

NON-ASYMPTOTIC ESTIMATES FOR ACCELERATED HIGH ORDER LANGEVIN MONTE CARLO ALGORITHMS

ARIEL NEUFELD AND YING ZHANG

ABSTRACT. In this paper, we propose two new algorithms, namely, aHOLA and aHOLLA, to sample from high-dimensional target distributions with possibly super-linearly growing potentials. We establish non-asymptotic convergence bounds for aHOLA in Wasserstein-1 and Wasserstein-2 distances with rates of convergence equal to $1 + q/2$ and $1/2 + q/4$, respectively, under a local Hölder condition with exponent $q \in (0, 1]$ and a convexity at infinity condition on the potential of the target distribution. Similar results are obtained for aHOLLA under certain global continuity conditions and a dissipativity condition. Crucially, we achieve state-of-the-art rates of convergence of the proposed algorithms in the non-convex setting which are higher than those of the existing algorithms. Examples from high-dimensional sampling and logistic regression are presented, and numerical results support our main findings.

1. INTRODUCTION

We consider the problem of sampling from a high-dimensional target distribution

$$\pi_\beta(d\theta) \propto \exp(-\beta U(\theta)) d\theta, \quad (1)$$

where $\theta \in \mathbb{R}^d$, $\beta > 0$ is the so-called inverse temperature parameter, and $U : \mathbb{R}^d \rightarrow \mathbb{R}$ is some (smooth enough) function. The distribution π_β in (1) can be viewed as the invariant measure of the Langevin stochastic differential equation (SDE) given by

$$Z_0 := \theta_0, \quad dZ_t = -h(Z_t) dt + \sqrt{2\beta^{-1}} dB_t, \quad t \geq 0, \quad (2)$$

where θ_0 is an \mathbb{R}^d -valued random variable, $h := \nabla U$, and $(B_t)_{t \geq 0}$ is a d -dimensional Brownian motion. Thus, to sample from π_β , one may consider using algorithms that track (2). Widely used algorithms of this type include the unadjusted Langevin algorithm (ULA) (or the Langevin Monte Carlo (LMC) algorithm) and the stochastic gradient Langevin dynamics (SGLD) algorithm, which are the Euler discretization of (2). Theoretical guarantees for ULA and SGLD to sample approximately from π_β have been well established in the literature under the conditions that the (stochastic) gradient of the potential of π_β is globally Lipschitz continuous and is strongly convex, see [2, 3, 10, 11, 13, 14]. Recent research focuses on the analysis of ULA and SGLD under weaker conditions so as to accommodate a variety of distributions. To relax the strong convexity condition of h , [6, 7, 32, 34, 38, 42, 45, 46] considered replacing it with certain (local) dissipativity or convexity at infinity condition, and obtain convergence results using techniques developed in [15, 16]. To relax the global Lipschitz condition (in θ) and replace it with a local Lipschitz (or Hölder) condition, certain techniques need to be applied to modify the algorithms (see, e.g., [21, 28, 30, 39] and references therein). This is due to the fact that the absolute moments of the aforementioned algorithms could diverge to infinity at a finite time point [20]. In [4], a taming technique proposed in [21, 39] is applied to obtain the tamed ULA algorithm (TULA) and convergence results are obtained under a local Lipschitz condition. Several variants of ULA and SGLD have been developed by applying the taming techniques while their convergence results are established under relaxed conditions to accommodate π_β with super-linearly growing potentials [25, 26, 27, 29, 33]. We also refer to [1, 5, 8, 12, 19, 24, 31] for high-order Langevin algorithms, to [9, 17, 32, 35, 43] for sampling guarantees under functional inequalities, and to [36, 44] for inverse problem sampling.

The algorithms mentioned above are first-order methods which make use of the (stochastic) gradient of the potential of π_β . The state-of-the-art rates of convergence of these algorithms in Wasserstein-2 distances are shown to be 1 in the convex case while they are 1/2 in the non-convex case, which are obtained under certain conditions imposed on the Hessian of the potential. In [11], an LMC algorithm

Key words and phrases. Sampling problem, non-asymptotic estimates, taming technique, super-linearly growing coefficients, high order algorithm.

Financial support by the MOE AcRF Tier 2 Grant MOE-T2EP20222-0013 and the Guangzhou-HKUST(GZ) Joint Funding Programs (No. 2024A03J0630 and No. 2025A03J3322) is gratefully acknowledged.

with Ozaki discretization (LMCO) and its variant LMCO' are proposed, which are second-order methods making use of the Hessian of the potential. It is shown in [11] that second-order methods improve on the first-order methods in ill-conditioned cases. However, no improvements are made on the rates of convergences of these methods in Wasserstein distances. In [40], a high order LMC algorithm (HOLA) is developed by applying a taming technique to an order 1.5 numerical scheme introduced in [22]. HOLA makes use of the third derivative of the potential of π_β , and [40] shows that the rate of convergence of HOLA in Wasserstein-2 distance in the convex setting is $3/2$, which is higher than that of the first and second-order methods.

In this paper, we mainly consider an algorithm that can be used to sample from π_β with a super-linearly growing potential. To this end, we propose an accelerated HOLA algorithm (aHOLA), which is a variant of HOLA in [40] obtained using a new taming factor. For completeness, we also propose an accelerated high order linear LMC algorithm (aHOLLA) which is the counterpart of aHOLA in the linear case (i.e., in the case where the derivatives of the potential of π_β are growing at most linearly). Crucially, we obtain non-asymptotic error bounds for aHOLA in Wasserstein distances under a local Hölder condition with exponent $q \in (0, 1]$ and a convexity at infinity condition. Our results are applicable to various distributions including, e.g., the double-well potential distribution, which cannot be covered by the corresponding results in Wasserstein distances for HOLA in [40]. In addition, in the non-convex setting, we obtain state-of-the-art rates of convergence of aHOLA in Wasserstein-1 and Wasserstein-2 distances equal to $1 + q/2$ and $1/2 + q/4$, respectively, which improve the rates of convergence of the first and second-order algorithms in the existing literature. We achieve similar convergence results for aHOLLA under certain global Hölder and Lipschitz conditions and a dissipativity condition. To illustrate the applicability of our results, we use aHOLA and aHOLLA to sample from high-dimensional distributions including a multivariate standard Gaussian distribution, a multivariate Gaussian mixture distribution, and a double-well potential distribution. Numerical results confirm the theoretical improvement of our proposed algorithms in terms of the rate of convergence in Wasserstein distances. Moreover, we present a binary logistic regression example to show that aHOLLA can be used to sample from the associated posterior distribution.

The rest of the paper is organised as follows. Section 2 presents the setting and assumptions together with the main results for aHOLA where the target distributions have super-linearly growing potentials. Section 3 presents the setting and assumptions together with the main results for aHOLLA where the target distributions have at most linearly growing potentials. Section 4 discusses the related results in the literature in comparison with our work to highlight our contributions. Section 5 illustrates numerical results which support our main findings. Section 6 contains the proofs of main results in Section 2. Finally, Appendix A contains the proofs for the auxiliary results in Sections 2, 5, and 6, while Appendices B and C present tables summarising full expressions of all constants appearing in the statements for aHOLA and aHOLLA, respectively.

