\right)\mathrm{d}t + \int_{nh}^{(n+1)h}\mathrm{d}W_t.
$$

Note that $\tilde{Y}_0 = \vartheta_n^{\mathsf{EM}}$, then, it holds that

$$\begin{aligned}
\left\|\tilde{Y}_h - \vartheta_{n+1}^{\mathsf{EM}}\right\|_{\mathbb{L}_2} &= \left\|\frac{1}{2}\int_0^h (\tilde{Y}_t - \vartheta_n^{\mathsf{EM}})\,\mathrm{d}t + \int_0^h (\nabla \log p_{T-nh-t}(\tilde{Y}_t) - s_*(T-nh, \vartheta_n^{\mathsf{EM}}))\,\mathrm{d}t\right\|_{\mathbb{L}_2} \\
&\leqslant \underbrace{\left\|\frac{1}{2}\int_0^h (\tilde{Y}_t - \tilde{Y}_0)\,\mathrm{d}t + \int_0^h (\nabla \log p_{T-nh-t}(\tilde{Y}_t) - \nabla \log p_{T-nh-t}(\tilde{Y}_0)\,\mathrm{d}t\right\|_{\mathbb{L}_2}}_{\mathrm{I}} \\
&\quad + \underbrace{\left\|\int_0^h (\nabla \log p_{T-nh-t}(\vartheta_n^{\mathsf{EM}}) - \nabla \log p_{T-nh}(\vartheta_n^{\mathsf{EM}}))\,\mathrm{d}t\right\|_{\mathbb{L}_2}}_{\mathrm{II}} \\
&\quad + \underbrace{\left\|\int_0^h (\nabla \log p_{T-nh}(\vartheta_n^{\mathsf{EM}}) - s_*(T-nh, \vartheta_n^{\mathsf{EM}}))\,\mathrm{d}t\right\|_{\mathbb{L}_2}}_{\mathrm{III}}.
\end{aligned} \tag{15}$$

Here, we decompose the term $\left\|\tilde{Y}_h - \vartheta_{n+1}^{\mathsf{EM}}\right\|_{\mathbb{L}_2}$ into a sum of three terms and then control each term individually.

For the term I of inequality (15), by Assumption 1 and Lipschitzness of $\nabla \log p_t$, we obtain

$$\begin{aligned}
\mathrm{I} &= \left\|\frac{1}{2}\int_0^h (\tilde{Y}_t - \tilde{Y}_0)\,\mathrm{d}t + \int_0^h (\nabla \log p_{T-nh-t}(\tilde{Y}_t) - \nabla \log p_{T-nh-t}(\tilde{Y}_0))\,\mathrm{d}t\right\|_{\mathbb{L}_2} \\
&\leqslant \frac{1}{2}\int_0^h \left\|\tilde{Y}_t - \tilde{Y}_0\right\|_{\mathbb{L}_2}\,\mathrm{d}t + \int_0^h L(T-nh-t)\left\|\tilde{Y}_t - \tilde{Y}_0\right\|_{\mathbb{L}_2}\,\mathrm{d}t \\
&\leqslant \left(\frac{1}{2}h + \int_{nh}^{(n+1)h} L(T-t)\,\mathrm{d}t\right) \sup_{0\leqslant t\leqslant h}\left\|\tilde{Y}_t - \tilde{Y}_0\right\|_{\mathbb{L}_2}.
\end{aligned}$$

We then proceed to derive the upper bound for the term $\sup_{0\leqslant t\leqslant h}\left\|\tilde{Y}_t - \tilde{Y}_0\right\|_{\mathbb{L}_2}$.

Lemma 10. *When p_0 satisfies Assumption 1, it holds that*

$$\begin{aligned}
\sup_{0\leqslant t\leqslant h}\left\|\tilde{Y}_t - \tilde{Y}_0\right\|_{\mathbb{L}_2} &\leqslant \left(\frac{1}{2}h + \int_{nh}^{(n+1)h} L(T-t)\,\mathrm{d}t\right)\left\|Y_{nh} - \vartheta_n^{\mathsf{EM}}\right\|_{\mathbb{L}_2} \\
&\quad + \left(\frac{1}{2}h + \int_{nh}^{(n+1)h} L(T-t)\,\mathrm{d}t\right) e^{-\int_0^{nh}(m(T-t)-\frac{1}{2})\,\mathrm{d}t}\left\|Y_0 - X_T\right\|_{\mathbb{L}_2} \\
&\quad + \frac{1}{2}h \sup_{0\leqslant t\leqslant T}\left\|X_t\right\|_{\mathbb{L}_2} + \int_{nh}^{(n+1)h} (dL(T-t))^{1/2}\,\mathrm{d}t + \sqrt{dh}.
\end{aligned}$$

Notice that we have no initial limit on the $\tilde{Y}_t$ in Lemma 10, which means that we can use this lemma to any discretization scheme.

For the term II of (15), we first rely on Assumption 2 to derive

$$\begin{aligned}
\mathrm{II} &= \left\|\int_0^h (\nabla \log p_{T-nh-t}(\vartheta_n^{\mathsf{EM}}) - \nabla \log p_{T-nh}(\vartheta_n^{\mathsf{EM}}))\,\mathrm{d}t\right\|_{\mathbb{L}_2} \\
&\leqslant \int_0^h \left\|\nabla \log p_{T-nh-t}(\vartheta_n^{\mathsf{EM}}) - \nabla \log p_{T-nh}(\vartheta_n^{\mathsf{EM}})\right\|_{\mathbb{L}_2}\,\mathrm{d}t \\
&\leqslant h^2 M_1(1 + \left\|\vartheta_n^{\mathsf{EM}}\right\|_{\mathbb{L}_2}).
\end{aligned}$$

Using the triangle inequality and (14), we obtain

$$\begin{aligned}
\left\|\vartheta_n^{\mathsf{EM}}\right\|_{\mathbb{L}_2} &\leqslant \left\|Y_{nh}-\vartheta_n^{\mathsf{EM}}\right\|_{\mathbb{L}_2} + \|Y_{nh}-X_{nh}^{\leftarrow}\|_{\mathbb{L}_2} + \|X_{nh}^{\leftarrow}\|_{\mathbb{L}_2} \\
&\leqslant \left\|Y_{nh}-\vartheta_n^{\mathsf{EM}}\right\|_{\mathbb{L}_2} + e^{-\int_0^{nh}(m(T-t)-\frac{1}{2})\,\mathrm{d}t}\|Y_0-X_T\|_{\mathbb{L}_2} + \sup_{0\leqslant t\leqslant T}\|X_t\|_{\mathbb{L}_2}
\end{aligned} \tag{16}$$

For the term III of (15), it follows from Assumption 3 that

$$\begin{aligned}
\mathrm{III} &= \left\|\int_0^h \left(\nabla \log p_{T-nh}(\vartheta_n^{\mathsf{EM}}) - s_*(T-nh,\vartheta_n^{\mathsf{EM}})\right)\mathrm{d}t\right\|_{\mathbb{L}_2} \\
&\leqslant \int_0^h \left\|\nabla \log p_{T-nh}(\vartheta_n^{\mathsf{EM}}) - s_*(T-nh,\vartheta_n^{\mathsf{EM}})\right\|_{\mathbb{L}_2}\mathrm{d}t \\
&\leqslant h\varepsilon_{sc}.
\end{aligned}$$

Combining these terms above, we obtain that

$$\begin{aligned}
\left\|\tilde{Y}_h-\vartheta_{n+1}^{\mathsf{EM}}\right\|_{\mathbb{L}_2} \leqslant & h^2(C_1(n)^2+M_1)\left\|Y_{nh}-\vartheta_n^{\mathsf{EM}}\right\|_{\mathbb{L}_2} \\
&+ h^2\left[C_1(n)\left(C_1(n)C_2(n)+\frac{1}{2}C_4+C_3(n)\right)+M_1\left(1+C_2(n)+C_4\right)\right] \\
&+ h^{3/2}\sqrt{d}C_1(n) \\
&+ h\varepsilon_{sc},
\end{aligned} \tag{17}$$

where

$$\begin{aligned}
C_1(n) &= \frac{1}{2}+\frac{1}{h}\int_{nh}^{(n+1)h} L(T-t)\,\mathrm{d}t, \\
C_2(n) &= e^{-\int_0^{nh}(m(T-t)-\frac{1}{2})\,\mathrm{d}t}\|Y_0-X_T\|_{\mathbb{L}_2}, \\
C_3(n) &= \frac{1}{h}\int_{nh}^{(n+1)h}(dL(T-t))^{1/2}\,\mathrm{d}t, \\
C_4 &= \sup_{0\leqslant t\leqslant T}\|X_t\|_{\mathbb{L}_2}.
\end{aligned}$$

This completes the proof of Claim **(1)**.

Proof of Claim (2). By the triangle inequality, we have

$$\left\|Y_{(n+1)h}-\vartheta_{n+1}^{\mathsf{EM}}\right\|_{\mathbb{L}_2} \leqslant \left\|Y_{(n+1)h}-\tilde{Y}_h\right\|_{\mathbb{L}_2} + \left\|\tilde{Y}_h-\vartheta_{n+1}^{\mathsf{EM}}\right\|_{\mathbb{L}_2}. \tag{18}$$

Applying (13) to the first term of (18), we obtain that

$$\left\|Y_{(n+1)h}-\tilde{Y}_h\right\|^2 \leqslant e^{-\int_{nh}^{(n+1)h}(2m(T-t)-1)\,\mathrm{d}t}\left\|Y_{nh}-\tilde{Y}_0\right\|^2. \tag{19}$$

Notice that $\tilde{Y}_0=\vartheta_n^{\mathsf{EM}}$, it then follows that

$$\left\|Y_{(n+1)h}-\tilde{Y}_h\right\|_{\mathbb{L}_2} \leqslant e^{-\int_{nh}^{(n+1)h}(m(T-t)-\frac{1}{2})\,\mathrm{d}t}\left\|Y_{nh}-\vartheta_n^{\mathsf{EM}}\right\|_{\mathbb{L}_2}.$$

Claim **(2)** follows directly from our previous results and Claim **(1)**. Since this step is independent of the discretization method, it applies to all the schemes discussed in this section. In the following analysis, we omit this step and proceed directly with the proof of the first claim. □

We now proceed to derive the upper bound of the Wasserstein distance between the sample distribution generated after N iterations and the target distribution p_0, based on the one-step discretization error bound given by Proposition 9.

First, note that

$$\mathsf{W}_2(\mathcal{L}(\vartheta_N^{\mathsf{EM}}),p_0) \leqslant \left\|\vartheta_N^{\mathsf{EM}}-X_0\right\|_{\mathbb{L}_2} \leqslant \left\|Y_{Nh}-\vartheta_N^{\mathsf{EM}}\right\|_{\mathbb{L}_2} + \|Y_{Nh}-X_0\|_{\mathbb{L}_2}.$$

Invoking Proposition 7 of [13], we have

$$\|Y_{Nh} - X_0\|_{\mathbb{L}_2} \leqslant e^{-\int_0^T m(t)\,\mathrm{d}t} \|X_0\|_{\mathbb{L}_2}. \tag{20}$$

According to Proposition 9, by induction, we obtain that

$$\begin{aligned}
\left\|Y_{Nh} - \vartheta_N^{\mathsf{EM}}\right\|_{\mathbb{L}_2} \leqslant & r_{N-1}^{\mathsf{EM}} \left\|Y_{(N-1)h} - \vartheta_{N-1}^{\mathsf{EM}}\right\|_{\mathbb{L}_2} + \left(h^2 C_{N-1}^{\mathsf{EM}} + h^{3/2}\sqrt{d} C_1(N-1) + h\varepsilon_{sc}\right) \\
\leqslant & \left(\prod_{j=0}^{N-1} r_j^{\mathsf{EM}}\right) \left\|Y_0 - \vartheta_0^{\mathsf{EM}}\right\|_{\mathbb{L}_2} + \sum_{k=0}^{N-1}\left(\prod_{j=k+1}^{N-1} r_j^{\mathsf{EM}}\right)\left(h^2 C_k^{\mathsf{EM}} + h^{3/2}\sqrt{d} C_1(k) + h\varepsilon_{sc}\right) \\
= & \sum_{k=0}^{N-1}\left(\prod_{j=k+1}^{N-1} r_j^{\mathsf{EM}}\right)\left(h^2 C_k^{\mathsf{EM}} + h^{3/2}\sqrt{d} C_1(k) + h\varepsilon_{sc}\right),
\end{aligned} \tag{21}$$

where we define $\prod_{j=N}^{N-1} r_j^{\mathsf{EM}} = 1$. Notice that

$$\begin{aligned}
\prod_{j=k+1}^{N-1} r_j^{\mathsf{EM}} = & \prod_{j=k+1}^{N-1} \left(e^{-\int_{jh}^{(j+1)h}(m(T-t)-\frac{1}{2})\,\mathrm{d}t} + h^2(C_1(k)^2 + M_1)\right) \\
\lesssim & \prod_{j=k+1}^{N-1} e^{-h(m_{\min}-\frac{1}{2})} = e^{-h(m_{\min}-\frac{1}{2})(N-k-1)}.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\left\|Y_{Nh} - \vartheta_N^{\mathsf{EM}}\right\|_{\mathbb{L}_2} \lesssim & \sum_{k=0}^{N-1} e^{-h(m_{\min}-\frac{1}{2})(N-k-1)} \left(h^2 C_k^{\mathsf{EM}} + h^{3/2}\sqrt{d} C_1(k) + h\varepsilon_{sc}\right) \\
\leqslant & \frac{1}{1-e^{-h(m_{\min}-\frac{1}{2})}} \left(h^2 \max_{0\leqslant k\leqslant N-1} C_k^{\mathsf{EM}} + h^{3/2}\sqrt{d} \max_{0\leqslant k\leqslant N-1} C_1(k) + h\varepsilon_{sc}\right) \\
\lesssim & \frac{1}{m_{\min}-1/2}\left(\sqrt{dh} \max_{0\leqslant k\leqslant N-1} C_1(k) + \varepsilon_{sc}\right).
\end{aligned} \tag{22}$$

Recall the definition of $C_1(k)$ and the upper bound of $L(t)$, it follows that

$$\max_{0\leqslant k\leqslant N-1} C_1(k) \leqslant \frac{1}{2} + L_{\max},$$

and thus we obtain that

$$\left\|Y_{Nh} - \vartheta_N^{\mathsf{EM}}\right\|_{\mathbb{L}_2} \lesssim \sqrt{dh}\cdot\frac{L_{\max}+1/2}{m_{\min}-1/2} + \varepsilon_{sc}\cdot\frac{1}{m_{\min}-1/2}.$$

Plugging this back into the previous display then we have

$$\mathsf{W}_2(\mathcal{L}(\vartheta_N^{\mathsf{EM}}), p_0) \lesssim e^{-\int_0^T m(t)\,\mathrm{d}t}\|X_0\|_{\mathbb{L}_2} + \sqrt{dh}\cdot\frac{L_{\max}+1/2}{m_{\min}-1/2} + \varepsilon_{sc}\cdot\frac{1}{m_{\min}-1/2},$$

which completes the first part of the proof for Theorem 1.

B.2 Proof of Theorem 1: Part II

This part aims to prove the Wasserstein convergence result for the Exponential Integrator (EI) scheme. We will prove this theorem using the same method as in Theorem 1. Following this approach, we first establish the one-step discretization error in the proposition below.

Proposition 11. *Suppose that Assumption 1, 3 and 2 hold, then one-step discretization error for Exponential Integrator scheme is obtained from the following two bounds.*

(1) It holds that

$$\begin{aligned}\left\|\tilde{Y}_h - \vartheta^{\mathsf{EI}}_{n+1}\right\|_{\mathbb{L}_2} \leqslant & h^2\left(C_5(n)C_1(n) + M_1\frac{2(e^{h/2}-1)}{h}\right)\left\|Y_{nh} - \vartheta^{\mathsf{EM}}_n\right\|_{\mathbb{L}_2} \\ & + h^2\left[C_5(n)\left(C_1(n)C_2(n) + \frac{1}{2}C_4 + C_3(n)\right) + \frac{2(e^{h/2}-1)}{h}M_1(1 + C_2(n) + C_4)\right] \\ & + h^{3/2}\sqrt{d}C_5(n) \\ & + h\cdot\frac{2(e^{h/2}-1)}{h}\varepsilon_{sc},\end{aligned}$$

where

$$C_5(n) = \frac{1}{h}\int_{nh}^{(n+1)h} e^{\frac{1}{2}((n+1)h-t)}L(T-t)\,\mathrm{d}t \approx C_1(n) - \frac{1}{2}.$$

(2) Therefore, we have the bound for one-step discretization error

$$\left\|Y_{(n+1)h} - \vartheta^{\mathsf{EI}}_{n+1}\right\|_{\mathbb{L}_2} \leqslant r^{\mathsf{EI}}_n\left\|Y_{nh} - \vartheta^{\mathsf{EI}}_n\right\|_{\mathbb{L}_2} + h^2C^{\mathsf{EI}}_n + h^{3/2}\sqrt{d}C_5(n) + h\cdot\frac{2(e^{h/2}-1)}{h}\varepsilon_{sc},$$

where

$$\begin{aligned} r^{\mathsf{EI}}_n =& e^{-\int_{nh}^{(n+1)h}(m(T-t)-\frac{1}{2})\,\mathrm{d}t} + h^2\left(C_5(n)C_1(n) + M_1\frac{2(e^{h/2}-1)}{h}\right), \\ C^{\mathsf{EI}}_n =& C_5(n)\left(C_1(n)C_2(n) + \frac{1}{2}C_4 + C_3(n)\right) + \frac{2(e^{h/2}-1)}{h}M_1(1 + C_2(n) + C_4).\end{aligned}$$

Here, the constants $C_i, i = 1,2,3,4$ are as defined in Proposition 9.

Proof. We prove two claims in succession.

Proof of Claim (1). Consider the process defined in (12), which satisfies the SDE

$$\mathrm{d}\tilde{Y}_t = \left[\frac{1}{2}\tilde{Y}_t + \nabla\log p_{T-nh-t}(\tilde{Y}_t)\right]\,dt + \,\mathrm{d}W_t,$$

Instead of integrating both sides of the SDE, we use Itô's formula to $e^{-\frac{t}{2}}\tilde{Y}_t$, then we have

$$\mathrm{d}(e^{-\frac{t}{2}}\tilde{Y}_t) = -\frac{1}{2}e^{-\frac{t}{2}}\tilde{Y}_t + e^{-\frac{t}{2}}\,\mathrm{d}\tilde{Y}_t = e^{-\frac{t}{2}}\left(\nabla\log p_{T-nh-t}(\tilde{Y}_t)\,\mathrm{d}t + \,\mathrm{d}W_t\right),$$

and we notice that we can write it in an integral form.

$$\tilde{Y}_t = e^{t/2}\tilde{Y}_0 + \int_0^t e^{\frac{1}{2}(t-s)}\nabla\log p_{T-nh-s}(\tilde{Y}_s)\,\mathrm{d}s + \int_{nh}^{nh+t} e^{\frac{1}{2}((n+1)h-s)}\,\mathrm{d}W_s.$$

Then we obtain that

$$\tilde{Y}_h - \vartheta^{\mathsf{EI}}_{n+1} = \int_0^h e^{\frac{1}{2}(h-t)}(\nabla\log p_{T-nh-t}(\tilde{Y}_t) - s_*(T-nh, \vartheta^{\mathsf{EI}}_n))\,\mathrm{d}t.$$

We make decomposition the same as the one in (15), that is

$$\begin{aligned}\nabla\log p_{T-nh-t}(\tilde{Y}_t) - s_*(T-nh,\vartheta^{\mathsf{EI}}_n) = & \nabla\log p_{T-nh-t}(\tilde{Y}_t) - \nabla\log p_{T-nh-t}(\tilde{Y}_0) \\ & + \nabla\log p_{T-nh-t}(\vartheta^{\mathsf{EI}}_n) - \nabla\log p_{T-nh}(\vartheta^{\mathsf{EI}}_n) \\ & + \nabla\log p_{T-nh}(\vartheta^{\mathsf{EI}}_n) - s_*(T-nh,\vartheta^{\mathsf{EI}}_n).\end{aligned}$$

It then follows that

$$\begin{aligned}\left\|\tilde{Y}_h - \vartheta^{\mathsf{EI}}_{n+1}\right\|_{\mathbb{L}_2} & \leqslant \int_0^h e^{\frac{1}{2}(h-t)}\left\|\nabla\log p_{T-nh-t}(\tilde{Y}_t) - s_*(T-nh,\vartheta^{\mathsf{EI}}_n)\right\|_{\mathbb{L}_2}\,\mathrm{d}t \\ & \leqslant \int_0^h e^{\frac{1}{2}(h-t)}\left\|\nabla\log p_{T-nh-t}(\tilde{Y}_t) - \nabla\log p_{T-nh-t}(\tilde{Y}_0)\right\|_{\mathbb{L}_2}\,\mathrm{d}t \\ & \quad + \int_0^h e^{\frac{1}{2}(h-t)}\left\|\nabla\log p_{T-nh-t}(\vartheta^{\mathsf{EI}}_n) - \nabla\log p_{T-nh}(\vartheta^{\mathsf{EI}}_n)\right\|_{\mathbb{L}_2}\,\mathrm{d}t \\ & \quad + \int_0^h e^{\frac{1}{2}(h-t)}\left\|\nabla\log p_{T-nh}(\vartheta^{\mathsf{EI}}_n) - s_*(T-nh,\vartheta^{\mathsf{EI}}_n)\right\|_{\mathbb{L}_2}\,\mathrm{d}t.\end{aligned}$$

Note that apart from the exponential term $e^{\frac{1}{2}(h-t)}$, the derivation of the remaining parts is completely consistent with that of (15), until we encounter the term involving $\vartheta_n^{\mathsf{EI}}$, at which point we obtain

$$\begin{aligned}\left\|\tilde{Y}_h - \vartheta_{n+1}^{\mathsf{EI}}\right\|_{\mathbb{L}_2} \leqslant & \left(\int_0^h e^{\frac{1}{2}(h-t)} L(T-nh-t)\,\mathrm{d}t\right) \sup_{0\leqslant t\leqslant h}\left\|\tilde{Y}_t - \tilde{Y}_0\right\|_{\mathbb{L}_2} \\ & + \left(\int_0^h e^{\frac{1}{2}(h-t)}\,\mathrm{d}t\right) M_1 h\Big(1 + \left\|\vartheta_n^{\mathsf{EI}}\right\|_{\mathbb{L}_2}\Big) \\ & + \left(\int_0^h e^{\frac{1}{2}(h-t)}\,\mathrm{d}t\right)\varepsilon_{sc}.\end{aligned}$$

By Lemma 10, we can bound the first term on the right-hand side of the previous display. Moreover, from (16), $\left\|\vartheta_n^{\mathsf{EI}}\right\|_{\mathbb{L}_2}$ can be bounded similarly. Substituting all coefficients with $C_i(n)$ from Proposition 9, we obtain

$$\begin{aligned}\left\|\tilde{Y}_h - \vartheta_{n+1}^{\mathsf{EI}}\right\|_{\mathbb{L}_2} \leqslant & h^2\cdot C_5(n)\left[C_1(n)\left\|Y_{nh} - \vartheta_n^{\mathsf{EI}}\right\|_{\mathbb{L}_2} + C_1(n)C_2(n) + \frac{1}{2}C_4 + C_3(n)\right] \\ & + h^2\cdot\frac{2(e^{h/2}-1)}{h}M_1\left[1 + \left\|Y_{nh} - \vartheta_n^{\mathsf{EM}}\right\|_{\mathbb{L}_2} + C_2(n) + C_4\right] \\ & + h^{3/2}\sqrt{d}C_5(n) \\ & + h\cdot\frac{2(e^{h/2}-1)}{h}\varepsilon_{sc} \\ = & h^2\left(C_5(n)C_1(n) + M_1\frac{2(e^{h/2}-1)}{h}\right)\left\|Y_{nh} - \vartheta_n^{\mathsf{EM}}\right\|_{\mathbb{L}_2} \\ & + h^2\left[C_5(n)\left(C_1(n)C_2(n) + \frac{1}{2}C_4 + C_3(n)\right) + \frac{2(e^{h/2}-1)}{h}M_1(1 + C_2(n) + C_4)\right] \\ & + h^{3/2}\sqrt{d}C_5(n) \\ & + h\cdot\frac{2(e^{h/2}-1)}{h}\varepsilon_{sc},\end{aligned}$$

where

$$C_5(n) = \frac{1}{h}\int_{nh}^{(n+1)h} e^{\frac{1}{2}((n+1)h-t)} L(T-t)\,\mathrm{d}t \approx C_1(n) - \frac{1}{2}.$$

Proof of Claim (2). The proof is omitted for brevity, as it merely requires incorporating $\left\|Y_{(n+1)h} - \tilde{Y}_h\right\|_{\mathbb{L}_2}$ into the conclusion of Claim **(1)**, following a similar argument as in the proof of Claim **(2)** in Proposition 9.