1.1. Notation. We conclude this section by introducing some notation. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We denote by $\mathbb{E}[Z]$ the expectation of a random variable Z . For $1 \leq p < \infty$, \mathcal{L}^p is used to denote the usual space of p -integrable real-valued random variables. Fix integers $d, m \geq 1$. For an \mathbb{R}^d -valued random variable Z , its law on $\mathcal{B}(\mathbb{R}^d)$, i.e. the Borel sigma-algebra of \mathbb{R}^d , is denoted by $\mathcal{L}(Z)$. For a positive real number a , we denote by $\lfloor a \rfloor$ its integer part, and $\lceil a \rceil := \min\{b \in \mathbb{Z} | b \geq a\}$. The Euclidean scalar product is denoted by $\langle \cdot, \cdot \rangle$, with $\|\cdot\|$ standing for the corresponding norm (where the dimension of the space may vary depending on the context). Denote by $\|M\|_F$ and M^\top the Frobenius norm and the transpose of any given matrix $M \in \mathbb{R}^{d \times m}$, respectively. Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ and $g : \mathbb{R}^d \rightarrow \mathbb{R}^d$ be twice continuously differentiable functions. Denote by ∇f , $\nabla^2 f$ and Δf the gradient of f , the Hessian of f , and the Laplacian of f , respectively. Denote by $\tilde{\Delta}g : \mathbb{R}^d \rightarrow \mathbb{R}^d$ the vector Laplacian of g , i.e., for all $\theta \in \mathbb{R}^d$, $\tilde{\Delta}g(\theta)$ is a vector in \mathbb{R}^d whose i -th entry is $\sum_{j=1}^d \partial_{\theta(j)}^2 g^{(i)}(\theta)$. For any integer $q \geq 1$, let $\mathcal{P}(\mathbb{R}^q)$ denote the set of probability measures on $\mathcal{B}(\mathbb{R}^q)$. For $\mu \in \mathcal{P}(\mathbb{R}^d)$ and for a μ -integrable function $f : \mathbb{R}^d \rightarrow \mathbb{R}$, the notation $\mu(f) := \int_{\mathbb{R}^d} f(\theta) \mu(d\theta)$ is used. For $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, let $\mathcal{C}(\mu, \nu)$ denote the set of probability measures ζ on $\mathcal{B}(\mathbb{R}^{2d})$ such that its respective marginals are μ, ν . For two Borel probability measures μ and ν defined on \mathbb{R}^d with finite p -th moments, the Wasserstein distance of order $p \geq 1$ is defined as

$$W_p(\mu, \nu) := \left(\inf_{\zeta \in \mathcal{C}(\mu, \nu)} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |\theta - \bar{\theta}|^p \zeta(d\theta d\bar{\theta}) \right)^{1/p}.$$

2. ASSUMPTIONS AND MAIN RESULTS FOR aHOLA

In this section, we introduce the newly designed aHOLA algorithm (5)-(7) to sample from π_β given in (1). The aHOLA algorithm is a modification of the HOLA algorithm proposed in [40], which is motivated by the order 1.5 numerical scheme of SDE (2) [22, Chapter 10]. We present the assumptions under which our main results are established, and provide convergence guarantees in Wasserstein distances for the aHOLA algorithm.

2.1. Setting. Let $U : \mathbb{R}^d \rightarrow \mathbb{R}$ be a three times continuously differentiable function satisfying $\int_{\mathbb{R}^d} e^{-\beta U(\theta)} d\theta < \infty$ for any $\beta > 0$ and denote by

$$h := \nabla U, \quad H := \nabla^2 U, \quad \Upsilon := \vec{\Delta}(\nabla U)$$

its gradient, Hessian, and vector Laplacian, respectively. Furthermore, define, for any $\beta > 0$,

$$\pi_\beta(A) := \frac{\int_A e^{-\beta U(\theta)} d\theta}{\int_{\mathbb{R}^d} e^{-\beta U(\theta)} d\theta}, \quad A \in \mathcal{B}(\mathbb{R}^d). \quad (3)$$

Let $\rho \in [2, \infty) \cap \mathbb{N}$ and $q \in (0, 1]$. Then, for any $i, j \in \mathbb{N}$, $\lambda > 0$, and for any function $f : \mathbb{R}^d \rightarrow \mathbb{R}^{i \times j}$, we denote by

$$f_\lambda(\theta) := \frac{f(\theta)}{(1 + \lambda^{3/2} |\theta|^{3(\rho+q-1)})^{1/3}}, \quad \theta \in \mathbb{R}^d. \quad (4)$$

The accelerated high order Langevin Monte Carlo algorithm (aHOLA) for SDE (2) is given by

$$\theta_0^{\text{aHOLA}} := \theta_0, \quad \theta_{n+1}^{\text{aHOLA}} = \theta_n^{\text{aHOLA}} + \lambda \phi^\lambda(\theta_n^{\text{aHOLA}}) + \sqrt{2\lambda\beta^{-1}} \psi^\lambda(\theta_n^{\text{aHOLA}}) \xi_{n+1}, \quad n \in \mathbb{N}_0, \quad (5)$$

where $\lambda > 0$ is the step size, $\beta > 0$, $(\xi_n)_{n \in \mathbb{N}_0}$ are i.i.d. standard d -dimensional Gaussian random variables, and where for all $\theta \in \mathbb{R}^d$,

$$\phi^\lambda(\theta) := -h_\lambda(\theta) + (\lambda/2) (H_\lambda(\theta)h_\lambda(\theta) - \beta^{-1}\Upsilon_\lambda(\theta)), \quad (6)$$

and

$$\psi^\lambda(\theta) := \sqrt{I_d - \lambda H_\lambda(\theta) + (\lambda^2/3)(H_\lambda(\theta))^2} \quad (7)$$

with I_d being the $d \times d$ identity matrix. We refer to Algorithm 1 for the implementation of aHOLA (5)-(7).

Remark 2.1. We note that, as mentioned in [11, 22], instead of taking the matrix square root as in (7), $\psi^\lambda(\theta_n^{\text{aHOLA}})\xi_{n+1}$ in (5) can be computed by considering the transformation

$$\left(I_d - (1/2)\lambda H_\lambda(\theta_n^{\text{aHOLA}}) \right) \widehat{\xi}_{n+1} + (\sqrt{3}/6)\lambda H_\lambda(\theta_n^{\text{aHOLA}}) \xi'_{n+1} \quad (8)$$

with $(\widehat{\xi}_n)_{n \in \mathbb{N}_0}$ and $(\xi'_n)_{n \in \mathbb{N}_0}$ being two independent standard Gaussian vectors independent of $(\xi_n)_{n \in \mathbb{N}_0}$ and θ_0 (see also Algorithm 2). This is due to the fact that (8) has the same distribution as (7).

Furthermore, we note that aHOLA (5)-(7) is developed based on the (tamed) order 1.5 scheme of SDE (2), see [22, Chapter 10] and [41], which is a class of explicit numerical schemes obtained using the Itô-Taylor expansion [37] and is given by

$$\begin{aligned} \theta_{n+1}^\lambda &= \theta_n^\lambda - \lambda h_\lambda(\theta_n^\lambda) + (\lambda^2/2) \left(H_\lambda(\theta_n^\lambda)h_\lambda(\theta_n^\lambda) - \beta^{-1}\Upsilon_\lambda(\theta_n^\lambda) \right) \\ &\quad + \sqrt{2\lambda\beta^{-1}} \widetilde{\xi}_{n+1} - \lambda \sqrt{2\lambda\beta^{-1}} H_\lambda(\theta_n^\lambda) \bar{\xi}_{n+1}, \quad n \in \mathbb{N}_0, \end{aligned} \quad (9)$$

where $\theta_0^\lambda := \theta_0$, $(\widetilde{\xi}_n)_{n \in \mathbb{N}_0}$ is a sequence of i.i.d. standard d -dimensional Gaussian random variables, and $(\bar{\xi}_n)_{n \in \mathbb{N}_0}$ is a sequence of i.i.d. d -dimensional Gaussian random variables with mean 0 and covariance $(1/3)I_d$. More precisely, for any $n \in \mathbb{N}_0$, $\bar{\xi}_{n+1} = \int_n^{n+1} \int_n^s dB_r^\lambda ds$, where $B_t^\lambda := B_{\lambda t}/\sqrt{\lambda}$, $t \geq 0$. One observes that the law of aHOLA (5)-(7) coincides with that of the algorithm (9) at grid points, i.e., for each $n \in \mathbb{N}_0$, $\mathcal{L}(\theta_n^{\text{aHOLA}}) = \mathcal{L}(\theta_n^\lambda)$. It is assumed throughout the paper that the \mathbb{R}^d -valued random variable θ_0 (the initial condition) is independent of $(\xi_n)_{n \in \mathbb{N}_0}$, $(\widetilde{\xi}_n)_{n \in \mathbb{N}_0}$, and $(\bar{\xi}_n)_{n \in \mathbb{N}_0}$.

Crucially, the (tamed) order 1.5 scheme converges in the strong \mathcal{L}^2 -sense with the rate of convergence equal to 1.5 [22, 41], which is higher than that for the corresponding Euler and Milstein schemes. Hence, we are interested in extending such a result to an infinite time horizon in Wasserstein distances under relaxed conditions. To this end, we propose the aHOLA algorithm (5)-(7), which can be viewed as a variant of the HOLA algorithm proposed in [40]. The key difference between the two algorithms is that aHOLA adopts a newly designed taming factor defined in (4). This allows us to derive uniform in time

moment estimates in Lemma 6.3, and to establish the convergence result in Wasserstein-1 distance with a desired rate of convergence as shown in Theorem 2.7 which cannot be achieved using HOLA [40].

Algorithm 1: aHOLA Algorithm (5)-(7)

- Input:** $\beta, \delta > 0$, $d \in \mathbb{N}$, measurable function $U : \mathbb{R}^d \rightarrow \mathbb{R}$, initialization $\theta_0 \in \mathbb{R}^d$.
Output: Estimator θ_n^{aHOLA} .
- 1 Set ρ, q, L, K_h, K_H to be the constants given by Assumption 2.
 - 2 Set a, b, r, \bar{r} to be the constants given by Assumption 3.
 - 3 Set C_0, C_1, C_2 to be the constants given in Theorem 2.7.
 - 4 Set $\lambda_{\max} := (14)$.
 - 5 Fix $\lambda \in (0, \min \{(\delta/(2C_2))^{2/(2+q)}, \lambda_{\max}\})$.
 - 6 Fix $n > \max \{\delta^{-2/(2+q)}(2C_2)^{2/(2+q)}/C_0, 1/(\lambda_{\max}C_0)\} \ln(2C_1(\mathbb{E}[|\theta_0|^{16(\rho+1)}] + 1)/\delta)$.
 - 7 Set $\phi^\lambda := (6)$, $\psi^\lambda := (7)$.
 - 8 Initialize $\theta_0^{\text{aHOLA}} \leftarrow \theta_0$.
 - 9 **for** $n = 0, \dots, n-1$ **do**
 - 10 Draw $\xi_{n+1} \sim \mathcal{N}(0, I_d)$.
 - 11 Set $\theta_{n+1}^{\text{aHOLA}} = \theta_n^{\text{aHOLA}} + \lambda\phi^\lambda(\theta_n^{\text{aHOLA}}) + \sqrt{2\lambda\beta^{-1}}\psi^\lambda(\theta_n^{\text{aHOLA}})\xi_{n+1}$.
 - 12 **return** θ_n^{aHOLA} .
-

Algorithm 2: aHOLA Algorithm (5)-(7) — A Variant Using (8)

- Input:** $\beta, \delta > 0$, $d \in \mathbb{N}$, measurable function $U : \mathbb{R}^d \rightarrow \mathbb{R}$, initialization $\theta_0 \in \mathbb{R}^d$.
Output: Estimator θ_n^{aHOLA} .
- 1 Set ρ, q, L, K_h, K_H to be the constants given by Assumption 2.
 - 2 Set a, b, r, \bar{r} to be the constants given by Assumption 3.
 - 3 Set C_0, C_1, C_2 to be the constants given in Theorem 2.7.
 - 4 Set $\lambda_{\max} := (14)$.
 - 5 Fix $\lambda \in (0, \min \{(\delta/(2C_2))^{2/(2+q)}, \lambda_{\max}\})$.
 - 6 Fix $n > \max \{\delta^{-2/(2+q)}(2C_2)^{2/(2+q)}/C_0, 1/(\lambda_{\max}C_0)\} \ln(2C_1(\mathbb{E}[|\theta_0|^{16(\rho+1)}] + 1)/\delta)$.
 - 7 Set $\phi^\lambda := (6)$.
 - 8 Initialize $\theta_0^{\text{aHOLA}} \leftarrow \theta_0$.
 - 9 **for** $n = 0, \dots, n-1$ **do**
 - 10 Draw independent standard Gaussian vectors $\hat{\xi}_{n+1}$ and ξ'_{n+1} .
 - 11 Set $\theta_{n+1}^{\text{aHOLA}} = \theta_n^{\text{aHOLA}} + \lambda\phi^\lambda(\theta_n^{\text{aHOLA}}) + \sqrt{2\lambda\beta^{-1}}(I_d - (1/2)\lambda H_\lambda(\theta_n^{\text{aHOLA}}))\hat{\xi}_{n+1} + (\sqrt{3}/6)\lambda H_\lambda(\theta_n^{\text{aHOLA}})\xi'_{n+1}$.
 - 12 **return** θ_n^{aHOLA} .
-

2.2. Assumptions. Let $U : \mathbb{R}^d \rightarrow \mathbb{R}$ be three times continuously differentiable, and let $\rho \in [2, \infty) \cap \mathbb{N}$ and $q \in (0, 1]$ be fixed. Moreover, we impose the following assumptions.

We first impose a condition on the initial value θ_0 .

Assumption 1. The initial condition θ_0 is independent of $(\xi_n)_{n \in \mathbb{N}_0}$ and has a finite $16(\rho+1)$ -th moment, i.e., $\mathbb{E}[|\theta_0|^{16(\rho+1)}] < \infty$.

Then, we impose a local Hölder condition on the third derivative of U .

Assumption 2. There exists $L > 0$ such that, for all $i = 1, \dots, d$, $\theta, \bar{\theta} \in \mathbb{R}^d$,

$$|\nabla^2 h^{(i)}(\theta) - \nabla^2 h^{(i)}(\bar{\theta})| \leq L(1 + |\theta| + |\bar{\theta}|)^{\rho-2} |\theta - \bar{\theta}|^q.$$

In addition, there exist $K_h, K_H > 0$ such that, for all $\theta \in \mathbb{R}^d$,

$$|h(\theta)| \leq K_h(1 + |\theta|^{\rho+q}), \quad |H(\theta)| \leq K_H(1 + |\theta|^{\rho+q-1}).$$

By Assumption 2, we obtain local Lipschitz (or Hölder) conditions and growth conditions on the first, second, and third derivatives of U as presented in the following remark. The proof is postponed to Appendix A.1.

Remark 2.2. By Assumption 2, for any $\theta, \bar{\theta} \in \mathbb{R}^d$, $i = 1, \dots, d$, we obtain the following estimates:

$$|\nabla^2 h^{(i)}(\theta)| \leq \mathsf{K}_0(1 + |\theta|)^{\rho+q-2}, \quad (10)$$

$$|\nabla h^{(i)}(\theta) - \nabla h^{(i)}(\bar{\theta})| \leq \mathsf{K}_0(1 + |\theta| + |\bar{\theta}|)^{\rho+q-2}|\theta - \bar{\theta}|, \quad (11)$$

$$|\nabla h^{(i)}(\theta)| \leq \mathsf{K}_1(1 + |\theta|)^{\rho+q-1},$$

$$|h^{(i)}(\theta) - h^{(i)}(\bar{\theta})| \leq \mathsf{K}_1(1 + |\theta| + |\bar{\theta}|)^{\rho+q-1}|\theta - \bar{\theta}|,$$

$$|h^{(i)}(\theta)| \leq \mathsf{K}_2(1 + |\theta|)^{\rho+q},$$

$$|\Upsilon(\theta) - \Upsilon(\bar{\theta})| \leq d^{3/2}L(1 + |\theta| + |\bar{\theta}|)^{\rho-2}|\theta - \bar{\theta}|^q, \quad (12)$$

$$|\Upsilon(\theta)| \leq \mathsf{K}_{3,d}(1 + |\theta|)^{\rho+q-2},$$

where $\mathsf{K}_0 := 2^{1-q} \max\{L, |\nabla^2 h^{(1)}(0)|, \dots, |\nabla^2 h^{(d)}(0)|\}$, $\mathsf{K}_1 := \max\{\mathsf{K}_0, |\nabla h^{(1)}(0)|, \dots, |\nabla h^{(d)}(0)|\}$, $\mathsf{K}_2 := \max\{\mathsf{K}_1, |h^{(1)}(0)|, \dots, |h^{(d)}(0)|\}$, and $\mathsf{K}_{3,d} := \max\{d^{3/2}L, \Upsilon(0)\}$.