□

For the second part of the proof for Theorem 1, recall that in the first part, the three key steps (20), (21) and (22) lead to the desired result. We now revisit these steps within the framework of other discretization schemes.

Since (20) is independent of the discretization scheme, we can directly apply it throughout the proofs of Theorems 1, 3 and 4. For (21), we note that the h^2 term in r_j^α is neglected, which results in the same upper bound for $\prod_{j=k+1}^{N-1} r_j^\alpha$ across all discretization schemes.

Given the consistency of these two steps, for the remaining discretization schemes, we can directly derive an analogue of (22) from Claim **(2)**. Therefore, in the subsequent proofs of these theorems, after establishing the corresponding proposition, we proceed directly from an expression similar to (22).

For this theorem, we begin the proof with the following inequality

$$\begin{aligned}\left\|Y_{Nh}-\vartheta_N^{\mathsf{EI}}\right\|_{\mathbb{L}_2} &\lesssim \frac{1}{m_{\min}-1/2}\left(hC_n^{\mathsf{EI}}+h^{1/2}\sqrt{d}\max_{0\leqslant k\leqslant N-1}C_5(k)+\varepsilon_{sc}\right)\\ &\lesssim \frac{1}{m_{\min}-1/2}\left(\sqrt{dh}\max_{0\leqslant k\leqslant N-1}C_5(k)+\varepsilon_{sc}\right)\\ &\leqslant \sqrt{dh}\cdot\frac{L_{\max}}{m_{\min}-1/2}+\varepsilon_{sc}\cdot\frac{1}{m_{\min}-1/2}.\end{aligned}$$

Combining this with the bound of $\|X_0-Y_{Nh}\|_{\mathbb{L}_2}$, we obtain

$$\mathsf{W}_2(\mathcal{L}(\vartheta_N^{\mathsf{EI}}),p_0)\lesssim e^{-m_{\min}T}\|X_0\|_{\mathbb{L}_2}+\sqrt{dh}\cdot\frac{L_{\max}}{m_{\min}-1/2}+\varepsilon_{sc}\cdot\frac{1}{m_{\min}-1/2}$$

as desired.

B.3 Proof of Theorem 3: Part I

In this section, we prove the Wasserstein distance between the generated distribution $\mathcal{L}(\vartheta_N^{\mathsf{REM}})$ and the target distribution. The following proposition is established for the one-step discretization error.

Proposition 12. *Suppose that Assumptions 1, 2 and 4 are satisfied, the following two claims hold.*

(1) *It holds that*

$$\begin{aligned}&\left\|\tilde{Y}_h-\vartheta_{n+1}^{\mathsf{REM}}\right\|_{\mathbb{L}_2}\\ &\leqslant h^2\Bigg\{\left[\int_0^1\int_0^1\left[|u-v|L(T-(n+u)h)\left(\frac{1}{2}+L(T-nh)\right)+M_1\right]^2\mathrm{d}u\,\mathrm{d}v\right]^{1/2}\\ &\qquad+\frac{1}{4\sqrt{3}}L(T-nh)+\frac{1}{8\sqrt{3}}\Bigg\}\left\|Y_{nh}-\vartheta_n^{\mathsf{REM}}\right\|_{\mathbb{L}_2}\\ &\quad+h^2\Bigg\{\Bigg\{\int_0^1\int_0^1\Bigg\{(u-v)\left[\left(\frac{1}{2}+L(T-nh)\right)C_2(n)+\frac{1}{2}C_4+(dL(T-nh))^{1/2}\right]L(T-(n+u)h)\\ &\qquad+M_1(1+C_2(n)+C_4)\Bigg\}^2\mathrm{d}u\,\mathrm{d}v\Bigg\}^{1/2}\\ &\qquad+\frac{1}{8\sqrt{3}}(C_2(n)+C_4)+\frac{1}{4\sqrt{3}}\left(L(T-nh)C_2(n)+(dL(T-nh))^{1/2}\right)\Bigg\}\\ &\quad+h^{3/2}\left\{\sqrt{d}\left[\int_0^1\int_0^1 L(T-(n+u)h)^2|u-v|\,\mathrm{d}u\,\mathrm{d}v\right]^{1/2}+\frac{1}{2\sqrt{3}}\right\}\\ &\quad+2h\varepsilon_{sc}.\end{aligned}$$

(2) *Moreover, it holds that*

$$\left\|Y_{(n+1)h}-\vartheta_{n+1}^{\mathsf{REM}}\right\|\leqslant r_n^{\mathsf{REM}}\|Y_{nh}-\vartheta_n\|_{\mathbb{L}_2}+h^2C_{n,1}^{\mathsf{REM}}+h^{3/2}C_{n,2}^{\mathsf{REM}}+3h\varepsilon_{sc},$$

where

$$\begin{aligned}r_n^{\mathsf{REM}}&=e^{-\int_{nh}^{(n+1)h}(m(T-t)-\frac{1}{2})\,\mathrm{d}t}\\ &\quad+h^2\Bigg\{\left[\int_0^1\int_0^1\left[|u-v|L(T-(n+u)h)\left(\frac{1}{2}+L(T-nh)\right)+M_1\right]^2\mathrm{d}u\,\mathrm{d}v\right]^{1/2}\\ &\qquad+\frac{1}{4\sqrt{3}}L(T-nh)+\frac{1}{8\sqrt{3}}\Bigg\},\end{aligned}$$

$$
\begin{aligned}
C_{n,1}^{\mathsf{REM}} =&\bigg\{\int_0^1\int_0^1\bigg\{(u-v)\bigg[\bigg(\frac{1}{2}+L(T-nh)\bigg)C_2(n)+\frac{1}{2}C_4+(dL(T-nh))^{1/2}\bigg]L(T-(n+u)h) \\
&\qquad\qquad + M_1(1+C_2(n)+C_4)\bigg\}^2\,\mathrm{d}u\,\mathrm{d}v\bigg\}^{1/2} \\
&\quad + \frac{1}{8\sqrt{3}}(C_2(n)+C_4)+\frac{1}{4\sqrt{3}}\left(L(T-nh)C_2(n)+(dL(T-nh))^{1/2}\right) \\
C_{n,2}^{\mathsf{REM}} =&\sqrt{d}\left[\int_0^1\int_0^1 L(T-(n+u)h)^2|u-v|\,\mathrm{d}u\,\mathrm{d}v\right]^{1/2}+\frac{1}{2\sqrt{3}}.
\end{aligned}
$$

Proof of Proposition 12. **Proof of Claim (1).** We make the following decomposition of one-step discretization error

$$
\left\|\tilde{Y}_h-\vartheta_{n+1}^{\mathsf{REM}}\right\|_{\mathbb{L}_2} \leqslant \left\|\tilde{Y}_h-\mathbb{E}_{U_n}\left[\vartheta_{n+1}^{\mathsf{REM}}\right]\right\|_{\mathbb{L}_2}+\left\|\mathbb{E}_{U_n}\left[\vartheta_{n+1}^{\mathsf{REM}}\right]-\vartheta_{n+1}^{\mathsf{REM}}\right\|_{\mathbb{L}_2}. \tag{23}
$$

We first derive the upper bound for the term $\left\|\tilde{Y}_h-\mathbb{E}_{U_n}\left[\vartheta_{n+1}^{\mathsf{REM}}\right]\right\|_{\mathbb{L}_2}$. By the definitions of $\vartheta_n^{\mathsf{REM}}$ and $\tilde{Y}_h$, we have

$$
\begin{aligned}
&\left\|\tilde{Y}_h-\mathbb{E}_{U_n}\left[\vartheta_{n+1}^{\mathsf{REM}}\right]\right\|_{\mathbb{L}_2} \\
=&\left\|\frac{1}{2}\int_0^h\tilde{Y}_t\,\mathrm{d}t+\int_0^h\nabla\log p_{T-nh-t}(\tilde{Y}_t)\,\mathrm{d}t-\frac{1}{2}h\mathbb{E}_{U_n}(\vartheta_{n+U}^{\mathsf{REM}})-h\mathbb{E}_{U_n}[s_*(T-(n+U_n)h,\vartheta_{n+U}^{\mathsf{REM}}])\right\|_{\mathbb{L}_2}.
\end{aligned}
$$

Notice that

$$
\int_0^h\tilde{Y}_t\,\mathrm{d}t=h\mathbb{E}_{U_n}[\tilde{Y}_{U_nh}],\qquad \int_0^h\nabla\log p_{T-nh-t}(\tilde{Y}_t)\,\mathrm{d}t=h\mathbb{E}_{U_n}[\nabla\log p_{T-(n+U_n)h}(\tilde{Y}_{U_nh})].
$$

Plugging this back into the previous display then gives

$$
\begin{aligned}
&\left\|\tilde{Y}_h-\mathbb{E}_{U_n}[\vartheta_{n+1}^{\mathsf{REM}}]\right\|_{\mathbb{L}_2} \\
=&\left\|\frac{1}{2}h\mathbb{E}_{U_n}[\tilde{Y}_{U_nh}]+h\mathbb{E}_{U_n}[\nabla\log p_{T-(n+U_n)h}(\tilde{Y}_{U_nh})]-\frac{1}{2}h\mathbb{E}_{U_n}(\vartheta_{n+U}^{\mathsf{REM}})-h\mathbb{E}_{U_n}(s_*(T-(n+U_n)h,\vartheta_{n+U}^{\mathsf{REM}}))\right\|_{\mathbb{L}_2} \\
\leqslant&\frac{1}{2}h\left\|\mathbb{E}_{U_n}[\tilde{Y}_{U_nh}-\vartheta_{n+U_n}^{\mathsf{REM}}]\right\|_{\mathbb{L}_2}+h\left\|\mathbb{E}_{U_n}[\nabla\log p_{T-(n+U_n)h}(\tilde{Y}_{U_nh})-s_*(T-(n+U_n)h,\vartheta_{n+U_n}^{\mathsf{REM}})]\right\|_{\mathbb{L}_2}.
\end{aligned}
$$

By the definition of $\tilde{Y}_{U_nh}$ and $\vartheta_{n+U_n}^{\mathsf{REM}}$, we have

$$
\begin{aligned}
&\left\|\mathbb{E}_{U_n}[\tilde{Y}_{U_nh}-\vartheta_{n+U_n}^{\mathsf{REM}}]\right\|_{\mathbb{L}_2} \\
=&\left\|\mathbb{E}_{U_n}\left[\frac{1}{2}\int_0^{U_nh}(\tilde{Y}_t-\vartheta_n^{\mathsf{REM}})\,\mathrm{d}t+\int_0^{U_nh}(\nabla\log p_{T-nh-t}(\tilde{Y}_t)-s_*(T-nh,\vartheta_n^{\mathsf{REM}}))\,\mathrm{d}t\right]\right\|_{\mathbb{L}_2} \\
\leqslant&\left\|\mathbb{E}_{U_n}\left[\frac{1}{2}\int_0^{U_nh}\left\|\tilde{Y}_t-\vartheta_n^{\mathsf{REM}}\right\|\,\mathrm{d}t+\int_0^{U_nh}\left\|\nabla\log p_{T-nh-t}(\tilde{Y}_t)-s_*(T-nh,\vartheta_n^{\mathsf{REM}})\right\|\,\mathrm{d}t\right]\right\|_{\mathbb{L}_2} \\
\leqslant&\left\|\mathbb{E}_{U_n}\left[\frac{1}{2}\int_0^{h}\left\|\tilde{Y}_t-\vartheta_n^{\mathsf{REM}}\right\|\,\mathrm{d}t+\int_0^{h}\left\|\nabla\log p_{T-nh-t}(\tilde{Y}_t)-s_*(T-nh,\vartheta_n^{\mathsf{REM}})\right\|\,\mathrm{d}t\right]\right\|_{\mathbb{L}_2} \\
\leqslant&\frac{1}{2}\int_0^h\left\|\tilde{Y}_t-\vartheta_n^{\mathsf{REM}}\right\|_{\mathbb{L}_2}\,\mathrm{d}t+\int_0^h\left\|\nabla\log p_{T-nh-t}(\tilde{Y}_t)-s_*(T-nh,\vartheta_n^{\mathsf{REM}})\right\|_{\mathbb{L}_2}\,\mathrm{d}t.
\end{aligned}
$$

The second inequality arises because the integrand is non-negative, the last inequality follows from the fact that the random variables inside the inner expectation $\mathbb{E}_{U_n}$ are independent of U_n, and thus the inner expectation can be ignored. Then using the same argument as in the proof of Proposition 9, especially adopting the same procedure as the one following (15), we can apply the conclusion of Proposition 9 to the term above, then we obtain that

$$
\begin{aligned}
\left\|\mathbb{E}_{U_n}\left[\tilde{Y}_{U_n h}-\vartheta_{n+U_n}^{\mathsf{REM}}\right]\right\|_{\mathbb{L}_2} \leqslant & \left(\frac{1}{2}h+\int_{nh}^{(n+1)h} L(T-t)\,\mathrm{d}t\right)\sup_{0\leqslant t\leqslant h}\left\|\tilde{Y}_t-\tilde{Y}_0\right\|_{\mathbb{L}_2} \\
& +\int_{nh}^{(n+1)h}\left\|\nabla\log p_{T-nh-t}(\vartheta_n^{\mathsf{REM}})-s_*(T-nh,\vartheta_n^{\mathsf{REM}})\right\|_{\mathbb{L}_2}\mathrm{d}t \\
\leqslant & h^2(C_1(n)^2+M_1)\left\|Y_{nh}-\vartheta_n^{\mathsf{REM}}\right\|_{\mathbb{L}_2} \\
& +h^2\left[C_1(n)\left(C_1(n)C_2(n)+\frac{1}{2}C_4+C_3(n)\right)+M_1(1+C_2(n)+C_4)\right] \\
& +h^{3/2}\sqrt{d}C_1(n) \\
& +h\varepsilon_{sc} \\
\triangleq & h^2 r_1\left\|Y_{nh}-\vartheta_n^{\mathsf{REM}}\right\|_{\mathbb{L}_2}+h^2 r_2+h^{3/2}\sqrt{d}C_1(n)+h\varepsilon_{sc},
\end{aligned}
$$

where

$$
\begin{aligned}
r_1 =& C_1(n)^2+M_1, \\
r_2 =& C_1(n)\left(C_1(n)C_2(n)+\frac{1}{2}C_4+C_3(n)\right)+M_1(1+C_2(n)+C_4).
\end{aligned}
$$

We now derive the upper bound of the second term in (23). Note that

$$
\begin{aligned}
& \left\|\mathbb{E}_{U_n}\left[\nabla\log p_{T-(n+U_n)h}(\tilde{Y}_{(n+U_n)h})-s_*(T-(n+U_n)h,\vartheta_{n+U_n}^{\mathsf{REM}})\right]\right\|_{\mathbb{L}_2} \\
=& \left\|\int_0^1\left(\nabla\log p_{T-(n+u)h}(\tilde{Y}_{(n+u)h})-s_*(T-(n+u)h,\vartheta_{n+u}^{\mathsf{REM}})\right)\mathrm{d}u\right\|_{\mathbb{L}_2} \\
\leqslant & \int_0^1\left\|\nabla\log p_{T-(n+u)h}(\tilde{Y}_{(n+u)h})-s_*(T-(n+u)h,\vartheta_{n+u}^{\mathsf{REM}})\right\|_{\mathbb{L}_2}\mathrm{d}u \\
\leqslant & \int_0^1\Bigg(\left\|\nabla\log p_{T-(n+u)h}(\tilde{Y}_{(n+u)h})-\nabla\log p_{T-(n+u)h}(\vartheta_{n+u}^{\mathsf{REM}})\right\|_{\mathbb{L}_2} \\
& \qquad +\left\|\nabla\log p_{T-(n+u)h}(\vartheta_{n+u}^{\mathsf{REM}})-s_*(T-(n+u)h,\vartheta_{n+u}^{\mathsf{REM}})\right\|_{\mathbb{L}_2}\Bigg)\mathrm{d}u \\
\leqslant & \int_0^1 L(T-(n+u)h)\left\|\tilde{Y}_{(n+u)h}-\vartheta_{n+u}^{\mathsf{REM}}\right\|_{\mathbb{L}_2}\mathrm{d}u+\varepsilon_{sc},
\end{aligned}
\tag{24}
$$

the second inequality follows from the triangle inequality, and the last inequality depends on Assumption 1 and 4. By (17), changing the value of h to uh, we have

$$
\begin{aligned}
& \left\|\tilde{Y}_{(n+u)h}-\vartheta_{n+u}^{\mathsf{REM}}\right\|_{\mathbb{L}_2} \\
\leqslant & (uh)^2(C_{1,n}(u)^2+M_1)\left\|Y_{nh}-\vartheta_n^{\mathsf{REM}}\right\|_{\mathbb{L}_2} \\
& +(uh)^2\left[C_{1,n}(u)\left(C_{1,n}(u)C_2(n)+\frac{1}{2}C_4+C_{3,n}(u)\right)+M_1(1+C_2(n)+C_4)\right] \\
& +(uh)^{3/2}\sqrt{d}C_{1,n}(u) \\
& +uh\varepsilon_{sc},
\end{aligned}
\tag{25}
$$

where $C_{1,n}(u)$ and $C_{3,n}(u)$ is the uh-version of $C_1(n)$ and $C_3(n)$, respectively, that is

$$
\begin{aligned}
C_{1,n}(u) =& \frac{1}{2}+\frac{1}{uh}\int_{nh}^{(n+u)h} L(T-t)\,\mathrm{d}t, \\
C_{3,n}(u) =& \frac{1}{uh}\int_{nh}^{(n+u)h}(dL(T-t))^{1/2}\,\mathrm{d}t,
\end{aligned}
$$