Remark 2.3. One may notice that in Assumption 2, we assume separately growth conditions of h and H , which could have been deduced directly by using the polynomial Hölder condition of $\nabla^2 h^{(i)}$, $i = 1, \dots, d$, as shown in Remark 2.2. The reason is that the stepsize restriction λ_{\max} given in (14) is reciprocally related to the growth constants of h and H (i.e., K_h and K_H , respectively), thus, separately imposing growth conditions allows us to optimize λ_{\max} . For example, consider the double well potential $U(\theta) = (1/4)|\theta|^4 - (1/2)|\theta|^2$, $\theta \in \mathbb{R}^d$. In this case, it can be shown that, for any $i = 1, \dots, d$, $\nabla^2 h^{(i)}(\theta) = 2\theta e_i^\top + 2\theta^{(i)}\mathbf{1}_d + 2e_i\theta^\top$, where e_i denotes the standard basis vector in \mathbb{R}^d with its i -th element being 1. We have that, for any $\theta, \bar{\theta} \in \mathbb{R}^d$,

$$|\nabla^2 h^{(i)}(\theta) - \nabla^2 h^{(i)}(\bar{\theta})| \leq 6|\theta - \bar{\theta}|,$$

which implies, by following the same arguments as in the proof of Remark 2.2, that

$$|h(\theta)| \leq 24\sqrt{d}(1 + |\theta|^3), \quad |H(\theta)| \leq 12\sqrt{d}(1 + |\theta|^2).$$

However, since $h(\theta) = \theta(|\theta|^2 - 1)$ and $H(\theta) = \mathbf{1}_d(|\theta|^2 - 1) + 2\theta\theta^\top$, we have that

$$|h(\theta)| \leq 2(1 + |\theta|^3), \quad |H(\theta)| \leq 3(1 + |\theta|^2).$$

From the above calculations, we observe that the growth constants of h and H obtained by using the expressions is much smaller than those deduced using the polynomial Hölder condition.

Next, we impose a convexity at infinity condition on U .

Assumption 3. There exist constants $a, b > 0$, and $\bar{r} \in [0, r)$ with $r := \rho + q - 1$ such that, for all $\theta, \bar{\theta} \in \mathbb{R}^d$,

$$\langle \theta - \bar{\theta}, h(\theta) - h(\bar{\theta}) \rangle \geq a|\theta - \bar{\theta}|^2(|\theta|^r + |\bar{\theta}|^r) - b|\theta - \bar{\theta}|^2(|\theta|^{\bar{r}} + |\bar{\theta}|^{\bar{r}}).$$

Under Assumptions 2 and 3, we obtain a dissipativity condition on h . The explicit statement is provided below and its proof is postponed to Appendix A.1.

Remark 2.4. By Assumption 3, for any $\theta \in \mathbb{R}^d$, we obtain

$$\langle \theta, h(\theta) \rangle \geq a_D|\theta|^{r+2} - b_D,$$

where $a_D := a/2$ and $b_D := (a/2 + b)R_D^{\bar{r}+2} + |h(0)|^2/(2a)$ with $R_D := \max\{(4b/a)^{1/(r-\bar{r})}, 2^{1/r}\}$. Moreover, for any $\theta \in \mathbb{R}^d$, we have that

$$\langle \theta, h(\theta) \rangle \geq a_D|\theta|^2 - \bar{b}_D, \quad (13)$$

where $\bar{b}_D := a_D + b_D$.

Furthermore, we obtain a one-sided Lipschitz condition on h as presented below. The proof is postponed to Appendix A.1.

Remark 2.5. By Assumptions 2 and 3, for any $\theta, \bar{\theta} \in \mathbb{R}^d$, we obtain

$$\langle \theta - \bar{\theta}, h(\theta) - h(\bar{\theta}) \rangle \geq -L_{\text{OS}} |\theta - \bar{\theta}|^2,$$

where $L_{\text{OS}} := \sqrt{d}K_1(1 + 2R_{\text{OS}})^{\rho+q-1} > 0$ with $R_{\text{OS}} := (b/a)^{1/(r-\bar{r})}$.

Remark 2.6. We comment on our Assumptions 1-3:

- Assumption 1 imposes conditions on the initial value θ_0 to make sense of the convergence upper bounds in Theorem 2.7 and Corollary 2.8. In particular, the requirement that θ_0 has a finite $16(\rho + 1)$ -th moment is due to Lemma A.2 and (96), which is the minimal moment estimate required on θ_0 to establish the aforementioned convergence results using our current proof technique. We note that this condition can be easily satisfied by, e.g., constants or Gaussian random variables, which are typically used for algorithm initialization.
- Assumption 2 imposes a local Hölder condition on the third derivative of U , which allows super-linear growths of derivatives of U (as shown in Remark 2.2) so as to accommodate a wide range of target distributions including, e.g., the double-well potential distribution. In addition, we impose growth conditions of h and H to avoid unnecessarily small stepsize restriction as discussed in Remark 2.3.
- Assumption 3 imposes a convexity at infinity condition on U , which is the same as [33, Assumption 3] and one may refer to [33, Remark 2.4] for a detailed discussion. This condition implies a dissipativity condition of h as shown in Remark 2.4, which (together with Assumption 2) ensures the existence of a unique solution to SDE (2), and which is crucial to obtain moment estimates of the aHOLA algorithm. Moreover, it implies a one-sided Lipschitz condition as shown in Remark 2.5, which is key in establishing convergence results in Wasserstein distances.

Note that, while our assumptions accommodate a wide range of distributions, especially those with highly non-linear potentials, they exclude discrete distributions and some heavy-tailed distributions including, e.g., the Cauchy distributions.

2.3. Main results. Denote by

$$\lambda_{\max} := \min\{1, a_{\text{D}}^{-1}, (19a_{\text{D}}/240K_h \max\{K_H, K_h\})^2, (a_{\text{D}}/120K_h^2K_H^2)^{2/3}, a_{\text{D}}/(480K_h^2K_H)\}. \quad (14)$$

Under Assumptions 1, 2, and 3, we obtain the following non-asymptotic error bound in Wasserstein-1 distance between the law of aHOLA (5)-(7) and π_{β} .

Theorem 2.7. Let Assumptions 1, 2, and 3 hold. Then, for any $\beta > 0$, there exist positive constants C_0, C_1, C_2 such that, for any $n \in \mathbb{N}_0$, $0 < \lambda \leq \lambda_{\max}$,

$$W_1(\mathcal{L}(\theta_n^{\text{aHOLA}}), \pi_{\beta}) \leq C_1 e^{-C_0 \lambda n} (\mathbb{E}[|\theta_0|^{16(\rho+1)}] + 1) + C_2 \lambda^{1+q/2},$$

where C_0, C_1, C_2 are given explicitly in (37) (see also Appendix B). Furthermore, for any $\beta > 0$, $\delta > 0$, if we choose

$$\begin{aligned} \lambda &\leq \min\left\{(\delta/(2C_2))^{2/(2+q)}, \lambda_{\max}\right\}, \\ n &\geq \max\left\{\delta^{-2/(2+q)}(2C_2)^{2/(2+q)}/C_0, 1/(\lambda_{\max}C_0)\right\} \ln(2C_1(\mathbb{E}[|\theta_0|^{16(\rho+1)}] + 1)/\delta), \end{aligned}$$

then, we have $W_1(\mathcal{L}(\theta_n^{\text{aHOLA}}), \pi_{\beta}) \leq \delta$.

Moreover, we can also obtain a non-asymptotic result in Wasserstein-2 distance between the law of aHOLA (5)-(7) and π_{β} as presented below.

Corollary 2.8. Let Assumptions 1, 2, and 3 hold. Then, for any $\beta > 0$, there exist positive constants C_3, C_4, C_5 such that, for any $n \in \mathbb{N}_0$, $0 < \lambda \leq \lambda_{\max}$,

$$W_2(\mathcal{L}(\theta_n^{\text{aHOLA}}), \pi_{\beta}) \leq C_4 e^{-C_3 \lambda n} (\mathbb{E}[|\theta_0|^{16(\rho+1)}] + 1)^{1/2} + C_5 \lambda^{1/2+q/4}$$

where C_3, C_4, C_5 are given explicitly in (42) (see also Appendix B). Furthermore, for any $\beta > 0$, $\delta > 0$, if we choose

$$\begin{aligned} \lambda &\leq \min\left\{(\delta/(2C_5))^{4/(2+q)}, \lambda_{\max}\right\}, \\ n &\geq \max\left\{\delta^{-4/(2+q)}(2C_5)^{4/(2+q)}/C_3, 1/(\lambda_{\max}C_3)\right\} \ln(2C_4(\mathbb{E}[|\theta_0|^{16(\rho+1)}] + 1)^{1/2}/\delta), \end{aligned}$$

then, we have $W_2(\mathcal{L}(\theta_n^{\text{aHOLA}}), \pi_{\beta}) \leq \delta$.

The proofs of Theorem 2.7 and Corollary 2.8 are deferred in Section 6.

3. ASSUMPTIONS AND MAIN RESULTS FOR aHOLLA

In this section, we present the accelerated high order linear Langevin Monte Carlo algorithm (aHOLLA), which is the counterpart of the aHOLA algorithm in the linear setting where the derivatives of U are growing at most linearly. We provide assumptions under which convergence results of aHOLLA are established, followed by the formal statement of our main results.

3.1. Setting. The setting in this case is similar to that described in Section 2 except that there is no need to use tamed coefficients as in aHOLA (5)-(7). More precisely, with the assumptions and notation defined up to (3), the aHOLLA algorithm is given by

$$\Theta_0^{\text{aHOLLA}} := \theta_0, \quad \Theta_{n+1}^{\text{aHOLLA}} = \Theta_n^{\text{aHOLLA}} + \lambda \phi_{\text{Lin}}^\lambda(\Theta_n^{\text{aHOLLA}}) + \sqrt{2\lambda\beta^{-1}} \psi_{\text{Lin}}^\lambda(\Theta_n^{\text{aHOLLA}}) \xi_{n+1}, \quad n \in \mathbb{N}_0, \quad (15)$$

where for all $\theta \in \mathbb{R}^d$,

$$\phi_{\text{Lin}}^\lambda(\theta) := -h(\theta) + (\lambda/2) (H(\theta)h(\theta) - \beta^{-1}\Upsilon(\theta)), \quad (16)$$

and

$$\psi_{\text{Lin}}^\lambda(\theta) := \sqrt{1_d - \lambda H(\theta) + (\lambda^2/3)(H(\theta))^2}. \quad (17)$$

Remark 3.1. The corresponding order 1.5 scheme of aHOLLA (15)-(17) in the linear case is given by

$$\begin{aligned} \Theta_0^\lambda := \theta_0, \quad \Theta_{n+1}^\lambda = \Theta_n^\lambda - \lambda h(\Theta_n^\lambda) + (\lambda^2/2) (H(\Theta_n^\lambda)h(\Theta_n^\lambda) - \beta^{-1}\Upsilon(\Theta_n^\lambda)) \\ + \sqrt{2\lambda\beta^{-1}} \tilde{\xi}_{n+1} - \lambda \sqrt{2\lambda\beta^{-1}} H(\Theta_n^\lambda) \bar{\xi}_{n+1}, \quad n \in \mathbb{N}_0. \end{aligned} \quad (18)$$

Then, it holds that $\mathcal{L}(\Theta_n^{\text{aHOLLA}}) = \mathcal{L}(\Theta_n^\lambda)$, for each $n \in \mathbb{N}_0$.

We provide below the assumptions and main results for the aHOLLA algorithm (15)-(17).

3.2. Assumptions. Let $U : \mathbb{R}^d \rightarrow \mathbb{R}$ be three times continuously differentiable and let $q \in (0, 1]$ be fixed. The following assumptions are stated, which can be viewed as counterparts to those stated in Section 2.

We first impose assumptions on the initial condition θ_0 .

Assumption 4. The initial condition θ_0 is independent of $(\xi_n)_{n \in \mathbb{N}_0}$ and has a finite fourth moment, i.e., $\mathbb{E}[|\theta_0|^4] < \infty$.

Then, we impose conditions on the first, second, and third derivatives of U .

Assumption 5. There exists $\bar{L}_1 > 0$ such that, for all $i = 1, \dots, d$, $\theta, \bar{\theta} \in \mathbb{R}^d$,

$$|\nabla^2 h^{(i)}(\theta) - \nabla^2 h^{(i)}(\bar{\theta})| \leq \bar{L}_1 |\theta - \bar{\theta}|^q.$$

In addition, there exist $\bar{L}_2, \bar{L}_3 > 0$ such that, for all $\theta, \bar{\theta} \in \mathbb{R}^d$,

$$\begin{aligned} |H(\theta) - H(\bar{\theta})| &\leq \bar{L}_2 |\theta - \bar{\theta}|, \\ |h(\theta) - h(\bar{\theta})| &\leq \bar{L}_3 |\theta - \bar{\theta}|. \end{aligned}$$

Remark 3.2. By Assumption 5, for any $\theta, \bar{\theta} \in \mathbb{R}^d$, $i = 1, \dots, d$, we obtain the following estimates:

$$\begin{aligned} |\nabla^2 h^{(i)}(\theta)| &\leq \bar{K}_0 (1 + |\theta|^q), & |H(\theta)\bar{\theta}| &\leq \bar{L}_3 |\bar{\theta}|, & |h(\theta)| &\leq \bar{K}_1 (1 + |\theta|), \\ |\Upsilon(\theta) - \Upsilon(\bar{\theta})| &\leq d^{3/2} \bar{L}_1 |\theta - \bar{\theta}|^q, & |\Upsilon(\theta)| &\leq d \bar{L}_2, \end{aligned}$$

where $\bar{K}_0 := \max\{\bar{L}_1, |\nabla^2 h^{(1)}(0)|, \dots, |\nabla^2 h^{(d)}(0)|\}$, $\bar{K}_1 := \max\{\bar{L}_3, |h(0)|\}$.

Finally, we impose a dissipativity condition on h .

Assumption 6. There exist constants $\bar{a}, \bar{b} > 0$ such that, for all $\theta \in \mathbb{R}^d$,

$$\langle \theta, h(\theta) \rangle \geq \bar{a} |\theta|^2 - \bar{b}.$$

3.3. **Main results.** Denote by

$$\bar{\lambda}_{\max} := \min \left\{ 1, 1/\bar{a}, \bar{a}/(16\bar{L}_3\bar{K}_1), \bar{a}/(16\bar{K}_1^2), \bar{a}^{1/3}/(4\bar{L}_3\bar{K}_1^2)^{1/3}, \bar{a}^{1/2}/(16\bar{L}_3\bar{K}_1^2)^{1/3} \right\}. \quad (19)$$

Then, under Assumptions 4, 5, and 6, we deduce the following non-asymptotic convergence results in Wasserstein distances for aHOLLA (15)-(17).

Theorem 3.3. *Let Assumptions 4, 5, and 6 hold. Then, for any $\beta > 0$, there exist positive constants $C_{\text{Lin},0}, C_{\text{Lin},1}, C_{\text{Lin},2}$ such that, for any $n \in \mathbb{N}_0$, $0 < \lambda \leq \bar{\lambda}_{\max}$,*

$$W_1(\mathcal{L}(\Theta_n^{\text{aHOLLA}}), \pi_\beta) \leq C_{\text{Lin},1} e^{-C_{\text{Lin},0}\lambda n} (\mathbb{E}[|\theta_0|^4] + 1) + C_{\text{Lin},2} \lambda^{1+q/2},$$

where $C_{\text{Lin},0}, C_{\text{Lin},1}, C_{\text{Lin},2}$ are given explicitly in Appendix C. Furthermore, for any $\beta > 0$, $\delta > 0$, if we choose

$$\begin{aligned} \lambda &\leq \min \left\{ (\delta/2C_{\text{Lin},2})^{2/(2+q)}, \bar{\lambda}_{\max} \right\}, \\ n &\geq \max \left\{ (2C_{\text{Lin},2}/\delta)^{2/(2+q)} / C_{\text{Lin},0}, 1/(\bar{\lambda}_{\max} C_{\text{Lin},0}) \right\} \ln(2C_{\text{Lin},1} (\mathbb{E}[|\theta_0|^4] + 1) / \delta), \end{aligned}$$

then, we have $W_1(\mathcal{L}(\Theta_n^{\text{aHOLLA}}), \pi_\beta) \leq \delta$.