Plugging the previous display (25) back into display (24), then rearranging and simplifying the expression, yields

$$
\begin{aligned}
&\left\|\mathbb{E}_{U_n}[\nabla \log p_{T-(n+U_n)h}(\tilde{Y}_{(n+U_n)h}) - s_*(T-(n+U_n)h, \vartheta^{\mathsf{REM}}_{n+U_n})]\right\|_{\mathbb{L}_2} \\
&\leqslant h^2 \left(\int_0^1 L(T-(n+u)h)u^2(C_{1,n}(u)^2 + M_1)\,\mathrm{d}u\right) \left\|Y_{nh} - \vartheta^{\mathsf{EM}}_n\right\|_{\mathbb{L}_2} \\
&\quad + h^2 \left\{\int_0^1 L(T-(n+u)h)u^2 \left[C_{1,n}(u)\left(C_{1,n}(u)C_2(n) + \frac{1}{2}C_4 + C_{3,n}(u)\right) + M_1(1 + C_2(n) + C_4)\right] \mathrm{d}u\right\} \\
&\quad + h^{3/2}\left(\int_0^1 L(T-(n+u)h)u^{3/2}\,\mathrm{d}u\right)\sqrt{d}C_1(n) \\
&\quad + h\left(\int_0^1 L(T-(n+u)h)u\,\mathrm{d}u\right)\varepsilon_{sc} \\
&\quad + \varepsilon_{sc} \\
&\triangleq h^2 r_3 \left\|Y_{nh} - \vartheta^{\mathsf{EM}}_n\right\|_{\mathbb{L}_2} + h^2 r_4 + h^{3/2} r_5 + h r_6 \varepsilon_{sc} + \varepsilon_{sc},
\end{aligned}
$$

where

$$
\begin{aligned}
r_3 &= \int_0^1 L(T-(n+u)h)u^2(C_{1,n}(u)^2 + M_1)\,\mathrm{d}u, \\
r_4 &= \int_0^1 L(T-(n+u)h)u^2 \left[C_{1,n}(u)\left(C_{1,n}(u)C_2(n) + \frac{1}{2}C_4 + C_{3,n}(u)\right)\right] \mathrm{d}u + M_1(1 + C_2(n) + C_4), \\
r_5 &= \left(\int_0^1 L(T-(n+u)h)u^{3/2}\,\mathrm{d}u\right)\sqrt{d}C_1(n), \\
r_6 &= \int_0^1 L(T-(n+u)h)u\,\mathrm{d}u.
\end{aligned}
$$

From the bounds we have obtained for two terms, it follows that

$$
\begin{aligned}
&\left\|\tilde{Y}_h - \mathbb{E}_{U_n}[\vartheta^{\mathsf{REM}}_{n+1}]\right\|_{\mathbb{L}_2} \\
&\leqslant h^3(\frac{1}{2}r_1 + r_3)\left\|Y_{nh} - \vartheta^{\mathsf{EM}}_n\right\|_{\mathbb{L}_2} + h^3(\frac{1}{2}r_2 + r_4) + h^{5/2}(\frac{1}{2}\sqrt{d}C_1(n) + r_5) + h^2(\frac{1}{2} + r_6)\varepsilon_{sc} + h\varepsilon_{sc}.
\end{aligned} \tag{26}
$$

Considering the second term of one-step discretization error

$$
\begin{aligned}
&\vartheta^{\mathsf{REM}}_{n+1} - \mathbb{E}_{U_n}[\vartheta^{\mathsf{REM}}_{n+1}] \\
&= \frac{1}{2}h\left[\vartheta^{\mathsf{REM}}_{n+U} - \mathbb{E}_{U_n}[\vartheta^{\mathsf{REM}}_{n+U}]\right] + h\left[s_*(T-(n+U_n)h, \vartheta^{\mathsf{REM}}_{n+U}) - \mathbb{E}_{U_n}[s_*(T-(n+U_n)h, \vartheta^{\mathsf{REM}}_{n+U})]\right] \\
&= \frac{1}{2}h\left[\frac{1}{2}h(U_n - \frac{1}{2})\vartheta^{\mathsf{REM}}_n + h(U_n - \frac{1}{2})s_*(T-nh, \vartheta^{\mathsf{REM}}_n)\right] \\
&\quad + \frac{1}{2}h\left[\int_{nh}^{(n+U_n)h} \mathrm{d}W_t - \int_0^1\left(\int_{nh}^{(n+u)h} \mathrm{d}W_t\right)\mathrm{d}u\right] \\
&\quad + h\left[s_*(T-(n+U_n)h, \vartheta^{\mathsf{REM}}_{n+U}) - \mathbb{E}_{U_n}[s_*(T-(n+U_n)h, \vartheta^{\mathsf{REM}}_{n+U})]\right].
\end{aligned} \tag{27}
$$

The second equality follows from the fact that

$$
\begin{aligned}
\mathbb{E}_{U_n}[\vartheta^{\mathsf{REM}}_{n+U}] &= \vartheta^{\mathsf{REM}}_n + \frac{1}{2}h\mathbb{E}_{U_n}[U_n]\vartheta^{\mathsf{REM}}_n + h\mathbb{E}_{U_n}[U_n]s_*(T-nh, \vartheta^{\mathsf{REM}}_n) + \mathbb{E}_{U_n}\int_{nh}^{(n+U_n)h} \mathrm{d}W_t \\
&= \vartheta^{\mathsf{REM}}_n + \frac{1}{4}h\vartheta^{\mathsf{REM}}_n + \frac{1}{2}hs_*(T-nh, \vartheta^{\mathsf{REM}}_n) + \int_0^1\left(\int_{nh}^{(n+u)h} \mathrm{d}W_t\right)\mathrm{d}u,
\end{aligned}
$$

since U_n is independent of $\vartheta^{\mathsf{REM}}_n$.

We proceed to bound each term in (27). For the first term, still notice that the independence between

U_n and $\vartheta_n^{\mathsf{REM}}$, then we find that

$$\begin{aligned}\left\|(U_n-\frac{1}{2})\vartheta_n^{\mathsf{REM}}\right\|_{\mathbb{L}_2} &= \left\{\mathbb{E}\left[\mathbb{E}_{U_n}\left\|(U_n-\frac{1}{2})\vartheta_n^{\mathsf{REM}}\right\|^2\right]\right\}^{1/2} \\ &= \left\{\mathbb{E}\left[\mathbb{E}_{U_n}[(U_n-\frac{1}{2})^2]\cdot\left\|\vartheta_n^{\mathsf{REM}}\right\|^2\right]\right\}^{1/2} \\ &= \left\{\mathbb{E}\left[\frac{1}{12}\left\|\vartheta_n^{\mathsf{REM}}\right\|^2\right]\right\}^{1/2} \\ &= \frac{1}{2\sqrt{3}}\left\|\vartheta_n^{\mathsf{REM}}\right\|_{\mathbb{L}_2}.\end{aligned}$$

The bounding of another part of the first term follows in a similar manner, we obtain that

$$\left\|(U_n-\frac{1}{2})s_*(T-nh,\vartheta_n^{\mathsf{REM}})\right\|_{\mathbb{L}_2} = \frac{1}{2\sqrt{3}}\left\|s_*(T-nh,\vartheta_n^{\mathsf{REM}})\right\|_{\mathbb{L}_2}.$$

For the second term of (27), notice that due to Itô's isometry formula, for any well-defined stochastic process X_t and its Itô stochastic integral $I_t(X)=\int_0^t X_u\,\mathrm{d}M_u$, we have

$$\mathbb{E}[I_t(X)^2]=\mathbb{E}\int_0^t X_u^2\,\mathrm{d}\langle M\rangle_u, \tag{28}$$

then we can establish a lemma.

Lemma 13. *Suppose W_t is a d-dim standard Brownian motion, then*

$$\left\|\int_{nh}^{(n+U_n)h}\mathrm{d}W_t-\int_0^1\left(\int_{nh}^{(n+u)h}\mathrm{d}W_t\right)\mathrm{d}u\right\|_{\mathbb{L}_2}^2 \leqslant \frac{h}{3}.$$

For the third term of (27), we get

$$\begin{aligned}&\left\|s_*(T-(n+U_n)h,\vartheta_{n+U}^{\mathsf{REM}})-\mathbb{E}_{U_n}[s_*(T-(n+U_n)h,\vartheta_{n+U}^{\mathsf{REM}})]\right\|_{\mathbb{L}_2} \\ &= \left\|\int_0^1 s_*(T-(n+U_n)h,\vartheta_{n+U}^{\mathsf{REM}})-s_*(T-(n+v)h,\vartheta_{n+v}^{\mathsf{REM}})\,\mathrm{d}v\right\|_{\mathbb{L}_2} \\ &= \left\{\mathbb{E}\int_0^1\left[\int_0^1 s_*(T-(n+u)h,\vartheta_{n+u}^{\mathsf{REM}})-s_*(T-(n+v)h,y_{(n+v)h})\,\mathrm{d}v\right]^2\mathrm{d}u\right\}^{1/2} \\ &\leqslant \left\{\mathbb{E}\int_0^1\int_0^1\left[s_*(T-(n+u)h,\vartheta_{n+u}^{\mathsf{REM}})-s_*(T-(n+v)h,\vartheta_{n+v}^{\mathsf{REM}})\right]^2\mathrm{d}u\,\mathrm{d}v\right\}^{1/2} \\ &= \left\{\int_0^1\int_0^1\left\|s_*(T-(n+u)h,\vartheta_{n+u}^{\mathsf{REM}})-s_*(T-(n+v)h,\vartheta_{n+v}^{\mathsf{REM}})\right\|_{\mathbb{L}_2}^2\mathrm{d}u\,\mathrm{d}v\right\}^{1/2}.\end{aligned}$$

Then by the triangle inequality and Assumption 4, we have

$$\begin{aligned}&\left\|s_*(T-(n+u)h,\vartheta_{n+u}^{\mathsf{REM}})-s_*(T-(n+v)h,\vartheta_{n+v}^{\mathsf{REM}})\right\|_{\mathbb{L}_2} \\ &\leqslant \left\|s_*(T-(n+u)h,\vartheta_{n+u}^{\mathsf{REM}})-\nabla\log p_{T-(n+u)h}(\vartheta_{n+u}^{\mathsf{REM}})\right\|_{\mathbb{L}_2} \\ &\quad+\left\|s_*(T-(n+v)h,\vartheta_{n+v}^{\mathsf{REM}})-\nabla\log p_{T-(n+v)h}(\vartheta_{n+v}^{\mathsf{REM}})\right\|_{\mathbb{L}_2} \\ &\quad+\left\|\nabla\log p_{T-(n+u)h}(\vartheta_{n+u}^{\mathsf{REM}})-\nabla\log p_{T-(n+v)h}(\vartheta_{n+v}^{\mathsf{REM}})\right\|_{\mathbb{L}_2} \\ &\leqslant 2\varepsilon_{sc}+\left\|\nabla\log p_{T-(n+u)h}(\vartheta_{n+u}^{\mathsf{REM}})-\nabla\log p_{T-(n+v)h}(\vartheta_{n+v}^{\mathsf{REM}})\right\|_{\mathbb{L}_2}.\end{aligned} \tag{29}$$

Combining the three terms of (27) together, we have

$$\begin{aligned}\left\|\vartheta_{n+1}^{\mathsf{REM}}-\mathbb{E}_{U_n}[\vartheta_{n+1}^{\mathsf{REM}}]\right\|_{\mathbb{L}_2} \leqslant{}& \frac{1}{8\sqrt{3}}h^2\left\|\vartheta_n^{\mathsf{REM}}\right\|_{\mathbb{L}_2}+\frac{1}{4\sqrt{3}}h^2\left\|s_*(T-nh,\vartheta_n^{\mathsf{REM}})\right\|_{\mathbb{L}_2}+\frac{1}{2\sqrt{3}}h^{3/2} \\ &+h\left\{\int_0^1\int_0^1\left\|\nabla\log p_{T-(n+v)h}(\vartheta_{n+v}^{\mathsf{REM}})-\nabla\log p_{T-(n+u)h}(\vartheta_{n+v}^{\mathsf{REM}})\right\|_{\mathbb{L}_2}^2\mathrm{d}u\,\mathrm{d}v\right\}^{1/2} \\ &+2h\varepsilon_{sc}.\end{aligned}$$

By applying the same technique used in the proofs of Proposition 9 and Proposition 11, the upper bounds for $\left\|\vartheta_n^{\mathsf{REM}}\right\|_{\mathbb{L}_2}$ and $\left\|s_*(T-nh, \vartheta_n^{\mathsf{REM}})\right\|_{\mathbb{L}_2}$ follows readily. Thus, the proposition follows immediately from the bound on the second last term. We now consider the case for $u > v$, due to Assumptions 1 and 2,

$$\begin{aligned}&\left\|\nabla \log p_{T-(n+u)h}(\vartheta_{n+u}^{\mathsf{REM}}) - \nabla \log p_{T-(n+v)h}(\vartheta_{n+v}^{\mathsf{REM}})\right\|_{\mathbb{L}_2}\\ \leqslant\ & L(T-(n+u)h)\left\|\vartheta_{n+u}^{\mathsf{REM}} - \vartheta_{n+v}^{\mathsf{REM}}\right\|_{\mathbb{L}_2} + M_1 h\Big(1 + \left\|\vartheta_{n+v}^{\mathsf{REM}}\right\|_{\mathbb{L}_2}\Big).\end{aligned}$$

Since

$$\begin{aligned}\left\|\vartheta_{n+u}^{\mathsf{REM}} - \vartheta_{n+v}^{\mathsf{REM}}\right\|_{\mathbb{L}_2} &\leqslant \frac{1}{2}(u-v)h\left\|\vartheta_n^{\mathsf{REM}}\right\|_{\mathbb{L}_2} + (u-v)h\left\|s_*(T-nh, \vartheta_n^{\mathsf{REM}})\right\|_{\mathbb{L}_2} + \left\|\int_{(n+v)h}^{(n+u)h} \mathrm{d}W_t\right\|_{\mathbb{L}_2}\\ &\leqslant \frac{1}{2}(u-v)h\left(\left\|Y_{nh} - \vartheta_n^{\mathsf{REM}}\right\|_{\mathbb{L}_2} + C_2(n) + C_4\right)\\ &\quad + (u-v)h\left[\varepsilon_{sc} + L(T-nh)\left(\left\|Y_{nh} - \vartheta_n^{\mathsf{REM}}\right\|_{\mathbb{L}_2} + C_2(n)\right) + (dL(T-nh))^{1/2}\right]\\ &\quad + \sqrt{(u-v)h}\\ &\leqslant (u-v)h\left[\frac{1}{2} + L(T-nh)\right]\left\|Y_{nh} - \vartheta_n^{\mathsf{REM}}\right\|_{\mathbb{L}_2}\\ &\quad + (u-v)h\left[(\frac{1}{2} + L(T-nh))C_2(n) + \frac{1}{2}C_4 + (dL(T-nh))^{1/2}\right]\\ &\quad + \sqrt{(u-v)dh} + (u-v)h\varepsilon_{sc}.\end{aligned}$$

The second inequality follows from (16), Assumptions 1, 3 and Lemma 18. Similarly,

$$\begin{aligned}\left\|\vartheta_{n+v}^{\mathsf{REM}}\right\|_{\mathbb{L}_2} &\leqslant \left\|\vartheta_{n+v}^{\mathsf{REM}} - \vartheta_n^{\mathsf{REM}}\right\|_{\mathbb{L}_2} + \left\|\vartheta_n^{\mathsf{REM}}\right\|_{\mathbb{L}_2}\\ &\leqslant vh\left[\frac{1}{2} + L(T-nh)\right]\left\|Y_{nh} - \vartheta_n^{\mathsf{REM}}\right\|_{\mathbb{L}_2}\\ &\quad + vh\left[(\frac{1}{2} + L(T-nh))C_2(n) + \frac{1}{2}C_4 + (dL(T-nh))^{1/2}\right]\\ &\quad + \sqrt{vdh} + vh\varepsilon_{sc}\\ &\quad + \left\|Y_{nh} - \vartheta_n^{\mathsf{REM}}\right\|_{\mathbb{L}_2} + C_2(n) + C_4.\end{aligned}$$

Therefore, we obtain that

$$\begin{aligned}&\left\|\nabla \log p_{T-(n+u)h}(\vartheta_{n+u}^{\mathsf{REM}}) - \nabla \log p_{T-(n+v)h}(\vartheta_{n+v}^{\mathsf{REM}})\right\|_{\mathbb{L}_2}\\ \leqslant\ & h\left\{(u-v)\left[\frac{1}{2} + L(T-nh)\right]L(T-(n+u)h) + M_1\right\}\left\|Y_{nh} - \vartheta_n^{\mathsf{REM}}\right\|_{\mathbb{L}_2}\\ &+ h^2 M_1 v\left[\frac{1}{2} + L(T-nh)\right]\left\|Y_{nh} - \vartheta_n^{\mathsf{REM}}\right\|_{\mathbb{L}_2}\\ &+ h^2 v M_1\left[\left(\frac{1}{2} + L(T-nh)\right)C_2(n) + \frac{1}{2}C_4 + (dL(T-nh))^{1/2}\right]\\ &+ h^{3/2} M_1\sqrt{vd}\\ &+ h\left\{(u-v)\left[\left(\frac{1}{2} + L(T-nh)\right)C_2(n) + \frac{1}{2}C_4 + (dL(T-nh))^{1/2}\right]L(T-(n+u)h)\right.\\ &\qquad\left. + M_1(1 + C_2(n) + C_4)\right\}\\ &+ h^{1/2} L(T-(n+u)h)\sqrt{(u-v)d}\\ &+ h(u-v)L(T-(n+u)h)\varepsilon_{sc} + h^2 M_1 v\varepsilon_{sc}.\end{aligned}$$

We claim that we only consider the lowest order of each part, which means the relative higher order term with the combination of d and h will be ignored. Then take the supremum with respect to v,

$$
\begin{aligned}
&\left\{\int_0^1\int_0^1 \left\|\nabla\log p_{T-(n+v)h}(\vartheta_{n+v}^{\mathsf{REM}}) - \nabla\log p_{T-(n+u)h}(\vartheta_{n+v}^{\mathsf{REM}})\right\|_{\mathbb{L}_2}^2 \,\mathrm{d}u\,\mathrm{d}v\right\}^{1/2} \\
&\leqslant h\left\{\int_0^1\int_0^1 \left[|u-v|L(T-(n+u)h)\left(\frac{1}{2}+L(T-nh)\right)+M_1\right]^2 \mathrm{d}u\,\mathrm{d}v\right\}^{1/2} \left\|Y_{nh}-\vartheta_n^{\mathsf{REM}}\right\|_{\mathbb{L}_2} \\
&\quad + h^{1/2}\sqrt{d}\left[\int_0^1\int_0^1 L(T-(n+u)h)^2|u-v|\,\mathrm{d}u\,\mathrm{d}v\right]^{1/2} \\
&\quad + h\left[\int_0^1\int_0^1 (u-v)^2 L(T-(n+u)h)^2\,\mathrm{d}u\,\mathrm{d}v\right]^{1/2}\varepsilon_{sc}.
\end{aligned}
\tag{30}
$$

Combining the above,

$$
\begin{aligned}
&\left\|\vartheta_{n+1}^{\mathsf{REM}} - \mathbb{E}_{U_n}[\vartheta_{n+1}^{\mathsf{REM}}]\right\|_{\mathbb{L}_2} \\
&\leqslant \frac{1}{8\sqrt{3}}h^2\left(\left\|Y_{nh}-\vartheta_n^{\mathsf{REM}}\right\|_{\mathbb{L}_2} + C_2(n) + C_4\right) \\
&\quad + \frac{1}{4\sqrt{3}}h^2\left[\varepsilon_{sc} + L(T-nh)\left(\left\|Y_{nh}-\vartheta_n^{\mathsf{REM}}\right\|_{\mathbb{L}_2} + C_2(n)\right) + (dL(T-nh))^{1/2}\right] \\
&\quad + \frac{1}{2\sqrt{3}}h^{3/2} \\
&\quad + h^2\left\{\int_0^1\int_0^1 \left[|u-v|L(T-(n+u)h)\left(\frac{1}{2}+L(T-nh)\right)+M_1\right]^2 \mathrm{d}u\,\mathrm{d}v\right\}^{1/2} \left\|Y_{nh}-\vartheta_n^{\mathsf{REM}}\right\|_{\mathbb{L}_2} \\
&\quad + h^2\left\{\int_0^1\int_0^1 \left\{(u-v)\left[\left(\frac{1}{2}+L(T-nh)\right)C_2(n) + \frac{1}{2}C_4 + (dL(T-nh))^{1/2}\right]L(T-(n+u)h)\right.\right. \\
&\qquad\qquad \left.\left. + M_1(1+C_2(n)+C_4)\right\}^2 \mathrm{d}u\,\mathrm{d}v\right\}^{1/2} \\
&\quad + h^{3/2}\sqrt{d}\left[\int_0^1\int_0^1 L(T-(n+u)h)^2|u-v|\,\mathrm{d}u\,\mathrm{d}v\right]^{1/2} \\
&\quad + h^2\left[\int_0^1\int_0^1 (u-v)^2 L(T-(n+u)h)^2\,\mathrm{d}u\,\mathrm{d}v\right]^{1/2}\varepsilon_{sc} \\
&\quad + 2h\varepsilon_{sc} \\
&\lesssim h^2\left\{\left[\int_0^1\int_0^1 \left[|u-v|L(T-(n+u)h)\left(\frac{1}{2}+L(T-nh)\right)+M_1\right]^2 \mathrm{d}u\,\mathrm{d}v\right]^{1/2}\right. \\
&\qquad \left. + \frac{1}{4\sqrt{3}}L(T-nh) + \frac{1}{8\sqrt{3}}\right\}\left\|Y_{nh}-\vartheta_n^{\mathsf{REM}}\right\|_{\mathbb{L}_2} \\
&\quad + h^2\left\{\left\{\int_0^1\int_0^1 \left\{(u-v)\left[\left(\frac{1}{2}+L(T-nh)\right)C_2(n) + \frac{1}{2}C_4 + (dL(T-nh))^{1/2}\right]L(T-(n+u)h)\right.\right.\right. \\
&\qquad\qquad \left.\left. + M_1(1+C_2(n)+C_4)\right\}^2 \mathrm{d}u\,\mathrm{d}v\right\}^{1/2} \\
&\qquad\quad \left. + \frac{1}{8\sqrt{3}}(C_2(n)+C_4) + \frac{1}{4\sqrt{3}}(L(T-nh)C_2(n) + (dL(T-nh))^{1/2})\right\} \\
&\quad + h^{3/2}\left\{\sqrt{d}\left[\int_0^1\int_0^1 L(T-(n+u)h)^2|u-v|\,\mathrm{d}u\,\mathrm{d}v\right]^{1/2} + \frac{1}{2\sqrt{3}}\right\} \\
&\quad + 2h\varepsilon_{sc}.
\end{aligned}
\tag{31}
$$

Compared to the term $\tilde{Y}_h - \mathbb{E}_{U_n}[\vartheta_{n+1}^{\mathsf{REM}}]$, we can focus on the lower-order terms, ignoring the score matching error. Therefore, we have

$$
\begin{aligned}
&\left\|\tilde{Y}_{(n+1)h} - \vartheta_{n+1}^{\mathsf{REM}}\right\|_{\mathbb{L}_2} \\
&\leqslant h^2\Bigg\{\left[\int_0^1\int_0^1\left[|u-v|L(T-(n+u)h)\left(\frac{1}{2}+L(T-nh)\right)+M_1\right]^2 \mathrm{d}u\,\mathrm{d}v\right]^{1/2}\left\|Y_{nh}-\vartheta_n^{\mathsf{REM}}\right\|_{\mathbb{L}_2} \\
&\qquad + \frac{1}{4\sqrt{3}}L(T-nh)+\frac{1}{8\sqrt{3}}\Bigg\} \\
&+ h^2\Bigg\{\Bigg\{\int_0^1\int_0^1\Bigg\{(u-v)\left[\left(\frac{1}{2}+L(T-nh)\right)C_2(n)+\frac{1}{2}C_4+(dL(T-nh))^{1/2}\right]L(T-(n+u)h) \\
&\qquad\qquad + M_1(1+C_2(n)+C_4)\Bigg\}^2\mathrm{d}u\,\mathrm{d}v\Bigg\}^{1/2} \\
&\qquad + \frac{1}{8\sqrt{3}}(C_2(n)+C_4)+\frac{1}{4\sqrt{3}}\left(L(T-nh)C_2(n)+(dL(T-nh))^{1/2}\right)\Bigg\} \\
&+ h^{3/2}\left\{\sqrt{d}\left[\int_0^1\int_0^1 L(T-(n+u)h)^2|u-v|\,\mathrm{d}u\,\mathrm{d}v\right]^{1/2}+\frac{1}{2\sqrt{3}}\right\} \\
&+ 3h\varepsilon_{sc}.
\end{aligned}
$$

□

Returning to the proof of Theorem 3, by the conclusion of Proposition 12, we have

$$
\begin{aligned}
\left\|Y_{Nh}-\vartheta_N^{\mathsf{REM}}\right\|_{\mathbb{L}_2} &\lesssim \frac{1}{m_{\min}-1/2}\left(h\max_{0\leqslant k\leqslant N-1}C_{k,1}^{\mathsf{REM}}+h^{1/2}\max_{0\leqslant k\leqslant N-1}C_{k,2}^{\mathsf{REM}}+3\varepsilon_{sc}\right) \\
&\lesssim \sqrt{h}\cdot\frac{\sqrt{d/3}L_{\max}+\frac{1}{2\sqrt{3}}}{m_{\min}-1/2}+\varepsilon_{sc}\cdot\frac{3}{m_{\min}-1/2}.
\end{aligned}
$$

This completes the first part of proof for Theorem 3.

B.4 Proof of Theorem 3: Part II

We begin with the following proposition.

Proposition 14. *Suppose that Assumptions 1, 2 and 4 are satisfied, the following two claims hold*

(1) *It holds that*

$$
\begin{aligned}
&\left\|\tilde{Y}_h - \vartheta_{n+1}^{\mathsf{REI}}\right\|_{\mathbb{L}_2} \\
&\leqslant h^2\Bigg\{\int_0^1\int_0^1\Bigg[|u-v|L(T-(n+u)h)\left(\frac{1}{2}+L(T-nh)\right)+M_1 \\
&\qquad + \frac{1}{2}|u-v|L(T-(n+v)h)r_n^{\mathsf{EI}}(v)\Bigg]^2\mathrm{d}u\,\mathrm{d}v\Bigg\}^{1/2}\left\|Y_{nh}-\vartheta_n^{\mathsf{REI}}\right\|_{\mathbb{L}_2} \\
&\quad + h^2\Bigg\{\frac{e^{\frac{1}{2}(1-v)h}-e^{\frac{1}{2}(1-u)h}}{h}e^{\frac{1}{2}vh}L(T-(n+u)h)\left[C_2(n)+C_4+2L(T-nh)C_2(n)+(dL(T-nh))^{1/2}\right] \\
&\qquad + e^{\frac{1}{2}(1-u)h}M_1\left[1+2e^{\frac{1}{2}vh}\left(L(T-nh)C_2(n)+(dL(T-nh))^{1/2}\right)+C_2(n)+C_4\right] \\
&\qquad + \frac{|e^{\frac{1}{2}(1-u)h}-e^{\frac{1}{2}(1-v)h}|}{h}\left(L(T-(n+v)h)C_2(n)+(dL(T-(n+v)h))^{1/2}\right)\Bigg\}
\end{aligned}
$$

$$+ h^{3/2}\sqrt{d}\bigg\{ \int_0^1 \int_0^1 L(T-(n+u)h)^2|u-v| \, \mathrm{d}u \, \mathrm{d}v \bigg\}^{1/2}$$
$$+ 3h\varepsilon_{sc}.$$

***(2)** Furthermore, it holds that*

$$\left\|Y_{(n+1)h} - \vartheta^{\mathsf{REI}}_{n+1}\right\|_{\mathbb{L}_2} \leqslant r^{\mathsf{REI}}_n \left\|Y_{nh} - \vartheta^{\mathsf{REI}}_n\right\|_{\mathbb{L}_2} + h^2 C^{\mathsf{REI}}_{n,1} + h^{3/2} C^{\mathsf{REI}}_{n,2} + 3h\varepsilon_{sc}\,,$$

where

$$\begin{aligned}
r^{\mathsf{REI}}_n &= e^{-\int_{nh}^{(n+1)h}(m(T-t)-\frac{1}{2})\,\mathrm{d}t} \\
&\quad + h^2\bigg\{ \int_0^1 \int_0^1 \Big[|u-v| L(T-(n+u)h)(\frac{1}{2}+L(T-nh)) \\
&\qquad\qquad + M_1 + \frac{1}{2}|u-v| L(T-(n+v)h) r^{\mathsf{EI}}_n(v)\Big]^2 \mathrm{d}u\,\mathrm{d}v \bigg\}^{1/2}, \\
C^{\mathsf{REI}}_{n,1} &= \frac{e^{\frac{1}{2}(1-v)h} - e^{\frac{1}{2}(1-u)h}}{h} e^{\frac{1}{2}vh} L(T-(n+u)h)\left[C_2(n) + C_4 + 2L(T-nh)C_2(n) + (dL(T-nh))^{1/2}\right] \\
&\quad + e^{\frac{1}{2}(1-u)h} M_1 \left[1 + 2e^{\frac{1}{2}vh}\left(L(T-nh)C_2(n) + (dL(T-nh))^{1/2}\right) + C_2(n) + C_4\right] \\
&\quad + \frac{|e^{\frac{1}{2}(1-u)h} - e^{\frac{1}{2}(1-v)h}|}{h}\left(L(T-(n+v)h)C_2(n) + (dL(T-(n+v)h))^{1/2}\right), \\
C^{\mathsf{REI}}_{n,2} &= \sqrt{d}\bigg\{ \int_0^1 \int_0^1 L(T-(n+u)h)^2|u-v| \, \mathrm{d}u \, \mathrm{d}v \bigg\}^{1/2}.
\end{aligned}$$

Proof of Proposition 14. This proposition can be proven following the same approach as in the proof of Proposition 12, with the only difference being the inclusion of the exponential coefficient term. However, this term does not significantly affect the overall proof.
Similarly, we make a decomposition as

$$\left\|\tilde{Y}_h - \vartheta^{\mathsf{REI}}_{n+1}\right\|_{\mathbb{L}_2} \leqslant \left\|\tilde{Y}_h - \mathbb{E}_{U_n}[\vartheta^{\mathsf{REI}}_{n+1}]\right\|_{\mathbb{L}_2} + \left\|\mathbb{E}_{U_n}[\vartheta^{\mathsf{REI}}_{n+1}]\right\|_{\mathbb{L}_2}. \tag{32}$$

Note that

$$\begin{aligned}
\tilde{Y}_h - \mathbb{E}_{U_n}[\vartheta^{\mathsf{REI}}_{n+1}] &= \int_0^h e^{\frac{1}{2}(h-t)} \nabla \log p_{T-nh-t}(\tilde{Y}_t)\,\mathrm{d}t - h\mathbb{E}_{U_n}\left[e^{\frac{1}{2}(1-U_n)h} s_*(T-nh-U_n h, \vartheta^{\mathsf{REI}}_{n+U_n})\right] \\
&= h\int_0^1 e^{\frac{1}{2}(1-u)h}\left(\nabla \log p_{T-nh-uh}(\tilde{Y}_{uh}) - s_*(T-nh-uh, \vartheta^{\mathsf{REI}}_{n+u})\right)\,\mathrm{d}u \\
&= h\int_0^1 e^{\frac{1}{2}(1-u)h}\left(\nabla \log p_{T-nh-uh}(\tilde{Y}_{uh}) - \nabla \log p_{T-nh-uh}(\vartheta^{\mathsf{REI}}_{n+u})\right)\,\mathrm{d}u \\
&\quad + h\int_0^1 e^{\frac{1}{2}(1-u)h}\left(\nabla \log p_{T-nh-uh}(\vartheta^{\mathsf{REI}}_{n+u}) - s_*(T-nh-uh, \vartheta^{\mathsf{REI}}_{n+u})\right)\,\mathrm{d}u.
\end{aligned}$$

Then, we obtain

$$\begin{aligned}
&\left\|\tilde{Y}_h - \mathbb{E}_{U_n}[\vartheta^{\mathsf{REI}}_{n+1}]\right\|_{\mathbb{L}_2} \\
&\leqslant h\int_0^1 \left\|e^{\frac{1}{2}(1-u)h}(\nabla \log p_{T-nh-uh}(\tilde{Y}_{uh}) - s_*(T-nh-uh, \vartheta^{\mathsf{REI}}_{n+u}))\right\|_{\mathbb{L}_2} \mathrm{d}u \\
&\leqslant h\int_0^1 e^{\frac{1}{2}(1-u)h} L(T-nh-uh)\left\|\tilde{Y}_{uh} - \vartheta^{\mathsf{REI}}_{n+u}\right\|_{\mathbb{L}_2} \mathrm{d}u + h\int_0^1 e^{\frac{1}{2}(1-u)h}\,\mathrm{d}u\,\varepsilon_{sc} \\
&\leqslant h^3\int_0^1 e^{\frac{1}{2}(1-u)h} L(T-nh-uh)u^2\left(C_{5,n}(u)C_{1,n}(u) + M_1\frac{2(e^{uh/2}-1)}{uh}\right)\mathrm{d}u\,\left\|Y_{nh} - \vartheta^{\mathsf{REI}}_n\right\|_{\mathbb{L}_2} \\
&\quad + h^3\int_0^1 e^{\frac{1}{2}(1-u)h} L(T-nh-uh)u^2\Big[C_{5,n}(u)\left(C_{1,n}(u)C_2(n) + \frac{1}{2}C_4 + C_{3,n}(u)\right)
\end{aligned}$$

$$+ \frac{2(e^{uh}-1)}{uh} M_1(1+C_2(n)+C_4)\Big] \, \mathrm{d}u$$

$$+ h^{5/2} \int_0^1 e^{\frac{1}{2}(1-u)h} L(T-nh-uh) u^{3/2} C_{5,n}(u) \, \mathrm{d}u \sqrt{d}$$

$$+ h^2 \int_0^1 e^{\frac{1}{2}(1-u)h} L(T-nh-uh) \frac{2(e^{uh/2}-1)}{h} \, \mathrm{d}u \, \varepsilon_{sc} + h \frac{2(e^{h/2}-1)}{h} \varepsilon_{sc}, \tag{33}$$

where

$$C_{5,n}(u) = \frac{1}{uh} \int_{nh}^{(n+u)h} e^{\frac{1}{2}((n+u)h-t)} L(T-t) \, \mathrm{d}t.$$

In the third inequality, we can directly bound $\left\| \tilde{Y}_{uh} - \vartheta_{n+u}^{\mathsf{REI}} \right\|_{\mathbb{L}_2}$, as it is a special case of Proposition 11, where the step size is replaced by uh.

For the second term of (32), we have

$$\begin{aligned}
&\vartheta_{n+1}^{\mathsf{REI}} - \mathbb{E}_{U_n}[\vartheta_{n+1}^{\mathsf{REI}}] \\
&= h e^{\frac{1}{2}(1-U_n)h} s_*(T-nh-U_nh, \vartheta_{n+U}^{\mathsf{REI}}) - h \mathbb{E}_{U_n}\left[e^{\frac{1}{2}(1-U_n)h} s_*(T-nh-U_nh, \vartheta_{n+U}^{\mathsf{REI}}) \right] \\
&= h \int_0^1 \left[e^{\frac{1}{2}(1-U_n)h} s_*(T-nh-U_nh, \vartheta_{n+U_n}^{\mathsf{REI}}) - e^{\frac{1}{2}(1-v)h} s_*(T-nh-vh, \vartheta_{n+v}^{\mathsf{REI}}) \right] \mathrm{d}v \, .
\end{aligned}$$

Similar to display (29), we then obtain

$$\begin{aligned}
&\left\| \vartheta_{n+1}^{\mathsf{REI}} - \mathbb{E}_{U_n}[\vartheta_{n+1}^{\mathsf{REI}}] \right\|_{\mathbb{L}_2} \\
&\leqslant \left\{ \mathbb{E} \int_0^1 \left[h \int_0^1 e^{\frac{1}{2}(1-u)h} s_*(T-(n+u)h, \vartheta_{n+u}^{\mathsf{REI}}) - e^{\frac{1}{2}(1-v)h} s_*(T-(n+v)h, \vartheta_{n+v}^{\mathsf{REI}}) \, \mathrm{d}v \right]^2 \mathrm{d}u \right\}^{1/2} \\
&\leqslant h \bigg\{ \int_0^1 \int_0^1 \left\| e^{\frac{1}{2}(1-u)h} s_*(T-(n+u)h, \vartheta_{n+u}^{\mathsf{REI}}) - e^{\frac{1}{2}(1-v)h} s_*(T-(n+v)h, \vartheta_{n+v}^{\mathsf{REI}}) \right\|_{\mathbb{L}_2}^2 \mathrm{d}u \, \mathrm{d}v \bigg\}^{1/2} \\
&\leqslant h \bigg\{ \int_0^1 \int_0^1 \left\| e^{\frac{1}{2}(1-u)h} \nabla \log p_{T-(n+u)h}(\vartheta_{n+u}^{\mathsf{REI}}) - e^{\frac{1}{2}(1-v)h} \nabla \log p_{T-(n+v)h}(\vartheta_{n+v}^{\mathsf{REI}}) \right\|_{\mathbb{L}_2}^2 \mathrm{d}u \, \mathrm{d}v \bigg\}^{1/2} \\
&\quad + 2h \left(\int_0^1 e^{(1-u)h} \, \mathrm{d}u \right)^{1/2} \varepsilon_{sc},
\end{aligned}$$

Using the same strategy as in display (30), we arrive at

$$\begin{aligned}
&\left\| e^{\frac{1}{2}(1-u)h} \nabla \log p_{T-(n+u)h}(\vartheta_{n+u}^{\mathsf{REI}}) - e^{\frac{1}{2}(1-v)h} \nabla \log p_{T-(n+v)h}(\vartheta_{n+v}^{\mathsf{REI}}) \right\|_{\mathbb{L}_2} \\
&\leqslant e^{\frac{1}{2}(1-u)h} \left\| \nabla \log p_{T-(n+u)h}(\vartheta_{n+u}^{\mathsf{REI}}) - \nabla \log p_{T-(n+u)h}(\vartheta_{n+v}^{\mathsf{REI}}) \right\|_{\mathbb{L}_2} \\
&\quad + e^{\frac{1}{2}(1-u)h} \left\| \nabla \log p_{T-(n+u)h}(\vartheta_{n+v}^{\mathsf{REI}}) - \nabla \log p_{T-(n+v)h}(\vartheta_{n+v}^{\mathsf{REI}}) \right\|_{\mathbb{L}_2} \\
&\quad + \left| e^{\frac{1}{2}(1-u)h} - e^{\frac{1}{2}(1-v)h} \right| \left\| \nabla \log p_{T-(n+v)h}(\vartheta_{n+v}^{\mathsf{REI}}) \right\|_{\mathbb{L}_2} \\
&\leqslant e^{\frac{1}{2}(1-u)h} L(T-(n+u)h) \left\| \vartheta_{n+u}^{\mathsf{REI}} - \vartheta_{n+v}^{\mathsf{REI}} \right\|_{\mathbb{L}_2} \\
&\quad + e^{\frac{1}{2}(1-u)h} M_1 h (1 + \left\| \vartheta_{n+v}^{\mathsf{REI}} \right\|_{\mathbb{L}_2}) \\
&\quad + \left| e^{\frac{1}{2}(1-u)h} - e^{\frac{1}{2}(1-v)h} \right| \left[L(T-(n+v)h) \left(\left\| Y_{(n+v)h} - \vartheta_{n+v}^{\mathsf{REI}} \right\|_{\mathbb{L}_2} + C_2(n) \right) + (dL(T-(n+v)h))^{1/2} \right].
\end{aligned}$$

The second inequality follows from Assumptions 1 and 2. We bound the term $\left\| \nabla \log p_{T-(n+v)h}(\vartheta_{n+v}^{\mathsf{REI}}) \right\|_{\mathbb{L}_2}$ by decomposing it as follows

$$\begin{aligned}
\left\| \nabla \log p_{T-(n+v)h}(\vartheta_{n+v}^{\mathsf{REI}}) \right\|_{\mathbb{L}_2} &\leqslant \left\| \nabla \log p_{T-(n+v)h}(\vartheta_{n+v}^{\mathsf{REI}}) - \nabla \log p_{T-(n+v)h}(Y_{(n+v)h}) \right\|_{\mathbb{L}_2} \\
&\quad + \left\| \nabla \log p_{T-(n+v)h}(Y_{(n+v)h}) - \nabla \log p_{T-(n+v)h}(X^{\leftarrow}_{((n+v)h)}) \right\|_{\mathbb{L}_2} \\
&\quad + \left\| \nabla \log p_{T-(n+v)h}(X_{T-(n+v)h}) \right\|_{\mathbb{L}_2} .
\end{aligned}$$

Without loss of generality, we consider the case where $u > v$; the other case follows similarly.

$$
\begin{aligned}
&\left\|\vartheta_{n+u}^{\mathsf{REI}} - \vartheta_{n+v}^{\mathsf{REI}}\right\|_{\mathbb{L}_2} \\
&= (e^{\frac{1}{2}uh} - e^{\frac{1}{2}vh}) \left\|\vartheta_n^{\mathsf{REI}}\right\|_{\mathbb{L}_2} + \int_{vh}^{uh} e^{\frac{1}{2}t}\,\mathrm{d}t \left\|s_*(T-nh, \vartheta_n^{\mathsf{REI}})\right\|_{\mathbb{L}_2} \\
&\quad + \left\|\int_{nh}^{(n+u)h} e^{\frac{1}{2}((n+u)h-t)}\,\mathrm{d}W_t - \int_{nh}^{(n+v)h} e^{\frac{1}{2}((n+v)h-t)}\,\mathrm{d}W_t\right\|_{\mathbb{L}_2} \\
&\leqslant \left(e^{\frac{1}{2}(u-v)h} - 1\right)e^{\frac{1}{2}vh}\left(\left\|Y_{nh} - \vartheta_n^{\mathsf{REI}}\right\|_{\mathbb{L}_2} + C_2(n) + C_4\right) \\
&\quad + 2\left(e^{\frac{1}{2}(u-v)h} - 1\right)e^{\frac{1}{2}vh}\left[\varepsilon_{sc} + L(T-nh)\left(\left\|Y_{nh} - \vartheta_n^{\mathsf{REI}}\right\|_{\mathbb{L}_2} + C_2(n)\right) + (dL(T-nh))^{1/2}\right] \\
&\quad + \left[(e^{uh}-1) + (e^{vh}-1) - 2(e^{\frac{u+v}{2}h} - e^{\frac{u-v}{2}h})\right]^{1/2}\sqrt{d}.
\end{aligned} \tag{34}
$$

Here, we apply the formula in (28) to bound the last term.

$$
\begin{aligned}
&\left\|\int_{nh}^{(n+u)h} \mathbf{1}_{\{t\leqslant (n+v)h\}}(e^{\frac{1}{2}((n+u)h-t)} - e^{\frac{1}{2}((n+v)h-t)}) + \mathbf{1}_{\{t>(n+v)h\}}e^{\frac{1}{2}((n+u)h-t)}\,\mathrm{d}W_t\right\|_{\mathbb{L}_2} \\
&= \sqrt{d}\left[\int_{nh}^{(n+v)h} (e^{\frac{1}{2}((n+u)h-t)} - e^{\frac{1}{2}((n+v)h-t)})^2\,\mathrm{d}t + \int_{(n+v)h}^{(n+u)h} e^{(n+u)h-t}\,\mathrm{d}t\right]^{1/2} \\
&= \sqrt{d}\left[(e^{uh}-1) + (e^{vh}-1) - 2(e^{\frac{u+v}{2}h} - e^{\frac{u-v}{2}h})\right]^{1/2},
\end{aligned}
$$

We then bound the term $\left\|\vartheta_{n+v}^{\mathsf{REI}}\right\|_{\mathbb{L}_2}$ following display (34) above. To this end, let $u = 0$, we then have

$$
\begin{aligned}
\left\|\vartheta_{n+v}^{\mathsf{REI}}\right\|_{\mathbb{L}_2} \leqslant\ & (e^{\frac{1}{2}vh} - 1)(\left\|Y_{nh} - \vartheta_n^{\mathsf{REI}}\right\|_{\mathbb{L}_2} + C_2(n) + C_4) \\
& + 2(e^{\frac{1}{2}vh} - 1)\left[\varepsilon_{sc} + L(T-nh)\left(\left\|Y_{nh} - \vartheta_n^{\mathsf{REI}}\right\|_{\mathbb{L}_2} + C_2(n)\right) + (dL(T-nh))^{1/2}\right] \\
& + \sqrt{d}(e^{vh}-1)^{1/2} \\
& + \left\|Y_{nh} - \vartheta_n^{\mathsf{REI}}\right\|_{\mathbb{L}_2} + C_2(n) + C_4.
\end{aligned}
$$