Corollary 3.4. *Let Assumptions 4, 5, and 6 hold. Then, for any $\beta > 0$, there exist positive constants $C_{\text{Lin},3}, C_{\text{Lin},4}, C_{\text{Lin},5}$ such that, for any $n \in \mathbb{N}_0$, $0 < \lambda \leq \bar{\lambda}_{\max}$,*

$$W_2(\mathcal{L}(\Theta_n^{\text{aHOLLA}}), \pi_\beta) \leq C_{\text{Lin},4} e^{-C_{\text{Lin},3}\lambda n} (\mathbb{E}[|\theta_0|^4] + 1)^{1/2} + C_{\text{Lin},5} \lambda^{1/2+q/4}$$

where $C_{\text{Lin},3}, C_{\text{Lin},4}, C_{\text{Lin},5}$ are given explicitly in Appendix C. Furthermore, for any $\beta > 0$, $\delta > 0$, if we choose

$$\begin{aligned} \lambda &\leq \min \left\{ (\delta/2C_{\text{Lin},5})^{4/(2+q)}, \bar{\lambda}_{\max} \right\}, \\ n &\geq \max \left\{ (2C_{\text{Lin},5}/\delta)^{4/(2+q)} / C_{\text{Lin},3}, 1/(\bar{\lambda}_{\max} C_{\text{Lin},3}) \right\} \ln(2C_{\text{Lin},4} (\mathbb{E}[|\theta_0|^4] + 1)^{1/2} / \delta), \end{aligned}$$

then, we have $W_2(\mathcal{L}(\Theta_n^{\text{aHOLLA}}), \pi_\beta) \leq \delta$.

The proofs of Theorem 3.3 and Corollary 3.4 follow the same ideas as in the proofs of Theorem 2.7 and Corollary 2.8 explained in Section 6, and hence are deferred in Appendix D.

4. RELATED WORK AND COMPARISON

In this section, we discuss related works in the literature and compare them with our results in Theorem 2.7, Corollary 2.8, Theorem 3.3, and Corollary 3.4. More precisely, we classify these results based on two types of assumptions: (i) the curvature condition and (ii) the smoothness condition, and then discuss the corresponding convergence guarantees.

Many works [2, 3, 8, 10, 11, 12, 13, 14, 31, 43] that investigate the sampling behavior of the Langevin-dynamics based algorithms impose a global Lipschitz condition and a strong convexity condition on h , i.e., there exist $\tilde{m}, \tilde{L} > 0$ such that for all $\theta, \bar{\theta} \in \mathbb{R}^d$,

$$\begin{cases} |h(\theta) - h(\bar{\theta})| \leq \tilde{L}|\theta - \bar{\theta}|, & \text{(Global Lipschitz continuity)} \\ \langle h(\theta) - h(\bar{\theta}), \theta - \bar{\theta} \rangle \geq \tilde{m}|\theta - \bar{\theta}|^2. & \text{(Strong convexity)} \end{cases}$$

Under these conditions, [10] shows that the LMC algorithm requires $\tilde{O}(d^3\delta^{-2})$ steps to reach a $\delta > 0$ precision level in total variation distance, while [2, 3, 11, 13, 14] prove that the (δ) -mixing time¹ scales as $\tilde{O}(d\delta^{-2})$ in Wasserstein-2 distance. Moreover, [43] obtains the convergence result in Kullback-Leibler (KL) divergence with mixing time $\tilde{O}(d\delta^{-1})$ under the additional assumption that the target distribution π_β of the form (1) satisfies a log-Sobolev inequality. For the underdamped Markov chain Monte Carlo algorithm (known also as the second-order Langevin algorithm), [8, 12] establish convergence results in Wasserstein-2 distance with mixing time $\tilde{O}(d^{1/2}\delta^{-1})$. Furthermore, [31] proposes a third-order Langevin algorithm and provides a convergence result in Wasserstein-2 distance for the algorithm to sample from the distribution whose potential U is of a ridge-separable form [31, Eq. (1)]. It is shown in [31, Theorem 1] that the mixing time is $\tilde{O}(d^{1/4}\delta^{-1/2})$.

¹The minimum number of steps required for an algorithm to reach within $\delta > 0$ of the target distribution in certain distance.

However, the global Lipschitz condition and the strong convexity condition are restrictive in the sense that they cannot accommodate key applications including, e.g., the problem of sampling from the double-well potential distribution and from the posterior distribution associated with the Bridge linear regression [18]. To address this issue, recent works focus on the relaxation of the aforementioned assumptions:

- (i) To relax the strong convexity condition, one direction is to replace it with a convexity at infinity condition. [7] shows that, under the condition that U is strongly convex outside a ball of radius c_R , the LMC algorithm converges in Wasserstein-1 distance with mixing time $\tilde{O}(de^{O(c_R^2)}\delta^{-2})$.

Another direction considers to replace the strong convexity condition with a dissipativity condition [6, 32, 33, 34, 38, 42, 45, 46]. In particular, [38, 45] establish convergence results in Wasserstein-2 distance for the SGLD algorithm, requiring respectively $\tilde{O}(d(c_{\text{spe}}\delta^4)^{-1})$ and $\tilde{O}(d(c_{\text{spe}}\delta)^{-1})$ steps together with a large enough minibatch size to reach a $\delta > 0$ precision level, where $c_{\text{spe}} > 0$ denotes the spectral gap which scales as $O(e^{-\tilde{O}(d)})$. Then, [6], and subsequently [34, 42, 46], improve substantially the results in [38, 45] by removing the requirement on the minibatch sizes, and show that the SGLD algorithm reaches δ precision level within $\tilde{O}(e^{O(d)}\delta^{-2})$ and $\tilde{O}(e^{O(d)}\delta^{-4})$ steps in Wasserstein-1 and Wasserstein-2 distances, respectively. These results are further improved in [33] to $\tilde{O}(e^{O(d)}\delta^{-1})$ and $\tilde{O}(e^{O(d)}\delta^{-2})$, under a Hessian Lipschitz condition on U . Moreover, for the LMC algorithm with drift $h/2$, by assuming additionally a log-Sobolev inequality on π_β , [32] obtains convergence results in Wasserstein-1 and Wasserstein-2 distances with mixing times $\tilde{O}(d^{3/2}(c_{\text{LSI}}^{3/2}\delta)^{-1})$ and $\tilde{O}(d(c_{\text{LSI}}^{5/2}\delta)^{-1})$, respectively, where $c_{\text{LSI}} > 0$ denotes the log-Sobolev constant.

In addition, a recent result [9] establishes a convergence result for the LMC algorithm in Rényi divergence under the conditions that π_β has a Lipschitz log-gradient and satisfies a log-Sobolev inequality with constant $c_{\text{LSI}} > 0$. It shows that the LMC algorithm requires $\tilde{O}(dc_{\text{LSI}}^2\delta^{-1})$ steps to reach a δ precision level.

- (ii) To relax the global Lipschitz condition, one line of research focuses on sampling from distributions with (a mixture of) weakly smooth potentials [17, 35]. Under the degenerate convexity at infinity condition, a c_1^{WS} -dissipativity condition with $c_1^{\text{WS}} \in [1, 2]$, and a c_2^{WS} -Hölder condition with $c_2^{\text{WS}} \in [c_3^{\text{WS}}, 1]$, $c_3^{\text{WS}} \in (0, c_1^{\text{WS}}/2]$, [17] establishes a convergence result for the LMC algorithm in Wasserstein- c_1^{WS} distance with a mixing time $\tilde{O}(d^{(3c_1^{\text{WS}} + (1+c_4^{\text{WS}})c_2^{\text{WS}})/(c_1^{\text{WS}}c_2^{\text{WS}})}\delta^{-2c_1^{\text{WS}}/c_2^{\text{WS}}})$ with $c_4^{\text{WS}} \geq 0$. Then, [35] considers to relax the aforementioned conditions and replaces them with the following weaker ones: a Poincaré inequality with constant $c_{\text{PI}} > 0$, a \bar{c}_2^{WS} -mixture weakly smooth condition with $0 < \bar{c}_2^{\text{WS}} = \bar{c}_{2,1}^{\text{WS}} < \dots < \bar{c}_{2,N}^{\text{WS}} \leq 1$, and a \bar{c}_1^{WS} -dissipativity condition with $\bar{c}_1^{\text{WS}} \geq \bar{c}_{2,N}^{\text{WS}}$. Under these conditions, [35] obtains Wasserstein- \bar{c}_1^{WS} convergence result with a mixing time $\tilde{O}(d^{(2+2(1+1/\bar{c}_2^{\text{WS}})/\bar{c}_1^{\text{WS}})\vee(2+4/\bar{c}_2^{\text{WS}}+3\bar{c}_{2,N}^{\text{WS}}/p)}c_{\text{PI}}^{-(1+1/\bar{c}_2^{\text{WS}})}\delta^{-2\bar{c}_1^{\text{WS}}(1+2/\bar{c}_2^{\text{WS}})})$ with $p > 1$.