Additionally, we can bound $\left\|Y_{(n+v)h} - \vartheta_{n+v}^{\mathsf{REI}}\right\|_{\mathbb{L}_2}$, as it is a special case of the one-step discretization error under the Exponential Integrator scheme, where the step size is replaced by vh. Specifically, we have

$$
\begin{aligned}
&\left\|Y_{(n+v)h} - \vartheta_{n+v}^{\mathsf{REI}}\right\|_{\mathbb{L}_2} \\
&\leqslant r_n^{\mathsf{EI}}(v)\left\|Y_{nh} - \vartheta_n^{\mathsf{REI}}\right\|_{\mathbb{L}_2} + h^2 C_n^{\mathsf{EI}}(v) + h^{3/2}u^{3/2}\sqrt{d}C_{5,n}(v) + vh\frac{2(e^{\frac{1}{2}vh}-1)}{vh}\varepsilon_{sc},
\end{aligned}
$$

where

$$
\begin{aligned}
r_n^{\mathsf{EI}}(v) &= e^{-\int_{nh}^{(n+v)h}(m(T-t)-\frac{1}{2})\,\mathrm{d}t} + v^2h^2\left(C_{5,n}(v)C_{1,n}(v) + M_1\frac{2(e^{\frac{1}{2}vh}-1)}{vh}\right), \\
C_n^{\mathsf{EI}}(v) &= C_{5,n}(v)\left(C_{1,n}(v)C_2(n) + \frac{1}{2}C_4 + C_{3,n}(v)\right) + \frac{2(e^{\frac{1}{2}vh}-1)}{vh}M_1(1 + C_2(n) + C_4).
\end{aligned}
$$

Then, we obtain

$$
\begin{aligned}
&\left\|e^{\frac{1}{2}(1-u)h}\nabla\log p_{T-(n+u)h}(\vartheta_{n+u}^{\mathsf{REI}}) - e^{\frac{1}{2}(1-v)h}\nabla\log p_{T-(n+v)h}(\vartheta_{n+v}^{\mathsf{REI}})\right\|_{\mathbb{L}_2} \\
&\leqslant \Bigg\{2(e^{\frac{1}{2}(1-v)h} - e^{\frac{1}{2}(1-u)h})e^{\frac{1}{2}vh}L(T-(n+u)h)\left[\frac{1}{2} + L(T-nh)\right] \\
&\qquad + e^{\frac{1}{2}(1-u)h}M_1 h\left[(e^{\frac{1}{2}vh}-1) + 2(e^{\frac{1}{2}vh}-1)L(T-nh) + 1\right]
\end{aligned}
$$

$$
\begin{aligned}
&\quad + \left| e^{\frac{1}{2}(1-u)h} - e^{\frac{1}{2}(1-v)h} \right| L(T-(n+v)h) r_n^{\mathsf{EI}}(v) \Bigg\} \left\| Y_{nh} - \vartheta_n^{\mathsf{REI}} \right\|_{\mathbb{L}_2} \\
&+ h^3 L(T-(n+v)h) \frac{|e^{\frac{1}{2}(1-u)h} - e^{\frac{1}{2}(1-v)h}|}{h} C_n^{\mathsf{EI}}(v) \\
&+ h^{5/2} L(T-(n+v)h) \frac{|e^{\frac{1}{2}(1-u)h} - e^{\frac{1}{2}(1-v)h}|}{h} u^{3/2} \sqrt{d} C_{5,n}(v) \\
&+ h^2 M_1 e^{\frac{1}{2}(1-u)h} \frac{e^{\frac{1}{2}vh} - 1}{h} (C_2(n) + C_4) \\
&+ h^{3/2} M_1 e^{\frac{1}{2}(1-u)h} \sqrt{vd} \left(\frac{e^{vh} - 1}{vh} \right)^{1/2} \\
&+ h \Bigg\{ \frac{e^{\frac{1}{2}(1-v)h} - e^{\frac{1}{2}(1-u)h}}{h} e^{\frac{1}{2}vh} L(T-(n+u)h) \left[C_2(n) + C_4 + 2L(T-nh)C_2(n) + (dL(T-nh))^{1/2} \right] \\
&\quad + e^{\frac{1}{2}(1-u)h} M_1 \left[1 + 2e^{\frac{1}{2}vh} \left(L(T-nh)C_2(n) + (dL(T-nh))^{1/2} \right) + C_2(n) + C_4 \right] \\
&\quad + \frac{|e^{\frac{1}{2}(1-u)h} - e^{\frac{1}{2}(1-v)h}|}{h} \left(L(T-(n+v)h)C_2(n) + (dL(T-(n+v)h))^{1/2} \right) \Bigg\} \\
&+ h^{1/2} L(T-(n+u)h) e^{\frac{1}{2}(1-u)h} \left[\frac{(e^{uh}-1) + (e^{vh}-1) - 2(e^{\frac{u+v}{2}h} - e^{\frac{u-v}{2}h})}{h} \right]^{1/2} \sqrt{d} \\
&+ h^2 \varepsilon_{sc} L(T-(n+v)h) \frac{|e^{\frac{1}{2}(1-u)h} - e^{\frac{1}{2}(1-v)h}|}{h} \frac{2(e^{\frac{1}{2}vh} - 1)}{h} \\
&+ h\varepsilon_{sc} \cdot 2e^{\frac{1}{2}(1-u)h} \left[L(T-(n+u)h) \frac{e^{\frac{1}{2}uh} - e^{\frac{1}{2}vh}}{h} + M_1 e^{\frac{1}{2}vh} \right].
\end{aligned}
$$

Ignoring the higher-order terms, we take the supremum with respect to v and substitute it back into the original expression, yielding

$$
\begin{aligned}
&\left\| \vartheta_{n+1}^{\mathsf{REI}} - \mathbb{E}_{U_n}[\vartheta_{n+1}^{\mathsf{REI}}] \right\|_{\mathbb{L}_2} \\
&\leqslant h \Bigg\{ \int_0^1 \int_0^1 \left\| e^{\frac{1}{2}(1-u)h} s_*(T-(n+u)h, \vartheta_{n+u}^{\mathsf{REI}}) - e^{\frac{1}{2}(1-v)h} s_*(T-(n+v)h, \vartheta_{n+v}^{\mathsf{REI}}) \right\|_{\mathbb{L}_2}^2 \mathrm{d}u\, \mathrm{d}v \Bigg\}^{1/2} \\
&\leqslant h^2 \Bigg\{ \int_0^1 \int_0^1 \Big[|u-v| L(T-(n+u)h) \left(\frac{1}{2} + L(T-nh) \right) + M_1 \\
&\qquad + \frac{1}{2}|u-v| L(T-(n+v)h) r_n^{\mathsf{EI}}(v) \Big]^2 \mathrm{d}u\, \mathrm{d}v \Bigg\}^{1/2} \left\| Y_{nh} - \vartheta_n^{\mathsf{REI}} \right\|_{\mathbb{L}_2} \\
&\quad + h^{3/2} \Bigg\{ \int_0^1 \int_0^1 dL(T-(n+u)h)^2 |u-v|\, \mathrm{d}u\, \mathrm{d}v \Bigg\}^{1/2} + 2h\varepsilon_{sc},.
\end{aligned} \tag{35}
$$

This completes the proof.

□

Now, we have

$$
\begin{aligned}
\left\| Y_{Nh} - \vartheta_N^{\mathsf{REI}} \right\|_{\mathbb{L}_2} &\lesssim \frac{1}{m_{\min} - 1/2} \left(h \max_{0 \leqslant k \leqslant N-1} C_{n,1}^{\mathsf{REI}} + \sqrt{h} \max_{0 \leqslant k \leqslant N-1} C_{n,2}^{\mathsf{REI}} + 3\varepsilon_{sc} \right) \\
&\lesssim \sqrt{dh} \frac{L_{\max}}{\sqrt{3}(m_{\min} - 1/2)} + \varepsilon_{sc} \frac{3}{m_{\min} - 1/2}.
\end{aligned}
$$

The desired result follows readily.

C The proof of the upper bound of error of the second-order acceleration scheme

This section is dedicated to proving the Wasserstein convergence result for second-order acceleration. To this end, we first establish the following proposition.

Proposition 15. *Suppose that Assumptions 1, 3, 5, 6, 7 are satisfied, the following results hold.*

(1) *First, we have an upper bound for $\left\|\tilde{Y}_h - \vartheta_{n+1}^{\mathsf{SO}}\right\|_{\mathbb{L}_2}$ as follows,*

$$\begin{aligned}\left\|\tilde{Y}_h - \vartheta_{n+1}^{\mathsf{SO}}\right\|_{\mathbb{L}_2} \leqslant & A_{n,1} e^{(L(nh)-\frac{1}{2})h} h^2 \left\|Y_{nh} - \vartheta_n^{\mathsf{SO}}\right\|_{\mathbb{L}_2} + A_{n,2} e^{(L(nh)-\frac{1}{2})h} h^2 \\ & + \left(h\varepsilon_{sc} + \frac{2}{3}\sqrt{d} h^{3/2} \varepsilon_{sc}^{(L)} + \frac{1}{2} h^2 \varepsilon_{sc}^{(M)}\right) e^{(L(T-nh)-\frac{1}{2})h},\end{aligned}$$

where

$$\begin{aligned}A_{n,1} = & \sup_{nh\leqslant t\leqslant (n+1)h} \frac{1}{t^2} \int_0^t \Big(\int_0^s \big[(1 + L(T-nh-u))L(T-nh-u) \\ & \qquad + (1+L(T-nh))L(T-nh)\big] \,\mathrm{d}u\Big) \,\mathrm{d}s, \\ A_{n,2} = & \sup_{nh\leqslant t\leqslant (n+1)h} \frac{1}{t^2} \Big[\int_0^t \Big(\int_0^s \big[(1 + L(T-nh-u))L(T-nh-u) \\ & \qquad + (1+L(T-nh))L(T-nh)\big] \,\mathrm{d}u\Big) \,\mathrm{d}s \cdot C_2(n) \\ & \qquad + \sqrt{d} \int_0^t \left(\int_0^s \Big[\Big(\frac{1}{2} + L(T-nh-u)\Big) L(T-nh-u)^{1/2} \right. \\ & \qquad \left. + \Big(\frac{1}{2} + L(T-nh)\Big) L(T-nh)^{1/2}\Big] \,\mathrm{d}u \right) \mathrm{d}s \\ & \qquad + \int_0^t \left(\int_0^s \frac{1}{2}(L(T-nh-u) + L(T-nh)) \,\mathrm{d}u \right) \mathrm{d}s \cdot C_4 \Big] \\ & + \frac{\sqrt{2}}{4} \sqrt{d} L_F\,.\end{aligned}$$

(2) *Furthermore, it holds that*

$$\begin{aligned}\left\|Y_{(n+1)h} - \vartheta_{n+1}^{\mathsf{SO}}\right\|_{\mathbb{L}_2} \leqslant\ & r_n^{\mathsf{SO}} \left\|Y_{nh} - \vartheta_n^{\mathsf{SO}}\right\|_{\mathbb{L}_2} + h^2 C_n^{\mathsf{SO}} \\ & + \left[h\varepsilon_{sc} + \frac{2}{3}\sqrt{d} h^{3/2} \varepsilon_{sc}^{(L)} + \frac{1}{2} h^2 \varepsilon_{sc}^{(M)}\right] e^{(L(T-nh)-\frac{1}{2})h},\end{aligned}$$

where

$$\begin{aligned}r_n^{\mathsf{SO}} =& e^{-\int_{nh}^{(n+1)h}(m(T-t)-\frac{1}{2})\,\mathrm{d}t} + h^2 A_{n,1} e^{(L(T-nh)-\frac{1}{2})h} \\ C_n^{\mathsf{SO}} =& A_{n,2} e^{(L(T-nh)-\frac{1}{2})h}\,.\end{aligned}$$

Proof. Recall the expression in display (44), which states that

$$x_t = \vartheta_n^{\mathsf{SO}} + \int_{nh}^t \left(\frac{1}{2}\vartheta_n^{\mathsf{SO}} + \nabla \log p_{T-nh}(\vartheta_n^{\mathsf{SO}}) + L_n(x_s - \vartheta_n^{\mathsf{SO}}) + M_n(s-nh)\right) \mathrm{d}s + \int_{nh}^t \mathrm{d}W_s$$

with

$$\begin{aligned}L_n &= \frac{1}{2} I_d + \nabla^2 \log p_{T-nh}(\vartheta_n^{\mathsf{SO}}) \in \mathbb{R}^{d\times d}, \\ M_n &= \frac{1}{2}\sum_{j=1}^d \frac{\partial^2}{\partial x_j^2} \nabla \log p_{T-nh}(\vartheta_n^{\mathsf{SO}}) - \frac{\partial}{\partial t} \nabla \log p_{T-nh}(\vartheta_n^{\mathsf{SO}}) \in \mathbb{R}^d.\end{aligned}$$

Plugging the estimates of $\nabla \log p_{T-nh}(\vartheta_n^{\mathsf{SO}})$, L_n and M_n into the previous display yields the following process for x_t^{SO}

$$
\begin{aligned}
x_t^{\mathsf{SO}} &= \vartheta_n^{\mathsf{SO}} + \int_{nh}^{t} \frac{1}{2}\vartheta_n^{\mathsf{SO}} + s_*(T-nh, \vartheta_n^{\mathsf{SO}})\,\mathrm{d}s \\
&\quad + \int_{nh}^{t} s_*^{(L)}(T-nh, \vartheta_n^{\mathsf{SO}})(x_s^{\mathsf{SO}} - \vartheta_n^{\mathsf{SO}}) + s_*^{(M)}(T-nh, \vartheta_n^{\mathsf{SO}})(s-nh)\,\mathrm{d}s + \int_{nh}^{t} \mathrm{d}W_s\,.
\end{aligned}
$$

Then, we obtain

$$
\begin{aligned}
x_t^{\mathsf{SO}} - x_t &= (t-nh)(s_*(T-nh, \vartheta_n^{\mathsf{SO}}) - \nabla \log p_{T-nh}(\vartheta_n^{\mathsf{SO}})) \\
&\quad + (s_*^{(L)}(T-nh, \vartheta_n^{\mathsf{SO}}) - L_n) \int_{nh}^{t} (x_s^{\mathsf{SO}} - \vartheta_n^{\mathsf{SO}})\,\mathrm{d}s + L_n \int_{nh}^{t} (x_s^{\mathsf{SO}} - x_s)\,\mathrm{d}s \\
&\quad + (s_*^{(M)}(T-nh, \vartheta_n^{\mathsf{SO}}) - M_n) \cdot \frac{1}{2}(t-nh)^2.
\end{aligned}
$$

Notice that

$$
\|L_n\|_{\mathbb{L}_2} = \left\| \frac{1}{2} I_d + \nabla^2 \log p_{T-nh}(\vartheta_n^{\mathsf{SO}}) \right\|_{\mathbb{L}_2} \leqslant L(T-nh) - \frac{1}{2}. \tag{36}
$$

Combining this with Assumptions 3 and 5 then provides us with

$$
\begin{aligned}
&\left\| x_t^{\mathsf{SO}} - x_t \right\|_{\mathbb{L}_2} \\
&\leqslant (t-nh)\varepsilon_{sc} + \varepsilon_{sc}^{(L)} \int_{nh}^{t} \left\| x_s^{\mathsf{SO}} - \vartheta_n^{\mathsf{SO}} \right\|_{\mathbb{L}_2} \mathrm{d}s + \|L_n\|_{\mathbb{L}_2} \int_{nh}^{t} \left\| x_s^{\mathsf{SO}} - x_s \right\| \mathrm{d}s + \frac{1}{2}(t-nh)^2 \varepsilon_{sc}^{(M)} \\
&\lesssim \left(L(T-nh) - \frac{1}{2} \right) \int_{nh}^{t} \left\| x_s^{\mathsf{SO}} - x_s \right\| \mathrm{d}s + (t-nh)\varepsilon_{sc} + \frac{2}{3}\sqrt{d}(t-nh)^{3/2}\varepsilon_{sc}^{(L)} + \frac{1}{2}(t-nh)^2 \varepsilon_{sc}^{(M)}.
\end{aligned}
$$

We omit the constant of term $\frac{2}{3}\sqrt{d}(t-nh)^{3/2}\varepsilon_{sc}^{(L)}$ in the last inequality. To handle the resulting integral inequality, we invoke the following Grönwall-type inequality.

Lemma 16. *Let $z(t) \geqslant t_0$ satisfy the following inequality:*

$$
z(t) \leqslant \alpha(t) + \int_{t_0}^{t} \beta(s) z(s)\,\mathrm{d}s, \quad t \geqslant t_0,
$$

where $\beta(s)$ is non-negative, and t_0 is the initial time. Then, the solution $z(t)$ satisfies the following bound:

$$
z(t) \leqslant \alpha(t) + \int_{t_0}^{t} \alpha(s)\beta(s) \exp\left(\int_{s}^{t} \beta(r)\,\mathrm{d}r \right) \mathrm{d}s, \quad t \geqslant t_0.
$$

Additionally, if $\alpha(t)$ is non-decreasing function, then

$$
z(t) \leqslant \alpha(t) \exp\left(\int_{t_0}^{t} \beta(s)\,\mathrm{d}s \right), \quad t \geqslant t_0.
$$

Let

$$
\begin{aligned}
z(t) &= \left\| x_t^{\mathsf{SO}} - x_t \right\|_{\mathbb{L}_2} \\
\alpha(t) &= (t-nh)\varepsilon_{sc} + \frac{2}{3}\sqrt{d}(t-nh)^{3/2}\varepsilon_{sc}^{(L)} + \frac{1}{2}(t-nh)^2 \varepsilon_{sc}^{(M)} \\
\beta(t) &= L(T-nh) - \frac{1}{2}\,,
\end{aligned}
$$

and set $t_0 = nh$. By Lemma 16, we have

$$
\left\| \vartheta_{n+1}^{\mathsf{SO}} - x_{(n+1)h} \right\|_{\mathbb{L}_2} \leqslant \left(h\varepsilon_{sc} + \frac{2}{3}\sqrt{d}h^{3/2}\varepsilon_{sc}^{(L)} + \frac{1}{2}h^2 \varepsilon_{sc}^{(M)} \right) e^{(L(T-nh)-\frac{1}{2})h}. \tag{37}
$$

The original SDE can be rewritten as follows

$$\tilde{Y}_t = \tilde{Y}_0 + \int_0^t \Big(\frac{1}{2}\tilde{Y}_s + \nabla \log p_{T-nh-s}(\tilde{Y}_s)\Big)\,\mathrm{d}s + \int_{nh}^{nh+t} \mathrm{d}W_s.$$

Combining this with the definition of x_t in (44), we then have

$$\begin{aligned}
\tilde{Y}_t - x_{nh+t} &= \int_0^t \bigg(\frac{1}{2}\tilde{Y}_s + \nabla \log p_{T-nh-s}(\tilde{Y}_s) - \frac{1}{2}\vartheta_n^{\mathsf{SO}} - \nabla \log p_{T-nh}(\vartheta_n^{\mathsf{SO}}) - L_n(x_{nh+s} - \vartheta_n^{\mathsf{SO}}) - M_n s\bigg)\,\mathrm{d}s \\
&= \int_0^t L_n(\tilde{Y}_s - x_{nh+s})\,\mathrm{d}s + \int_0^t \bigg(\frac{1}{2}\tilde{Y}_s + \nabla \log p_{T-nh-s}(\tilde{Y}_s) - \frac{1}{2}\vartheta_n^{\mathsf{SO}} - \nabla \log p_{T-nh}(\vartheta_n^{\mathsf{SO}})\bigg)\,\mathrm{d}s \\
&\quad - \int_0^t \bigg(\int_0^s L_n\,\mathrm{d}\tilde{Y}_u\,\mathrm{d}u\bigg)\,\mathrm{d}s - \int_0^t \bigg(\int_0^s M_n\,\mathrm{d}u\bigg)\,\mathrm{d}s \\
&= \int_0^t L_n(\tilde{Y}_s - x_{nh+s})\,\mathrm{d}s + \int_0^t \bigg(\int_0^s \mathrm{d}\left(\frac{1}{2}\tilde{Y}_u + \nabla \log p_{T-nh-u}(\tilde{Y}_u)\right)\bigg)\,\mathrm{d}s \\
&\quad - \int_0^t \bigg(\int_0^s L_n\,\mathrm{d}\tilde{Y}_u\,\mathrm{d}u\bigg)\,\mathrm{d}s - \int_0^t \bigg(\int_0^s M_n\,\mathrm{d}u\bigg)\,\mathrm{d}s\,.
\end{aligned}$$

We then apply the Itô formula to the term $\mathrm{d}\left(\frac{1}{2}\tilde{Y}_u + \nabla \log p_{T-nh-u}(\tilde{Y}_u)\right)$. Recall the definitions of L_n and M_n, and after rearranging the expression, we obtain

$$\begin{aligned}
\tilde{Y}_t - x_{nh+t} &= \underbrace{\int_0^t L_n(\tilde{Y}_s - x_{nh+s})\,\mathrm{d}s}_{\mathrm{I}} + \underbrace{\int_0^t \bigg(\int_0^s \nabla^2 \log p_{T-nh-u}(\tilde{Y}_u) - \nabla^2 \log p_{T-nh}(\tilde{Y}_0)\,\mathrm{d}\tilde{Y}_u\bigg)\,\mathrm{d}s}_{\mathrm{II}} \\
&\quad + \underbrace{\int_0^t \bigg[\int_0^s \bigg(\frac{1}{2}\sum_{j=1}^d \frac{\partial^2}{\partial x_j^2}\nabla \log p_{T-nh-u}(\tilde{Y}_u) - \partial_t \nabla \log p_{T-nh-u}(\tilde{Y}_u)\bigg)}_{\mathrm{III}} \\
&\qquad\qquad \underbrace{- \bigg(\frac{1}{2}\sum_{j=1}^d \frac{\partial^2}{\partial x_j^2}\nabla \log p_{T-nh}(\tilde{Y}_0) - \partial_t \nabla \log p_{T-nh}(\tilde{Y}_0)\bigg)\,\mathrm{d}u\bigg]\,\mathrm{d}s}_{\mathrm{III}}\,.
\end{aligned}$$

In what follows, we derive the upper bounds for each term on the right-hand side of the previous display.