Another line of research focuses on sampling from distributions with super-linearly growing log-gradients [25, 26, 27, 29, 33, 40]. [27, 29] propose the TUSLA algorithm which is a variant of the SGLD algorithm, and establish convergence results in Wasserstein-1 and Wasserstein-2 distances with mixing times $\tilde{O}(e^{O(d)}\delta^{-2})$ and $\tilde{O}(e^{O(d)}\delta^{-4})$, respectively. These results are achieved under a polynomial Lipschitz condition imposed on the gradient of U . [25, 26] propose another variant of the SGLD algorithm, called TH ϵ O POULA, which possesses superior empirical performance, and obtain similar convergence results to those in [27, 29]. In addition, [33] proposes the mTULA algorithm, which can be viewed as a variant of the LMC algorithm. Theoretical guarantees are established in Wasserstein-1 and Wasserstein-2 distances with improved mixing times $\tilde{O}(e^{O(d)}\delta^{-1})$ and $\tilde{O}(e^{O(d)}\delta^{-2})$, respectively, but under a polynomial Lipschitz condition on the Hessian of U . [40] proposes the HOLA algorithm that utilizes the third derivative of U . Under the conditions that the third derivative of U is polynomially q -Hölder continuous, $q \in (0, 1]$, and that U is strongly convex, [40] shows that the HOLA algorithm converges in Wasserstein-2 distance within $\tilde{O}(e^{O(d)}\delta^{-2/(2+q)})$ steps.

In Table 1, we compare our assumptions and results to those in the aforementioned works in (i) and (ii). In particular, our Theorem 2.7, Corollary 2.8, Theorem 3.3, and Corollary 3.4 provide convergence results for aHOLA (5)-(7) and aHOLLA (15)-(17) in Wasserstein-1 and Wasserstein-2 distances with

mixing times $\tilde{O}(e^{O(d)}\delta^{-2/(2+q)})$ and $\tilde{O}(e^{O(d)}\delta^{-4/(2+q)})$. The results are obtained under a polynomial q -Hölder condition on the third derivative of U (Assumption 2) and a convexity at infinity condition (Assumption 3). The latter condition is weaker than the strong convexity condition imposed in [40], hence allowing the application of our results to a broader range of distributions that cannot be covered by [40, Theorems 1 and 3] including, e.g., the double-well distribution. In addition, our results improve the rates of convergence obtained in [33], from $\tilde{O}(e^{O(d)}\delta^{-1})$ and $\tilde{O}(e^{O(d)}\delta^{-2})$ to $\tilde{O}(e^{O(d)}\delta^{-2/(2+q)})$ and $\tilde{O}(e^{O(d)}\delta^{-4/(2+q)})$. These improvements are achieved due to the assumptions imposed on the high order derivatives of U .

WORK	CONVERGENCE RATE	SMOOTHNESS	CURVATURE	OTHERS	DIST.
[7]	$\tilde{O}(de^{O(c_R^2)}\delta^{-2})$	Lipschitz h	Convexity at ∞	—	W-1
[38]	$\tilde{O}(d(c_{\text{spe}}\delta^4)^{-1})$	Lipschitz h	—	Dissipativity	W-2
[45]	$\tilde{O}(d(c_{\text{spe}}\delta)^{-1})$	Lipschitz h	—	Dissipativity	W-2
[6, 34, 42, 46]	$\tilde{O}(e^{O(d)}\delta^{-2})$ $\tilde{O}(e^{O(d)}\delta^{-4})$	Lipschitz h	—	Dissipativity	W-1 W-2
[33]	$\tilde{O}(e^{O(d)}\delta^{-1})$ $\tilde{O}(e^{O(d)}\delta^{-2})$	Lipschitz h, H	—	Dissipativity	W-1 W-2
[40]	$\tilde{O}(e^{O(d)}\delta^{-2/3})$	Lipschitz $h, H, \nabla^2 h^{(i)}$	Strong convexity	—	W-2
[32]	$\tilde{O}(d^{3/2}(c_{\text{LSI}}^{3/2}\delta)^{-1})$ $\tilde{O}(d(c_{\text{LSI}}^{5/2}\delta)^{-1})$	Lipschitz h, H	—	Dissipativity, LSI, LMC with drift $h/2$	W-1 W-2
[9]	$\tilde{O}(dc_{\text{LSI}}^2\delta^{-1})$	Lipschitz h	—	LSI	Rényi
Thm 3.3, Cor 3.4	$\tilde{O}(e^{O(d)}\delta^{-2/(2+q)})$ $\tilde{O}(e^{O(d)}\delta^{-4/(2+q)})$	Lipschitz $h, H,$ q -Hölder $\nabla^2 h^{(i)}$	—	Dissipativity	W-1 W-2
[17]	$\tilde{O}\left(\frac{d(3c_1^{\text{WS}} + (1+c_4^{\text{WS}})c_2^{\text{WS}})/(c_1^{\text{WS}}c_2^{\text{WS}})}{\delta^2c_1^{\text{WS}}/c_2^{\text{WS}}}\right)$	c_2^{WS} -Hölder h	Degenerate convexity at ∞	c_1^{WS} -dissipativity	W- c_1^{WS}
[35]	$\tilde{O}\left(\frac{d\left(2 + \frac{2(1+1/\bar{c}_1^{\text{WS}})}{\bar{c}_1^{\text{WS}}}\right) \vee \left(2 + \frac{4}{\bar{c}_2^{\text{WS}}} + \frac{3\bar{c}_2^{\text{WS}}}{p}\right)}{c_{\text{pl}}^{(1+1/\bar{c}_1^{\text{WS}})}\delta^2\bar{c}_1^{\text{WS}}(1+2/\bar{c}_2^{\text{WS}})}\right)$	\bar{c}_2^{WS} -mixture weakly smooth h	—	\bar{c}_1^{WS} -dissipativity, Poincaré inequality	W- \bar{c}_1^{WS}
[25, 26, 27, 29]	$\tilde{O}(e^{O(d)}\delta^{-2})$ $\tilde{O}(e^{O(d)}\delta^{-4})$	Polynomial Lipschitz h	Convexity at ∞	—	W-1 W-2
[33]	$\tilde{O}(e^{O(d)}\delta^{-1})$ $\tilde{O}(e^{O(d)}\delta^{-2})$	Polynomial Lipschitz h, H	Convexity at ∞	—	W-1 W-2
[40]	$\tilde{O}(e^{O(d)}\delta^{-2/(2+q)})$	Polynomial q -Hölder $\nabla^2 h^{(i)}$	Strong convexity	—	W-2
Thm 2.7, Cor 2.8	$\tilde{O}(e^{O(d)}\delta^{-2/(2+q)})$ $\tilde{O}(e^{O(d)}\delta^{-4/(2+q)})$	Polynomial q -Hölder $\nabla^2 h^{(i)}$	Convexity at ∞	—	W-1 W-2

TABLE 1. Comparison between our main results, i.e., Theorem 2.7, Corollary 2.8, Theorem 3.3, Corollary 3.4 and results in other works. $h, H, \nabla^2 h^{(i)}$ denotes the first, second, third derivative of U , respectively.

5. NUMERICAL EXPERIMENTS

In this section, we illustrate the applicability of our results in Theorem 2.7, Corollary 2.8, Theorem 3.3, and Corollary 3.4. We show that aHOLA (5)-(7) and aHOLLA (15)-(17) can be used to solve the problems of high-dimensional sampling and binary logistic regression. We then compare the performance of our algorithms with existing algorithms including LMC [11], LMCO' [11], mTULA [33], and HOLA [40].²

5.1. High-dimensional sampling. We use aHOLA (5)-(7) and aHOLLA (15)-(17) to draw samples from various distributions. More precisely, we consider the following high-dimensional target distributions:

- (i) a multivariate standard Gaussian distribution with its potential given by

$$U(\theta) := \frac{1}{2}|\theta|^2, \quad \theta \in \mathbb{R}^d, \quad (20)$$

²The python code is available at <https://github.com/tracyyingzhang/aHOLA>.