Upper bound for term I**:** The upper bound of the term I follows directly from the fact that

$$\|L_n\|_{\mathbb{L}_2} \leqslant L(T-nh) - \frac{1}{2}\,.$$

Upper bound for term II**:** To derive the upper bound for the second term, we expand the term $\mathrm{d}\tilde{Y}_u$, yielding

$$\begin{aligned}
&\bigg\|\int_0^s \nabla^2 \log p_{T-nh-u}(\tilde{Y}_u) - \nabla^2 \log p_{T-nh}(\tilde{Y}_0)\,\mathrm{d}\tilde{Y}_u\bigg\|_{\mathbb{L}_2} \\
&\leqslant \bigg\|\int_0^s \big(\nabla^2 \log p_{T-nh-u}(\tilde{Y}_u) - \nabla^2 \log p_{T-nh}(\tilde{Y}_0)\big)\Big(\frac{1}{2}\tilde{Y}_u + \nabla \log p_{T-nh-u}(\tilde{Y}_u)\Big)\,\mathrm{d}u\bigg\|_{\mathbb{L}_2} \\
&\quad + \bigg\|\int_0^s \big(\nabla^2 \log p_{T-nh-u}(\tilde{Y}_u) - \nabla^2 \log p_{T-nh}(\tilde{Y}_0)\big)\,\mathrm{d}W_u\bigg\|_{\mathbb{L}_2} \\
&\leqslant \int_0^s \Big\|\nabla^2 \log p_{T-nh-u}(\tilde{Y}_u) - \nabla^2 \log p_{T-nh}(\tilde{Y}_0)\Big\|_{\mathbb{L}_2} \cdot \Big\|\frac{1}{2}\tilde{Y}_u + \nabla \log p_{T-nh-u}(\tilde{Y}_u)\Big\|_{\mathbb{L}_2}\,\mathrm{d}u \\
&\quad + \bigg(\int_0^s \Big\|\nabla^2 \log p_{T-nh-u}(\tilde{Y}_u) - \nabla^2 \log p_{T-nh}(\tilde{Y}_0)\Big\|_{\mathbb{L}_2}^2\,\mathrm{d}u\bigg)^{1/2}\,.
\end{aligned}$$

The second inequality follows from display (28). We note that by Assumptions 6 and 7, it holds that

$$
\begin{aligned}
&\left\|\nabla^2 \log p_{T-nh-u}(\tilde{Y}_u) - \nabla^2 \log p_{T-nh}(\tilde{Y}_0)\right\|_{\mathbb{L}_2} \\
\leqslant &\left\|\nabla^2 \log p_{T-nh-u}(\tilde{Y}_u) - \nabla^2 \log p_{T-nh}(\tilde{Y}_u)\right\|_{\mathbb{L}_2} + \left\|\nabla^2 \log p_{T-nh}(\tilde{Y}_u) - \nabla^2 \log p_{T-nh}(\tilde{Y}_0)\right\|_{\mathbb{L}_2} \\
\leqslant &M_2 h(1 + \left\|\tilde{Y}_u\right\|_{\mathbb{L}_2}) + L_F \left\|\tilde{Y}_u - \tilde{Y}_0\right\|_{\mathbb{L}_2} \\
\leqslant &M_2 h + (L_F + M_2 h) \left\|\tilde{Y}_u - \tilde{Y}_0\right\|_{\mathbb{L}_2} + M_2 h \left\|\vartheta_n^{\mathsf{SO}}\right\|_{\mathbb{L}_2} \\
\lesssim &L_F \sqrt{du},
\end{aligned}
$$

Combining this with the previous display provides us with

$$
\begin{aligned}
\left\|\int_0^s \nabla^2 \log p_{T-nh-u}(\tilde{Y}_u) - \nabla^2 \log p_{T-nh}(\tilde{Y}_0)\, \mathrm{d}\tilde{Y}_u\right\|_{\mathbb{L}_2} \lesssim \int_0^s L_F\sqrt{du} \cdot \left\|\frac{1}{2}\tilde{Y}_u + \nabla \log p_{T-nh-u}(\tilde{Y}_u)\right\|_{\mathbb{L}_2} \mathrm{d}u \\
+ \left(\int_0^s L_F^2 du\, \mathrm{d}u\right)^{1/2}.
\end{aligned}
$$

Hence, we obtain

$$
\begin{aligned}
\|\mathrm{II}\|_{\mathbb{L}_2} &= \left\|\int_0^t \left(\int_0^s \nabla^2 \log p_{T-nh-u}(\tilde{Y}_u) - \nabla^2 \log p_{T-nh}(\tilde{Y}_0)\, \mathrm{d}\tilde{Y}_u\right) \mathrm{d}s\right\|_{\mathbb{L}_2} \\
&\leqslant \int_0^t \left\|\int_0^s \nabla^2 \log p_{T-nh-u}(\tilde{Y}_u) - \nabla^2 \log p_{T-nh}(\tilde{Y}_0)\, \mathrm{d}\tilde{Y}_u\right\|_{\mathbb{L}_2} \mathrm{d}s \\
&\leqslant \int_0^t \left[\int_0^s L_F\sqrt{du} \cdot \left\|\frac{1}{2}\tilde{Y}_u + \nabla \log p_{T-nh-u}(\tilde{Y}_u)\right\|_{\mathbb{L}_2} \mathrm{d}u + \left(\int_0^s L_F^2 du\, \mathrm{d}u\right)^{1/2}\right] \mathrm{d}s \\
&\lesssim \frac{\sqrt{2}}{4} L_F \sqrt{d} t^2.
\end{aligned} \tag{38}
$$

Upper bound for term III:

In Section 4, it is claimed that the partial derivative of $\nabla \log p_t$ with respect to t can be estimated without requiring additional assumptions. This is achieved by transforming the t-derivative into x-derivative via the Fokker-Planck equation, as detailed below.

$$
\partial_t p_t(x) = \frac{d}{2} p_t(x) + \frac{1}{2} x^\top \nabla p_t(x) + \frac{1}{2} \sum_{i=1}^d \frac{\partial^2 p_t(x)}{\partial x_i^2}. \tag{39}
$$

We need the following auxiliary lemma.

Lemma 17. *Let p_t be the probability density function of X_t, then*

$$
\begin{aligned}
\sum_{i=1}^d \frac{\partial^2 p_t(x)}{\partial x_i^2} \cdot \frac{1}{p_t(x)} &= \mathrm{Tr}\left(\nabla^2 \log p_t(x)\right) + \|\nabla \log p_t(x)\|^2, \\
\nabla \left(\sum_{i=1}^d \frac{\partial^2 p_t(x)}{\partial x_i^2}\right) \cdot \frac{1}{p_t(x)} &= \nabla\left(\mathrm{Tr}(\nabla^2 \log p_t(x))\right) + \nabla(\|\nabla \log p_t(x)\|^2) \\
&\quad + \left[\mathrm{Tr}(\nabla^2 \log p_t(x)) + \|\nabla \log p_t(x)\|^2\right] \cdot \nabla \log p_t(x).
\end{aligned}
$$

We begin by taking the gradient of $\log p_t$, and then compute the partial derivative of $\nabla \log p_t$ with respect to t. This results in

$$
\partial_t \nabla \log p_t(x) = \partial_t \left(\frac{\nabla p_t(x)}{p_t(x)}\right) = \frac{\partial_t \nabla p_t(x)}{p_t(x)} - \frac{\nabla p_t(x)}{p_t(x)} \cdot \frac{\partial_t p_t(x)}{p_t(x)}.
$$

Under certain regularity conditions, we can interchange the operators ∂_t and ∇ in the term $\partial_t \nabla p_t(x)$, and substitute $\partial_t p_t(x)$ by (39), it follows that

$$\partial_t \nabla \log p_t(x) = \frac{\nabla \partial_t p_t(x)}{p_t(x)} - \nabla \log p_t(x) \cdot \left(\frac{d}{2} + \frac{1}{2} x^\top \nabla \log p_t(x) + \frac{1}{2} \sum_{i=1}^d \frac{\partial^2 p_t(x)}{\partial x_i^2} \cdot \frac{1}{p_t(x)} \right).$$

and

$$\begin{aligned}
\frac{\nabla \partial_t p_t(x)}{p_t(x)} &= \frac{1}{p_t(x)} \cdot \nabla \left(\frac{d}{2} p_t(x) + \frac{1}{2} x^\top \nabla p_t(x) + \frac{1}{2} \sum_{i=1}^d \frac{\partial^2 p_t(x)}{\partial x_i^2} \right) \\
&= \frac{d}{2} \frac{\nabla p_t(x)}{p_t(x)} + \frac{1}{2} \frac{\nabla p_t(x)}{p_t(x)} + \frac{1}{2} \frac{\nabla^2 p_t(x) x}{p_t(x)} + \frac{1}{2} \nabla \left(\sum_{i=1}^d \frac{\partial^2 p_t(x)}{\partial x_i^2} \right) \cdot \frac{1}{p_t(x)} \\
&= \frac{d+1}{2} \nabla \log p_t(x) + \frac{1}{2} \frac{\nabla^2 p_t(x) x}{p_t(x)} + \frac{1}{2} \nabla \left(\sum_{i=1}^d \frac{\partial^2 p_t(x)}{\partial x_i^2} \right) \cdot \frac{1}{p_t(x)}.
\end{aligned}$$

Therefore, we obtain that

$$\begin{aligned}
\partial_t \nabla \log p_t(x) &= \frac{d+1}{2} \nabla \log p_t(x) + \frac{1}{2} \frac{\nabla^2 p_t(x) x}{p_t(x)} + \frac{1}{2} \nabla \left(\sum_{i=1}^d \frac{\partial^2 p_t(x)}{\partial x_i^2} \right) \cdot \frac{1}{p_t(x)} \\
&\quad - \nabla \log p_t(x) \cdot \left(\frac{d}{2} + \frac{1}{2} x^\top \nabla \log p_t(x) + \frac{1}{2} \sum_{i=1}^d \frac{\partial^2 p_t(x)}{\partial x_i^2} \cdot \frac{1}{p_t(x)} \right) \\
&= \frac{1}{2} \nabla \log p_t(x) + \frac{1}{2} \left(\frac{\nabla^2 p_t(x) x}{p_t(x)} - \nabla \log p_t(x) \nabla \log p_t(x)^\top x \right) \\
&\quad + \frac{1}{2} \nabla \left(\sum_{i=1}^d \frac{\partial^2 p_t(x)}{\partial x_i^2} \right) \cdot \frac{1}{p_t(x)} - \frac{1}{2} \nabla \log p_t(x) \sum_{i=1}^d \frac{\partial^2 p_t(x)}{\partial x_i^2} \cdot \frac{1}{p_t(x)}.
\end{aligned}$$

By Lemma 17, the last two terms above can be calculated. Additionally, it holds that

$$\nabla^2 \log p_t(x) = \frac{\nabla^2 p_t(x)}{p_t(x)} - \frac{\nabla p_t(x) \nabla p_t(x)^\top}{p_t(x)^2}.$$

Thus, $\partial_t \nabla \log p_t(x)$ can be simplified to

$$\begin{aligned}
\partial_t \nabla \log p_t(x) &= \frac{1}{2} \nabla \log p_t(x) + \frac{1}{2} \nabla^2 \log p_t(x) x \\
&\quad + \frac{1}{2} \left[\nabla \left(\mathrm{Tr}(\nabla^2 \log p_t(x)) \right) + \nabla (\|\nabla \log p_t(x)\|^2) \right] \\
&\quad + \frac{1}{2} \left(\mathrm{Tr}(\nabla^2 \log p_t(x)) + \|\nabla \log p_t(x)\|^2 \right) \cdot \nabla \log p_t(x) \\
&\quad - \frac{1}{2} \nabla \log p_t(x) \left(\mathrm{Tr}(\nabla^2 \log p_t(x)) + \|\nabla \log p_t(x)\|^2 \right) \\
&= \frac{1}{2} \nabla \log p_t(x) + \frac{1}{2} \nabla^2 \log p_t(x) x + \frac{1}{2} \nabla (\mathrm{Tr}(\nabla^2 \log p_t(x))) + \frac{1}{2} \nabla (\|\nabla \log p_t(x)\|^2).
\end{aligned}$$

Notice that

$$\sum_{j=1}^d \frac{\partial^2}{\partial x_j^2} \nabla \log p_t(x) = \frac{1}{2} \nabla (\mathrm{Tr}(\nabla^2 \log p_t(x))).$$

Then, it follows that

$$\begin{aligned}
&\frac{1}{2} \sum_{j=1}^d \frac{\partial^2}{\partial x_j^2} \nabla \log p_t(x) - \partial_t \nabla \log p_t(x) \\
&= - \left(\frac{1}{2} \nabla \log p_t(x) + \frac{1}{2} \nabla^2 \log p_t(x) x + \frac{1}{2} \nabla (\|\nabla \log p_t(x)\|^2) \right) \\
&= -\frac{1}{2} \left(\nabla \log p_t(x) + \nabla^2 \log p_t(x) x \right) + \nabla^2 \log p_t(x) \cdot \nabla \log p_t(x).
\end{aligned}$$

Therefore we have

$$\begin{aligned}
&\left\|\frac{1}{2}\sum_{j=1}^{d}\frac{\partial^2}{\partial x_j^2}\nabla\log p_{T-nh-u}(\tilde{Y}_u)-\partial_t\nabla\log p_{T-nh-u}(\tilde{Y}_u)\right\|_{\mathbb{L}_2}\\
&\leqslant \frac{1}{2}\left\|\nabla\log p_{T-nh-u}(\tilde{Y}_u)\right\|_{\mathbb{L}_2}+\frac{1}{2}\left\|\nabla^2\log p_{T-nh-u}(\tilde{Y}_u)\right\|_{\mathbb{L}_2}\left\|\tilde{Y}_u\right\|_{\mathbb{L}_2}\\
&\quad+\left\|\nabla^2\log p_{T-nh-u}(\tilde{Y}_u)\right\|_{\mathbb{L}_2}\left\|\nabla\log p_{T-nh-u}(\tilde{Y}_u)\right\|_{\mathbb{L}_2}\\
&\leqslant \frac{1}{2}\left\|\nabla\log p_{T-nh-u}(\tilde{Y}_u)\right\|_{\mathbb{L}_2}+\frac{1}{2}L(T-nh-u)\left\|\tilde{Y}_u\right\|_{\mathbb{L}_2}+L(T-nh-u)\left\|\nabla\log p_{T-nh-u}(\tilde{Y}_u)\right\|_{\mathbb{L}_2}.
\end{aligned}$$

The second inequality follows from Assumption 1. The bounds for $\left\|\tilde{Y}_u\right\|_{\mathbb{L}_2}$ and $\left\|\nabla\log p_{T-nh-u}(\tilde{Y}_u)\right\|_{\mathbb{L}_2}$ can be derived according to the proof of Lemma 10. We then find

$$\begin{aligned}
&\left\|\frac{1}{2}\sum_{j=1}^{d}\frac{\partial^2}{\partial x_j^2}\nabla\log p_{T-nh-u}(\tilde{Y}_u)-\partial_t\nabla\log p_{T-nh-u}(\tilde{Y}_u)\right\|_{\mathbb{L}_2}\\
&\leqslant \frac{1}{2}\left\|\nabla\log p_{T-nh-u}(\tilde{Y}_u)\right\|_{\mathbb{L}_2}+\frac{1}{2}L(T-nh-u)\left\|\tilde{Y}_u\right\|_{\mathbb{L}_2}+L(T-nh-u)\left\|\nabla\log p_{T-nh-u}(\tilde{Y}_u)\right\|_{\mathbb{L}_2}\\
&\leqslant\left(\frac{1}{2}+L(T-nh-u)\right)\left[L(T-nh-u)(\|Y_{nh}-\vartheta_n^{\mathsf{SO}}\|_{\mathbb{L}_2}+C_2(n))+\big(dL(T-nh-u)\big)^{1/2}\right]\\
&\quad+\frac{1}{2}L(T-nh-u)\Big(\|Y_{nh}-\vartheta_n^{\mathsf{SO}}\|_{\mathbb{L}_2}+C_2(n)+C_4\Big).
\end{aligned}$$

Therefore, we obtain

$$\begin{aligned}
\|\mathrm{III}\|_{\mathbb{L}_2}\leqslant&\int_0^t\int_0^s\left\|\frac{1}{2}\sum_{j=1}^{d}\frac{\partial^2}{\partial x_j^2}\nabla\log p_{T-nh-u}(\tilde{Y}_u)-\partial_t\nabla\log p_{T-nh-u}(\tilde{Y}_u)\right\|_{\mathbb{L}_2}\mathrm{d}u\,\mathrm{d}s\\
&+\int_0^t\int_0^s\left\|\frac{1}{2}\sum_{j=1}^{d}\frac{\partial^2}{\partial x_j^2}\nabla\log p_{T-nh}(\tilde{Y}_0)-\partial_t\nabla\log p_{T-nh}(\tilde{Y}_0)\right\|_{\mathbb{L}_2}\mathrm{d}u\,\mathrm{d}s\\
\leqslant&\int_0^t\left(\int_0^s[(1+L(T-nh-u))L(T-nh-u)+(1+L(T-nh))L(T-nh)]\,\mathrm{d}u\right)\mathrm{d}s\cdot\|Y_{nh}-\vartheta_n^{\mathsf{SO}}\|_{\mathbb{L}_2}\\
&+\int_0^t\left(\int_0^s[(1+L(T-nh-u))L(T-nh-u)+(1+L(T-nh))L(T-nh)]\,\mathrm{d}u\right)\mathrm{d}s\cdot C_2(n)\\
&+\sqrt{d}\int_0^t\left(\int_0^s\left[(\frac{1}{2}+L(T-nh-u))L(T-nh-u)^{1/2}+(\frac{1}{2}+L(T-nh))L(T-nh)^{1/2}\right]\mathrm{d}u\right)\mathrm{d}s\\
&+\int_0^t\left(\int_0^s\frac{1}{2}(L(T-nh-u)+L(T-nh))\,\mathrm{d}u\right)\mathrm{d}s\cdot C_4.
\end{aligned}\tag{40}$$

For simplicity, we focus on the lowest-order term. Recall equations (36), (38) and (40), which lead to the following expression

$$\left\|\tilde{Y}_t-x_{nh+t}\right\|_{\mathbb{L}_2}\leqslant\Big(L(T-nh)-\frac{1}{2}\Big)\int_0^t\left\|\tilde{Y}_s-x_{nh+s}\right\|_{\mathbb{L}_2}\mathrm{d}s+\Big(A_{n,1}\|Y_{nh}-\vartheta_n^{\mathsf{SO}}\|_{\mathbb{L}_2}+A_{n,2}\Big)\cdot t^2,$$

where

$$A_{n,1}=\sup_{nh\leqslant t\leqslant(n+1)h}\frac{1}{t^2}\int_0^t\left(\int_0^s[(1+L(T-nh-u))L(T-nh-u)+(1+L(T-nh))L(T-nh)]\,\mathrm{d}u\right)\mathrm{d}s,$$

and

$$A_{n,2} = \sup_{nh\leqslant t\leqslant (n+1)h} \frac{1}{t^2}\Bigg[\int_0^t \left(\int_0^s [(1+L(T-nh-u))L(T-nh-u)+(1+L(T-nh))L(T-nh)]\,\mathrm{d}u\right)\mathrm{d}s\cdot C_2(n)$$
$$+\sqrt{d}\int_0^t\left(\int_0^s\left[(\frac{1}{2}+L(T-nh-u))L(T-nh-u)^{1/2}+(\frac{1}{2}+L(T-nh))L(T-nh)^{1/2}\right]\mathrm{d}u\right)\mathrm{d}s$$
$$+\int_0^t\left(\int_0^s \frac{1}{2}(L(T-nh-u)+L(T-nh))\,\mathrm{d}u\right)\mathrm{d}s\cdot C_4\Bigg]$$
$$+\frac{\sqrt{2}}{4}\sqrt{d}L_F.$$

Using Lemma 16 with

$$z(t) = \left\|\tilde{Y}_t - x_{nh+t}\right\|_{\mathbb{L}_2}$$
$$\alpha(t) = (A_{n,1}\left\|Y_{nh}-\vartheta_n^{\mathsf{SO}}\right\|_{\mathbb{L}_2}+A_{n,2})\cdot t^2$$
$$\beta(t) = L(T-nh)-\frac{1}{2},$$

set $t_0 = nh$, we then obtain

$$\left\|\tilde{Y}_h - x_{(n+1)h}\right\|_{\mathbb{L}_2} \leqslant (A_{n,1}\left\|Y_{nh}-\vartheta_n^{\mathsf{SO}}\right\|_{\mathbb{L}_2}+A_{n,2})h^2\exp\left((L(T-nh)-\frac{1}{2})h\right)$$
$$= A_{n,1}e^{\left(L(T-nh)-\frac{1}{2}\right)h}h^2\left\|Y_{nh}-\vartheta_n^{\mathsf{SO}}\right\|_{\mathbb{L}_2}+A_{n,2}e^{(L(T-nh)-\frac{1}{2})h}h^2.$$

Invoking display (37), we arrive at

$$\left\|\tilde{Y}_h - \vartheta_{n+1}^{\mathsf{SO}}\right\|_{\mathbb{L}_2} \leqslant \left\|\tilde{Y}_h - x_{(n+1)h}\right\|_{\mathbb{L}_2} + \left\|\vartheta_{n+1}^{\mathsf{SO}} - x_{(n+1)h}\right\|_{\mathbb{L}_2}$$
$$\leqslant A_{n,1}e^{(L(T-nh)-\frac{1}{2})h}h^2\left\|Y_{nh}-\vartheta_n^{\mathsf{SO}}\right\|_{\mathbb{L}_2}+A_{n,2}e^{(L(T-nh)-\frac{1}{2})h}h^2$$
$$+\left[h\varepsilon_{sc}+\frac{2}{3}h^{3/2}\sqrt{d}\varepsilon_{sc}^{(L)}+\frac{1}{2}h^2\varepsilon_{sc}^{(M)}\right]e^{(L(T-nh)-\frac{1}{2})h}.$$

Furthermore, we can bound the coefficients $A_{n,1}$ and $A_{n,2}$ as follows,

$$A_{n,1} \leqslant \frac{1}{t^2}\int_0^t\left(\int_0^s 2(1+L_{\max})L_{\max}\,\mathrm{d}u\right)\mathrm{d}s = (1+L_{\max})L_{\max},$$
$$A_{n,2} \leqslant (1+L_{\max})L_{\max}C_2(n)+\sqrt{d}(\frac{1}{2}+L_{\max})L_{\max}^{1/2}+\frac{1}{2}L_{\max}C_4+\frac{\sqrt{2}}{4}\sqrt{d}L_F$$
$$\lesssim \sqrt{d}(\frac{1}{2}+L_{\max})L_{\max}^{1/2}+\frac{\sqrt{2}}{4}\sqrt{d}L_F.$$

Collecting all the pieces then gives

$$\left\|Y_{nh}-\vartheta_{n+1}^{\mathsf{SO}}\right\|_{\mathbb{L}_2} \leqslant r_n^{\mathsf{SO}}\left\|Y_{nh}-\vartheta_n^{\mathsf{SO}}\right\|_{\mathbb{L}_2}+C_n^{\mathsf{SO}}h^2+\left[h\varepsilon_{sc}+\frac{2}{3}h^{3/2}\varepsilon_{sc}^{(L)}+\frac{1}{2}h^2\varepsilon_{sc}^{(M)}\right]e^{(L(T-nh)-\frac{1}{2})h},$$

where

$$r_n^{\mathsf{SO}} = e^{-\int_{nh}^{(n+1)h}(m(T-t)-\frac{1}{2})\,\mathrm{d}t}+A_{n,1}e^{(L(T-nh)-\frac{1}{2})h}h^2,$$
$$C_n^{\mathsf{SO}} = A_{n,2}e^{(L(T-nh)-\frac{1}{2})h}.$$

□

From the result above, we finally obtain

$$\left\|Y_{Nh}-\vartheta_N^{\mathsf{SO}}\right\|_{\mathbb{L}_2} \lesssim \frac{1}{m_{\min}-1/2}\left[h\max_{0\leqslant k\leqslant N-1}C_k^{\mathsf{SO}}+\left(\varepsilon_{sc}+\frac{2}{3}h^{1/2}\varepsilon_{sc}^{(L)}+\frac{1}{2}h\varepsilon_{sc}^{(M)}\right)e^{(L_{\max}-\frac{1}{2})h}\right]$$
$$\lesssim h\cdot\frac{\sqrt{d}(L_{\max}^{3/2}+\sqrt{2}L_F/4)e^{(L_{\max}-\frac{1}{2})h}}{m_{\min}-1/2}+\left(\varepsilon_{sc}+\frac{2}{3}\sqrt{hd}\varepsilon_{sc}^{(L)}+\frac{1}{2}h\varepsilon_{sc}^{(M)}\right)e^{(L_{\max}-\frac{1}{2})h}.$$

This completes the proof of Theorem 4.

D Proof of Auxiliary Lemma

D.1 Proof of Lemma 8

We have

$$\begin{aligned}\frac{\mathrm{d}\left\|H_t-G_t\right\|^2}{\mathrm{d}t} &= 2\left\langle H_t-G_t,\, \frac{\mathrm{d}(H_t-G_t)}{\mathrm{d}t}\right\rangle \\ &= 2\left\langle H_t-G_t,\, \frac{1}{2}(H_t-G_t)+\left(\nabla\log p_{T-t}(H_t)-\nabla\log p_{T-t}(G_t)\right)\right\rangle \\ &= \left\|H_t-G_t\right\|^2+2\langle H_t-G_t,\nabla\log p_{T-t}(H_t)-\nabla\log p_{T-t}(G_t)\rangle \\ &\leqslant \left(1-2m(T-t)\right)\left\|H_t-G_t\right\|^2.\end{aligned} \tag{41}$$

The last inequality follows from Lemma 6. Then, we take the derivative of $e^{-\int_{t_1}^t(2m(T-s)-1)\,\mathrm{d}s}\left\|H_t-G_t\right\|^2$

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}\left(e^{\int_{t_1}^t(2m(T-s)-1)\,\mathrm{d}s}\left\|H_t-G_t\right\|^2\right) \\ &=(2m(T-t)-1)e^{-\int_{t_1}^t(2m(T-s)-1)\,\mathrm{d}s}\left\|H_t-G_t\right\|^2+e^{-\int_{t_1}^t(2m(T-s)-1)\,\mathrm{d}s}\frac{\mathrm{d}\left\|H_t-G_t\right\|^2}{\mathrm{d}t} \\ &\leqslant 0.\end{aligned}$$

Therefore, we obtain that

$$e^{\int_{t_1}^t(2m(T-s)-1)\,\mathrm{d}s}\left\|H_t-G_t\right\|^2\leqslant\left\|H_{t_t}-G_{t_1}\right\|^2.$$

taking the expectation of both sides and then applying the square root yields the desired result.

D.2 Proof of Lemma 10

By the definition of $\tilde{Y}_h$, we have

$$\begin{aligned}\left\|\tilde{Y}_t-\tilde{Y}_0\right\|_{\mathbb{L}_2} &= \left\|\int_0^t\left(\frac{1}{2}\tilde{Y}_s+\nabla\log p_{T-nh-s}(\tilde{Y}_s)\right)\mathrm{d}s+\int_{nh}^{nh+t}\mathrm{d}W_s\right\|_{\mathbb{L}_2} \\ &\leqslant \int_0^t\frac{1}{2}\left\|\tilde{Y}_s\right\|_{\mathbb{L}_2}\mathrm{d}t+\int_0^t\left\|\nabla\log p_{T-nh-s}(\tilde{Y}_s)\right\|_{\mathbb{L}_2}\mathrm{d}s+\left\|\int_{nh}^{nh+t}\mathrm{d}W_s\right\|_{\mathbb{L}_2}.\end{aligned}$$

To bound the first term, we observe that for any $s\in[0,h]$, the following holds

$$\begin{aligned}\left\|\tilde{Y}_s\right\|_{\mathbb{L}_2} &\leqslant \left\|Y_{nh+s}\right\|_{\mathbb{L}_2}+\left\|\tilde{Y}_s-Y_{nh+s}\right\|_{\mathbb{L}_2} \\ &\leqslant \left\|Y_{nh+s}-X^{\leftarrow}_{nh+s}\right\|_{\mathbb{L}_2}+\left\|X^{\leftarrow}_{nh+s}\right\|_{\mathbb{L}_2}+\left\|\tilde{Y}_s-Y_{nh+s}\right\|_{\mathbb{L}_2} \\ &\leqslant e^{-\int_0^{nh+s}(m(T-t)-\frac{1}{2})\,\mathrm{d}t}\left\|Y_0-X^{\leftarrow}_0\right\|_{\mathbb{L}_2}+\left\|X_{T-(nh+s)}\right\|_{\mathbb{L}_2}+e^{-\int_{nh}^{s}(m(T-u)-\frac{1}{2})\,\mathrm{d}u}\left\|Y_{nh}-\vartheta_n^{\mathsf{EM}}\right\|_{\mathbb{L}_2} \\ &\leqslant e^{-\int_0^{nh}(m(T-t)-\frac{1}{2})\,\mathrm{d}t}\left\|Y_0-X_T\right\|_{\mathbb{L}_2}+\sup_{0\leqslant t\leqslant T}\left\|X_t\right\|_{\mathbb{L}_2}+\left\|Y_{nh}-\vartheta_n^{\mathsf{EM}}\right\|_{\mathbb{L}_2}.\end{aligned}$$

Here, the second inequality follows from the Grönwall inequality applied on $\left\|Y_{nh+s}-X^{\leftarrow}_{nh+s}\right\|_{\mathbb{L}_2}$ and $\left\|\tilde{Y}_s-Y_{nh+s}\right\|_{\mathbb{L}_2}$, and the fact that $\left\|X_t\right\|_{\mathbb{L}_2}=\left\|X^{\leftarrow}_{T-t}\right\|_{\mathbb{L}_2}$. To bound the second term, we need the following lemma.

Lemma 18. *If the target distribution p_0 satisfies Assumption 1, it holds that*

$$\left\|\nabla\log p_t(X_t)\right\|_{\mathbb{L}_2}\leqslant(dL(t))^{1/2}.$$

According to Lemma 18, it follows that

$$
\begin{aligned}
&\left\|\nabla \log p_{T-nh-s}(\tilde{Y}_s)\right\|_{\mathbb{L}_2} \\
\leqslant\ & \left\|\nabla \log p_{T-nh-s}(\tilde{Y}_s) - \nabla \log p_{T-nh-s}(X^{\leftarrow}_{nh+s})\right\|_{\mathbb{L}_2} + \left\|\nabla \log p_{T-nh-s}(X^{\leftarrow}_{nh+s})\right\|_{\mathbb{L}_2} \\
\leqslant\ & L(T-nh-s)\left\|\tilde{Y}_s - X^{\leftarrow}_{nh+s}\right\|_{\mathbb{L}_2} + (dL(T-nh-s))^{1/2} \\
\leqslant\ & L(T-nh-s)\left\|\tilde{Y}_0 - X^{\leftarrow}_{nh}\right\|_{\mathbb{L}_2} + (dL(T-nh-s))^{1/2} \\
\leqslant\ & L(T-nh-s)\left(\left\|\tilde{Y}_0 - Y_{nh}\right\|_{\mathbb{L}_2} + \|Y_{nh} - X^{\leftarrow}_{nh}\|_{\mathbb{L}_2}\right) + (dL(T-nh-s))^{1/2} \\
\leqslant\ & L(T-nh-s)\left(\|Y_{nh} - \vartheta^{\mathsf{EM}}_n\|_{\mathbb{L}_2} + e^{-\int_0^{nh}(m(T-t)-\frac{1}{2})\,\mathrm{d}t}\,\|Y_0 - X_T\|_{\mathbb{L}_2}) + (dL(T-nh-s)\right)^{1/2}.
\end{aligned}
$$

Here, we use the fact that $\tilde{Y}_0 = \vartheta^{\mathsf{EM}}_n$, and Grönwall inequality are used in the third inequality and the last one. This completes the proof.

D.3 Proof of Lemma 13

For the stochastic integral of process X, we have

$$
\mathbb{E}(I_t(X))^2 = \mathbb{E}\int_0^t X_u^2 d\langle M\rangle_u.
$$

Then, we obtain

$$
\begin{aligned}
&\left\|\int_{nh}^{(n+U_n)h} \mathrm{d}W_t - \int_0^1\left(\int_{nh}^{(n+u)h} \mathrm{d}W_t\right)\mathrm{d}u\right\|^2_{\mathbb{L}_2} \\
=\ & \mathbb{E}\left(\int_0^1\int_{nh}^{(n+1)h} -\mathbf{1}_{\{U_n\leqslant u\}}\mathbf{1}_{\{(n+U_n)h\leqslant t\leqslant(n+u)h\}} + \mathbf{1}_{\{U_n>u\}}\mathbf{1}_{\{(n+u)h\leqslant t\leqslant(n+U_n)h\}}\,\mathrm{d}W_t\,\mathrm{d}u\right)^2 \\
\leqslant\ & \int_0^1 \mathbb{E}\left(\int_{nh}^{(n+1)h} -\mathbf{1}_{\{U_n\leqslant u\}}\mathbf{1}_{\{(n+U_n)h\leqslant t\leqslant(n+u)h\}} + \mathbf{1}_{\{U_n>u\}}\mathbf{1}_{\{(n+u)h\leqslant t\leqslant(n+U_n)h\}}\,\mathrm{d}W_t\right)^2 \mathrm{d}u \\
=\ & \int_0^1\left(\mathbb{E}\int_{nh}^{(n+1)h} \mathbf{1}_{\{U_n\leqslant u\}}\mathbf{1}_{\{(n+U_n)h\leqslant t\leqslant(n+u)h\}} + \mathbf{1}_{\{U_n>u\}}\mathbf{1}_{\{(n+u)h\leqslant t\leqslant(n+U_n)h\}}\,\mathrm{d}t\right)\mathrm{d}u \\
=\ & \int_0^1 \left(\mathbb{E}\left(\mathbf{1}_{\{U_n\leqslant u\}}(u-U_n)h + \mathbf{1}_{\{U_n>u\}}(U_n-u)h\right)\right)\mathrm{d}u \\
=\ & h\int_0^1\left(u^2 - u + \frac{1}{2}\right)\mathrm{d}u \\
=\ & \frac{1}{3}h.
\end{aligned}
$$

D.4 Proof of Lemma 16

Define the function $w(s)$ via

$$
w(s) = \exp\left(-\int_{t_0}^s \beta(r)dr\right)\int_{t_0}^s \beta(r)z(r)\,\mathrm{d}r, \quad \forall s \geqslant t_0.
$$

Differentiating this function gives

$$
w'(s) = \left(z(s) - \int_{t_0}^s \beta(r)z(r)\,\mathrm{d}r\right)\beta(s)\exp\left(-\int_{t_0}^s \beta(r)\,\mathrm{d}r\right) \leqslant \alpha(s)\beta(s)\exp\left(-\int_{t_0}^s \beta(r)\,\mathrm{d}r\right).
$$

Note that $w(t_0) = 0$. Integrating the function w from t_0 to t yields

$$w(t) \leqslant \int_{t_0}^{t} \alpha(s)\beta(s) \exp\left(-\int_{t_0}^{s} \beta(r)\,\mathrm{d}r\right) \mathrm{d}s.$$

By the definition of $w(s)$, we also have

$$\int_{t_0}^{t} \beta(s)z(s)\,\mathrm{d}s = \exp\left(\int_{t_0}^{t} \beta(r)\,\mathrm{d}r\right) w(t).$$

Combining the previous two displays provides us with

$$\int_{t_0}^{t} \beta(s)z(s)\,\mathrm{d}s \leqslant \int_{t_0}^{t} \alpha(s)\beta(s) \exp\left(\int_{s}^{t} \beta(r)\,\mathrm{d}r\right) \mathrm{d}s.$$

By substituting this estimate into the inequality, we can obtain the first desired result. Furthermore, if α is non-decreasing, then for any $s \leqslant t$, it holds that $\alpha(s) \leqslant \alpha(t)$. This leads to

$$z(t) \leqslant \alpha(t) + \alpha(t) \int_{t_0}^{t} \beta(s) \exp\left(\int_{s}^{t} \beta(r)\,\mathrm{d}r\right) \mathrm{d}s.$$

which can be simplified to

$$z(t) \leqslant \alpha(t) \exp\left(\int_{t_0}^{t} \beta(r)\,\mathrm{d}r\right), \quad t \geqslant t_0.$$

This completes the proof.

D.5 Proof of Lemma 17

Notice that

$$\begin{aligned}
\nabla^2 \log p_t(x) =& -\frac{1}{p_t(x)^2} \nabla p_t(x) \nabla p_t(x)^\top + \frac{1}{p_t(x)} \nabla^2 p_t(x) \\
=& -\nabla \log p_t(x) \nabla \log p_t(x)^\top + \frac{1}{p_t(x)} \nabla^2 p_t(x),
\end{aligned}$$

which indicates

$$\begin{aligned}
\frac{1}{2} \sum_{i=1}^{d} \frac{\partial^2 p_t(x)}{\partial x_i^2} \cdot \frac{1}{p_t(x)} &= \frac{1}{2} \operatorname{Tr}\left(\frac{1}{p_t(x)} \nabla^2 p_t(x)\right) \\
&= \frac{1}{2} \operatorname{Tr}\left(\nabla^2 \log p_t(x) + \nabla \log p_t(x) \nabla \log p_t(x)^\top\right) \\
&= \frac{1}{2} \operatorname{Tr}\left(\nabla^2 \log p_t(x)\right) + \frac{1}{2} \left\|\nabla \log p_t(x)\right\|^2,
\end{aligned}$$

Additionally, we have

$$\begin{aligned}
\nabla\left(\frac{\partial^2 \log p_t(x)}{\partial x_i^2}\right) &= \nabla\left(\frac{\partial^2 p_t(x)}{\partial x_i^2} \cdot \frac{1}{p_t(x)} - \left(\frac{\partial p_t(x)}{\partial x_i} \cdot \frac{1}{p_t(x)}\right)^2\right) \\
&= \nabla\left(\frac{\partial^2 p_t(x)}{\partial x_i^2}\right) \cdot \frac{1}{p_t(x)} - \frac{\partial^2 p_t(x)}{\partial x_i^2} \cdot \frac{1}{p_t(x)} \cdot \nabla \log p_t(x) - \nabla\left(\left(\frac{\partial \log p_t(x)}{\partial x_i}\right)^2\right).
\end{aligned}$$

Then, we obtain

$$\begin{aligned}
&\nabla\left(\sum_{i=1}^{d} \frac{\partial^2 p_t(x)}{\partial x_i^2}\right) \cdot \frac{1}{p_t(x)} \\
=&\nabla\left(\operatorname{Tr}(\nabla^2 \log p_t(x))\right) + \left[\operatorname{Tr}(\nabla^2 \log p_t(x)) + \|\nabla \log p_t(x)\|^2\right] \cdot \nabla \log p_t(x) + \nabla(\|\nabla \log p_t(x)\|^2).
\end{aligned}$$

D.6 Proof of Lemma 18

Note that

$$\begin{aligned}\mathbb{E}(\|\nabla \log p_t(X_t)\|^2) &= \int_{\mathbb{R}^d} \|\nabla \log p_t(x)\|^2 \, p_t(x)\,\mathrm{d}x \\ &= \lim_{R\to\infty} \int_{B(0,R)} \langle \nabla \log p_t(x), \nabla \log p_t(x)\rangle p_t(x)\,\mathrm{d}x \\ &= \lim_{R\to\infty} \int_{B(0,R)} \langle \nabla \log p_t(x), \nabla p_t(x)\rangle\,\mathrm{d}x\,,\end{aligned}$$

where $B(0,R)$ denotes the Euclidean ball with radius $R>0$ centered at the origin. Using integration by parts, we then obtain

$$\begin{aligned}\mathbb{E}(\|\nabla \log p_t(X_t)\|^2) &= \lim_{R\to\infty} \int_{B(0,R)} -p_t(x)\Delta \log p_t(x)\,\mathrm{d}x + \int_{\partial B(0,R)} p_t(x)\frac{\partial \log p_t(x)}{\partial \vec{n}}\,\mathrm{d}S \\ &= \int_{\mathbb{R}^d} p_t(x)\cdot(-\Delta \log p_t(x))\,\mathrm{d}x \\ &\leqslant dL(t),\end{aligned}$$

where $\dfrac{\partial f}{\partial \vec{n}} = \nabla f \cdot \vec{n}$ represents the directional derivative along the normal vector $\vec{n}$ and $\mathrm{d}S$ denotes the surface integral over the spherical surface. Here we use the fact that $p_t(x)$ converges to 0 at an exponential rate as $\|x\|$ approaches infinity, and the fact that

$$-\Delta \log p_t(x) = -\operatorname{Tr}(\nabla^2 \log p_t(x)) \in [0, dL(t)]\,,$$

which follows from Lemma 6.

E Details for second-order acceleration

In this section, we present a complete derivation of the second-order acceleration scheme, detailing the implementation of Itô-Taylor expansions and Itô's formula. Building upon the general backward process framework

$$dx_t = \gamma(T-t, x_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t\,, \tag{42}$$

where $\sigma > 0$ and W_t is the d-dimensional Brownian motion. We apply Itô's formula to $\gamma(T-t,x)$. This procedure generates an approximated structure of SDE (42),

$$dx_t = [\gamma(T-s, x_s) + L_s(x_t - x_s) + M_s(t-s)]\,\mathrm{d}t + \sigma\,\mathrm{d}W_t. \tag{43}$$

with

$$L_s = \frac{\partial \gamma}{\partial x}(T-s, x_s) \quad \text{and} \quad M_s = \frac{\sigma^2}{2}\frac{\partial^2 \gamma}{\partial x^2}(T-s, x_s) - \frac{\partial \gamma}{\partial t}(T-s, x_s)$$

which serves as the foundation for our subsequent second-order discretization.In Section 4, we demonstrate that this approximation preserves the core dynamical structure of the original SDE (42) while admitting a closed-form solution. This is achieved by replacing the intractable drift term $\gamma(T-t, x_s)$ with its Itô-expanded counterpart, which remains analytical tractable through explicit integration.

Applying Itô's formula to $e^{-L_s t}x_t$ yields

$$\mathrm{d}(e^{-L_s t}x_t) = e^{-L_s t}(\gamma(T-s, x_s) - L_s x_s + M_s(t-s))\,\mathrm{d}t + e^{-L_s t}\sigma\,\mathrm{d}W_t.$$

For fixed s, both sides of the equation permit closed-form integration. While the Brownian integral $\int_s^{s+\Delta t} e^{-L_s t}\sigma\,\mathrm{d}W_t$ formally appears non-analytic, it equivalently manifests as a Gaussian random variable with explicitly computable variance. This enables full analytical representation for $x_{s+\Delta t}$

when integrating over $[s, s+\Delta t]$,

$$\begin{aligned} x_{s+\Delta t} = e^{L_s \Delta t} x_s &+ \int_s^{s+\Delta t} e^{L_s(s+\Delta t - t)} \, \mathrm{d}t (\gamma(T-s, x_s) - L_s x_s) \\ &+ \int_s^{s+\Delta t} e^{L_s(s+\Delta t - t)} (t-s) \, \mathrm{d}t \, M_s + \sigma \int_s^{s+\Delta t} e^{L_s(s+\Delta t - t)} \, \mathrm{d}W_t \\ &= x_s + L_s^{-1}(e^{L_s \Delta t} - 1)\gamma(T-s, x_s) + L_s^{-2}\big[(e^{L_s \Delta t} - 1) - L_s \Delta t\big] M_s \\ &+ \sigma \int_s^{s+\Delta t} e^{L_s(s+\Delta t - u)} dW_t \, . \end{aligned}$$

Having established the general framework, we now specialize to our core case through the parameterization: set $\gamma(T-t, x) = \frac{1}{2}x + \nabla \log p_{T-t}(x), \sigma = 1$, let $\Delta t \in [0, h]$ and $s = nh$.

In the resulting expression, we denote x_s by $\vartheta_n^{\mathsf{SO}}$. Then for any $t \in [nh, (n+1)h]$, the solution admits the semi-analytic representation

$$\begin{aligned} x_t = \vartheta_n^{\mathsf{SO}} + \int_{nh}^t (\frac{1}{2}\vartheta_n^{\mathsf{SO}} + \nabla \log p_{T-nh}(\vartheta_n^{\mathsf{SO}}) \\ + L_n(x_u - \vartheta_n^{\mathsf{SO}}) + M_n(u-nh)) \, \mathrm{d}u + \int_{nh}^t \mathrm{d}W_u \end{aligned} \tag{44}$$

where

$$\begin{aligned} L_n &= \frac{1}{2} I_d + \nabla^2 \log p_{T-nh}(\vartheta_n^{\mathsf{SO}}) \\ M_n &= \frac{1}{2}\sum_{j=1}^d \frac{\partial^2}{\partial x_j^2} \nabla \log p_{T-nh}(\vartheta_n^{\mathsf{SO}}) - \frac{\partial}{\partial t} \nabla \log p_{T-nh}(\vartheta_n^{\mathsf{SO}}) \, . \end{aligned}$$

Though L_n and M_n are theoretically defined through exact derivatives in SDE (43), their practical evaluation requires approximations due to the score function's computational intractability. We implement these approximations via numerical methods or neural networks, with concrete techniques for L_n and M_n estimation provided separately in Appendix F. Substituting these approximations into the SDE (44) yields

$$\begin{aligned} x_t = \vartheta_n^{\mathsf{SO}} + \int_{nh}^t \Big(\gamma(T-nh, \vartheta_n^{\mathsf{SO}}) + s_*^{(L)}(T-nh, \vartheta_n^{\mathsf{SO}})(x_u - \vartheta_n^{\mathsf{SO}}) \\ + s_*^{(M)}(T-nh, \vartheta_n^{\mathsf{SO}})(u-nh) \Big) \, \mathrm{d}u + \int_{nh}^t \mathrm{d}W_u \, . \end{aligned}$$

Crucially, this substitution preserves the closed-form integrability of the original framework. Adopting the same exponential integration strategy as above, we derive a closed form of x_t. Let $\vartheta_{n+1}^{\mathsf{SO}}$ denote $x_{(n+1)h}$, the second-order discretization scheme is given by

$$\begin{aligned} \vartheta_{n+1}^{\mathsf{SO}} = \vartheta_n^{\mathsf{SO}} &+ s_*^{(L)}(T-nh, \vartheta_n^{\mathsf{SO}})^{-1} \left(e^{s_*^{(L)}(T-nh, \vartheta_n^{\mathsf{SO}})h} - I_d \right) \left(\frac{1}{2}\vartheta_n^{\mathsf{SO}} + s_*(T-nh, \vartheta_n^{\mathsf{SO}}) \right) \\ &+ s_*^{(L)}(T-nh, \vartheta_n^{\mathsf{SO}})^{-2} \left(e^{s_*^{(L)}(T-nh, \vartheta_n^{\mathsf{SO}})h} - s_*^{(L)}(T-nh, \vartheta_n^{\mathsf{SO}})h - I_d \right) s_*^{(M)}(T-nh, \vartheta_n^{\mathsf{SO}}) \\ &+ \int_{nh}^{(n+1)h} e^{s_*^{(L)}(T-nh, \vartheta_n^{\mathsf{SO}})[(n+1)h - t]} \, \mathrm{d}W_t \, . \end{aligned}$$

Implementation specifics for handling the matrix exponentials and stochastic integral are addressed in Appendix G.

F Numerical Studies on Synthetic Data

We apply the five schemes to the posterior density of penalized logistic regression, defined by $p_0(\theta) \propto \exp(-f(\theta))$ with the potential function

$$f(\theta) = \frac{\lambda}{2}\|\theta\|^2 + \frac{1}{n_{\mathsf{data}}} \sum_{i=1}^{n_{\mathsf{data}}} \log(1 + \exp(-y_i x_i^\top \theta)),$$

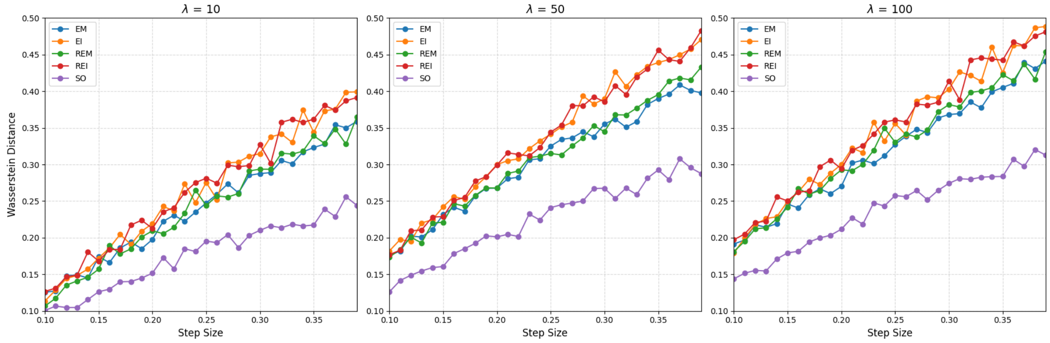

Figure 2: Error of various discretization schemes and second-order sampler with different choice of step size.

where $\lambda > 0$ denotes the tuning parameter. The data $\{x_i, y_i\}_{i=1}^{n_{\text{data}}}$, composed of binary labels $y_i \in \{-1, 1\}$ and features $x_i \in \mathbb{R}^d$ generated from $x_{i,j} \overset{iid}{\sim} \mathcal{N}(0, 100)$.

F.1 Implementation Details

In the numerical studies, we set $T = 10$, and the number of Monte Carlo iterations is chosen as the floor of T/h, where h varies according to the step size indicated in the figure. Figure 2 shows the Wasserstein distance measured along the first dimension between the empirical distributions of the N-th outputs from SGMs and the target distribution, with different choices of the step size h. In this simulation, we use the Monte-Carlo method to estimate the score function and the Hessian matrix.

F.2 Calculation

In this part, we derive explicit formulas for each coefficient term we need. First, the score function can be computed as

$$\begin{aligned}\nabla \log p_0(\theta) &= -\left(\lambda\theta + \frac{1}{n_{\text{data}}}\sum_{i=1}^{n_{\text{data}}} \frac{-y_i x_i \exp(-y_i x_i^\top \theta)}{1 + \exp(-y_i x_i^\top \theta)}\right) \\ &= -\left(\lambda\theta + \frac{1}{n_{\text{data}}}\sum_{i=1}^{n_{\text{data}}} \frac{-y_i x_i}{1 + \exp(y_i x_i^\top \theta)}\right).\end{aligned}$$

For simplicity, we denote the logistic sigmoid function $\sigma(u) = \dfrac{1}{1+e^{-u}}$, then

$$\nabla \log p_0(\theta) = -\left(\lambda\theta + \frac{1}{n_{\text{data}}}\sum_{i=1}^{n_{\text{data}}} -y_i x_i \sigma(-y_i x_i^\top \theta)\right).$$

Since $\sigma'(u) = \sigma(u)[1 - \sigma(u)]$, we have

$$\begin{aligned}\nabla^2 \log p_0(\theta) &= -\left(\lambda I_d + \frac{1}{n_{\text{data}}}\sum_{i=1}^{n_{\text{data}}} y_i^2 \sigma(-y_i x_i^\top \theta)\left[1 - \sigma(-y_i x_i^\top \theta)\right] x_i x_i^\top\right) \\ &= -\lambda I_d - \frac{1}{n_{\text{data}}}\sum_{i=1}^{n_{\text{data}}} \sigma(-y_i x_i^\top \theta)\left[1 - \sigma(-y_i x_i^\top \theta)\right] x_i x_i^\top.\end{aligned}$$

As $x_i x_i^\top \succcurlyeq 0$, $\nabla^2 \log p_0(\theta) \preccurlyeq -\lambda I_d$. We also have that $\sigma(1 - y_i x_i^\top \theta) \in (0, 1)$, then

$$\begin{aligned}\nabla^2 \log p_0(\theta) &\succcurlyeq -\lambda I_d - \frac{1}{4n_{\text{data}}}\sum_{i=1}^{n_{\text{data}}} x_i x_i^\top \\ &\succcurlyeq -(\lambda + \frac{1}{n_{\text{data}}}\lambda_{\max}(\sum_{i=1}^{n_{\text{data}}} x_i x_i^\top)) I_d\,.\end{aligned}$$

Therefore,

$$m_0 = \lambda, \quad L_0 = \lambda + \frac{1}{n_{\mathsf{data}}}\lambda_{\max}(\sum_i^{n_{\mathsf{data}}} x_i x_i^\top).$$

Recall that the transition probability $p_{t|0}(\theta_t|\theta_0) = \phi(\theta_t; \mu_t, \Sigma_t)$, where $\mu_t = e^{-\frac{1}{2}t}\theta_0, \Sigma_t = (1 - e^{-t})I_d$, and $\phi(\theta, \mu, \Sigma)$ denotes the probability density function of $\mathcal{N}(\mu, \Sigma)$, then we have

$$\begin{aligned}
p_t(\theta_t) &= \int_{\mathbb{R}^d} p_{t|0}(\theta_t|\theta_0) p_0(\theta_0)\, \mathrm{d}\theta_0 \\
&= \int_{\mathbb{R}^d} \frac{1}{\sqrt{(2\pi)^d |\Sigma_t|}} \exp(-\frac{1}{2}(\theta_t - \mu_t)^\top \Sigma_t^{-1}(\theta_t - \mu_t)) p_0(\theta_0)\, \mathrm{d}\theta_0 \\
&= \int_{\mathbb{R}^d} \frac{1}{[2\pi(1-e^{-t})]^{d/2}} \exp(-\frac{1}{2(1-e^{-t})}\|\theta_t - e^{-\frac{1}{2}t}\theta_0\|^2) p_0(\theta_0)\, \mathrm{d}\theta_0 \\
&= \frac{1}{[2\pi(1-e^{-t})]^{d/2}} \mathbb{E}_{\theta_0 \sim p_0}\left[\exp(-\frac{1}{2(1-e^{-t})}\|\theta_t - e^{-\frac{1}{2}t}\theta_0\|^2)\right].
\end{aligned}$$

Hence,

$$\begin{aligned}
\nabla p_t(\theta_t) &= \frac{1}{[2\pi(1-e^{-t})]^{d/2}} \mathbb{E}_{\theta_0 \sim p_0}\left[\nabla\left(\exp(-\frac{1}{2(1-e^{-t})}\|\theta_t - e^{-\frac{1}{2}t}\theta_0\|^2)\right)\right] \\
&= \frac{1}{[2\pi(1-e^{-t})]^{d/2}} \mathbb{E}_{\theta_0 \sim p_0}\left[\exp(-\frac{1}{2(1-e^{-t})}\|\theta_t - e^{-\frac{1}{2}t}\theta_0\|^2) \cdot \frac{-(\theta_t - e^{-\frac{1}{2}t}\theta_0)}{1-e^{-t}}\right], \\
\nabla^2 p_t(\theta_t) &= \frac{1}{[2\pi(1-e^{-t})]^{d/2}} \mathbb{E}_{\theta_0 \sim p_0}\Bigg[\exp(-\frac{1}{2(1-e^{-t})}\|\theta_t - e^{-\frac{1}{2}t}\theta_0\|^2) \\
&\qquad \cdot \left(\frac{(\theta_t - e^{-\frac{1}{2}t}\theta_0)(\theta_t - e^{-\frac{1}{2}t}\theta_0)^\top}{(1-e^{-t})^2} - \frac{1}{1-e^{-t}} I_d\right)\Bigg].
\end{aligned}$$

We can approximate $p_t(\theta_t), \nabla p_t(\theta_t)$ and $\nabla^2 p_t(\theta_t)$ or even higher order derivative tensor of $p_t(\theta_t)$ by Monte Carlo method, therefore, we can compute score function and its high order derivative by

$$\nabla \log p_t(\theta_t) = \frac{\nabla p_t(\theta_t)}{p_t(\theta_t)}, \quad \nabla^2 \log p_t(\theta_t) = \frac{\nabla^2 p_t(\theta_t)}{p_t(\theta_t)} - \frac{\nabla p_t(\theta_t) \nabla p_t(\theta_t)^\top}{p_t(\theta_t)^2}.$$

G Real Data Analysis

G.1 Implementation Details

We set the step size $h = 0.2$ and $N = 2/h$. We conducted experiments on an NVIDIA RTX 4060 GPU (16GB VRAM). The training process required 2 GPU hours over 100 epochs with a batch size of 32, using CUDA 12.4, PyTorch 2.4, and torchvision 0.20.0. Figure 3 shows the digits generated by five algorithms, using the same score functions. The execution times for the algorithms are as follows: EM method 2 hour 12 min 49 s, EI method 2 hour 12 min 50 s, REM method 2 hour 13 min 30 s, REI method 2 hour 13 min 47 s, SO method 2 hour 14 min 05 s.

G.2 Score matching function for second order acceleration

For the MNIST dataset, we have demonstrated in the proof of Proposition 15 that computing third-order derivatives is unnecessary. Unlike existing high-order methods for estimating second-order scores [28], which require the joint training of score functions and Hessian matrices and consequently incur substantial computational overhead, our second-order algorithm avoids explicit computation of the Jacobian matrix. Furthermore, by employing Hessian-vector products (HVPs), we efficiently capture higher-order information, enabling our second-order acceleration method to achieve improved performance with reduced iteration complexity and manageable computational cost.

More specifically, in the experiments of the MNIST dataset, we construct a U-Net architecture incorporating time and label embeddings to train the score function, where the time embedding operates

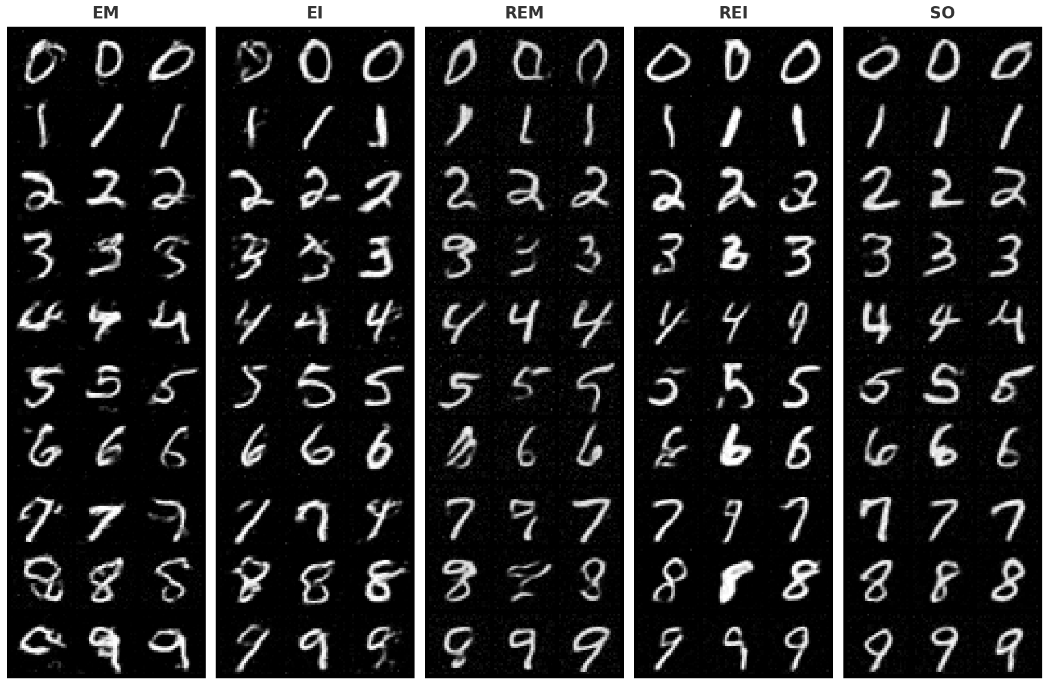

Figure 3: Comparative visualization of generated MNIST digits under various discretization schemes.

on the temporal variable t of the score function, while the label embedding leverages MNIST's categorical digit labels. This conditional formulation expresses the score function as $\nabla \log p(t, x|\text{label})$, enabling per-class score estimation through discriminative embedding propagation.

Recall iteration rule of the SO algorithm, we have

$$\begin{aligned}\vartheta_{n+1}^{\mathsf{SO}} = \vartheta_n^{\mathsf{SO}} &+ s_*^{(L)}(T-nh, \vartheta_n^{\mathsf{SO}})^{-1}\left(e^{s_*^{(L)}(T-nh,\vartheta_n^{\mathsf{SO}})h} - I_d\right)\left(\frac{1}{2}\vartheta_n^{\mathsf{SO}} + s_*(T-nh, \vartheta_n^{\mathsf{SO}})\right)\\ &+ s_*^{(L)}(T-nh, \vartheta_n^{\mathsf{SO}})^{-2}\left(e^{s_*^{(L)}(T-nh,\vartheta_n^{\mathsf{SO}})h} - s_*^{(L)}(T-nh, \vartheta_n^{\mathsf{SO}})h - I_d\right) s_*^{(M)}(T-nh, \vartheta_n^{\mathsf{SO}})\\ &+ \int_{nh}^{(n+1)h} e^{s_*^{(L)}(T-nh,\vartheta_n^{\mathsf{SO}})[(n+1)h-t]}\,\mathrm{d}W_t\,.\end{aligned}$$

Note that

$$\int_{nh}^{(n+1)h} e^{s_*^{(L)}(T-nh,\vartheta_n^{\mathsf{SO}})[(n+1)h-t]}\,\mathrm{d}W_t \sim \mathcal{N}\left(0, \frac{1}{2}s_*^{(L)}(T-nh, \vartheta_n^{\mathsf{SO}})^{-1}\left(e^{2s_*^{(L)}(T-nh,\vartheta_n^{\mathsf{SO}})h} - I_d\right)\right).$$

Let $g_n(\cdot) := s_*(T-nh, \cdot)$ denote the score matching function at time $T-nh$. Although the approximation of the Hessian matrix $\nabla^2 \log p_{T-nh}(\cdot)$ will not explicitly appear in the algorithmic implementation, we formally designate it as $H_n(\cdot)$ for notational clarity. Consequently, the estimators of L_n and M_n are chosen to be

$$\begin{aligned}s_*^{(L)}(T-nh, \vartheta_n^{\mathsf{SO}}) &:= \frac{1}{2}I_d + H_n(\vartheta_n^{\mathsf{SO}})\\ s_*^{(M)}(T-nh, \vartheta_n^{\mathsf{SO}}) &:= -\frac{1}{2}g_n(\vartheta_n^{\mathsf{SO}}) - H_n(\vartheta_n^{\mathsf{SO}})(\frac{1}{2}\vartheta_n^{\mathsf{SO}} + g_n(\vartheta_n^{\mathsf{SO}})).\end{aligned}$$

Employing the Taylor expansion, we have

$$\begin{aligned}(s_*^{(L)})^{-1}\left(e^{hs_*^{(L)}} - I_d\right) &= \sum_{k=1}^{\infty} \frac{h^k (s_*^{(L)})^{k-1}}{k!},\\ (s_*^{(L)})^{-2}\left(e^{hs_*^{(L)}} - hs_*^{(L)} - I_d\right) &= \sum_{k=2}^{\infty} \frac{h^k (s_*^{(L)})^{k-2}}{k!},\end{aligned}$$

$$\left[\frac{1}{2}(s_*^{(L)})^{-1}\left(e^{2hs_*^{(L)}} - I_d\right)\right]^{1/2} = \sqrt{h}\sum_{k=0}^{\infty} a_k (hs_*^{(L)})^k,$$

where $(a_0, a_1, a_2, a_3, a_4, a_5, a_6 \cdots) = (1, \frac{1}{2}, \frac{5}{24}, \frac{1}{16}, \frac{79}{5760}, \frac{3}{1280}, \frac{71}{193536}, \cdots)$. We thus reformulate all operators in the discretization scheme using matrix multiplications, which is a crucial step that avoids the explicit computation and storage of the full Hessian matrix H_t. This is achieved by leveraging Hessian-vector products (HVPs) via automatic differentiation, which reduces memory complexity to $\mathcal{O}(d)$ while retaining second-order curvature information. Specifically, given that g_n corresponds to the neural network's output and H_n represents its Jacobian matrix, we compute $H_n v$ for any vector v through PyTorch's reverse-mode differentiation (`torch.autograd.grad`). By iteratively applying this HVP procedure k times, we efficiently construct $H_n^k v$ for any $k \geqslant 0$. Through Taylor series expansion, these HVP-powerd computations enable precise evaluation of each term in the discretization scheme.