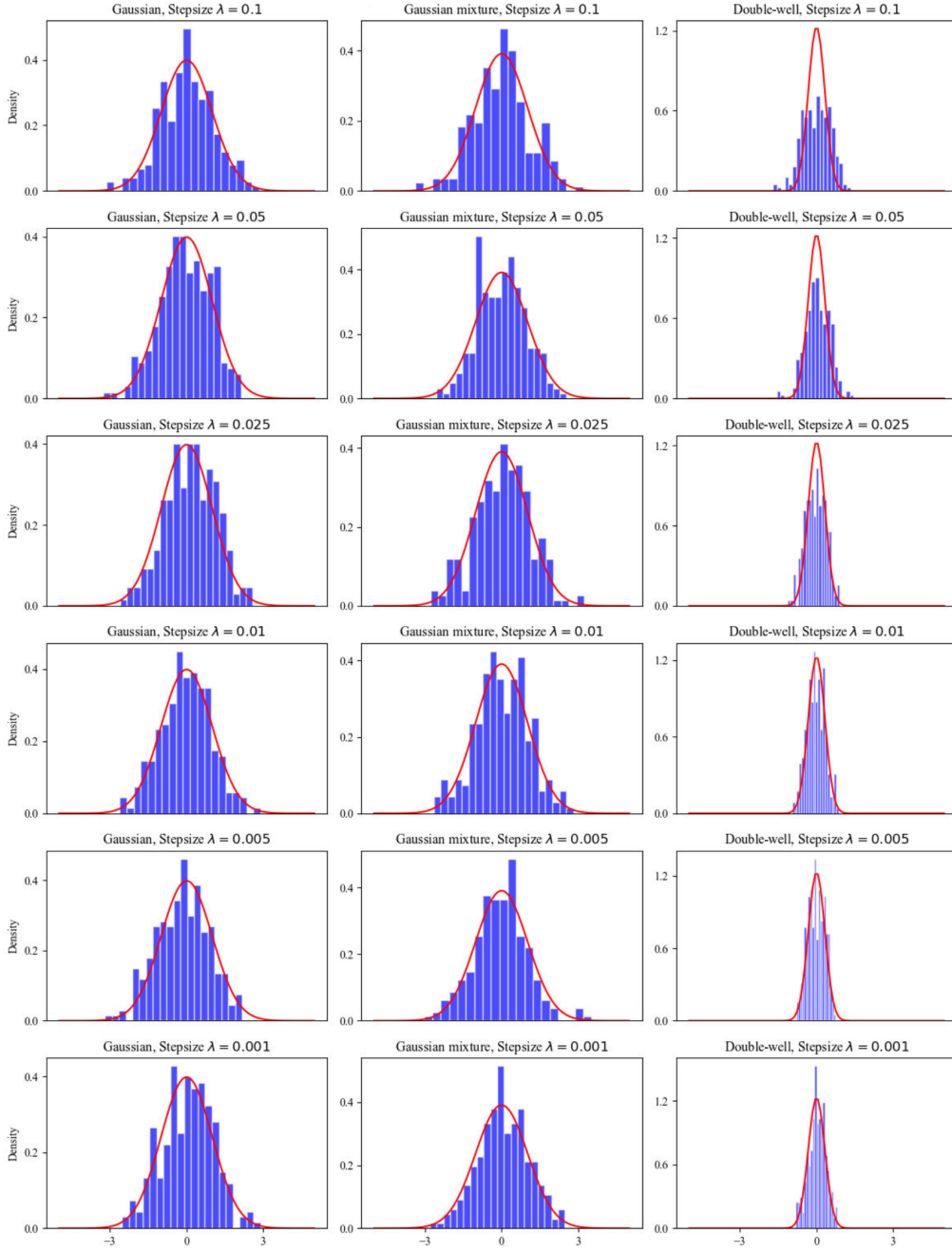


FIGURE 1. Normalised histograms of the first components of the samples drawn using aHOLA and aHOLLA.

(ii) a multivariate Gaussian mixture distribution with its potential given by

$$U(\theta) := \frac{1}{2}|\theta - \hat{a}|^2 - \log(1 + \exp(-2\langle \theta, \hat{a} \rangle)), \quad \theta \in \mathbb{R}^d, \quad (21)$$

where $\hat{a} \in \mathbb{R}^d$ is a given vector with $|\hat{a}| > 1$, and

(iii) a double-well potential distribution with its potential given by

$$U(\theta) := \frac{1}{4}|\theta|^4 - \frac{1}{2}|\theta|^2, \quad \theta \in \mathbb{R}^d. \quad (22)$$

We note that the first two distributions in (i) and (ii) have potentials (20) and (21) whose gradients are growing at most linearly, while the last distribution in (iii) has potential (22) whose gradient is growing

super-linearly. We show in the following proposition that the examples (i) and (ii) satisfy Assumptions 5 and 6, and (iii) satisfies Assumptions 2 and 3.

Proposition 5.1. *The functions $U : \mathbb{R}^d \rightarrow \mathbb{R}$ given in (20) and (21) satisfy Assumptions 5 and 6 while that given in (22) satisfies Assumptions 2 and 3.*

Proof. See Appendix A.2. □

For the numerical experiments, we set $d = 100$. The initial value θ_0 is set to be the null vector in \mathbb{R}^d so as to satisfy Assumptions 1 and 4. This, together with Proposition 5.1, ensures that the proposed algorithms aHOLA (5)-(7) and aHOLLA (15)-(17) can sample approximately from the target distributions due to our main results. In addition, we set $\beta = 1$ and consider different stepsizes $\lambda = \{0.001, 0.005, 0.01, 0.025, 0.05, 0.1\}$. The number of iterations n is chosen such that $\lambda n = 400$ is fixed. For each distribution and stepsize mentioned above, we run 250 independent aHOLA or aHOLLA Markov chains and collect outputs from the last iterations of these Markov chains.

To illustrate the performance of the proposed algorithms, in Figure 1, we plot the normalised histograms of the first components of the 250 samples generated using the method described above, and compare the histograms obtained numerically with their corresponding marginal density functions which are red lines superimposed on the plots. We note that the marginal density functions for the first component of a multivariate standard Gaussian distribution, a multivariate Gaussian mixture distribution, and a double-well potential distribution are given by

$$\begin{aligned} \theta^{(1)} &\ni \frac{1}{\sqrt{2\pi}} e^{-(\theta^{(1)})^2/2}, \\ \theta^{(1)} &\ni \frac{1}{2\sqrt{2\pi}} \left(e^{-(\theta^{(1)} - \hat{a}^{(1)})^2/2} + e^{-(\theta^{(1)} + \hat{a}^{(1)})^2/2} \right), \\ \theta^{(1)} &\ni \frac{\Gamma(d/2)}{\sqrt{\pi}\Gamma((d-1)/2)} \frac{\int_0^\infty x^{(d-3)/2} \exp\left(-\frac{(x + (\theta^{(1)})^2)^2}{4} + \frac{(x + (\theta^{(1)})^2)}{2}\right) dx}{\int_0^\infty x^{d/2-1} \exp(-x^2/4 + x/2) dx}, \end{aligned} \quad (23)$$

respectively, where we denote by $\theta^{(1)}, \hat{a}^{(1)} \in \mathbb{R}$ the first component of $\theta, \hat{a} \in \mathbb{R}^d$, respectively, and where Γ denotes the gamma function. In addition, for the multivariate Gaussian mixture distribution, we set $|\hat{a}| = 2$ in the experiments with all its components being equal.

In Figure 1, we observe that the histograms obtained using samples generated by aHOLA (5)-(7) and aHOLLA (15)-(17) are close to their corresponding theoretical marginal density functions. This illustrates that the proposed algorithms can generate samples approximately from given target distributions with appropriately chosen stepsizes. The numerical results support our main findings, i.e., Theorem 2.7, Corollary 2.8, Theorem 3.3, and Corollary 3.4.

Moreover, we compare the sampling performance of aHOLLA and aHOLA with that of existing algorithms including LMC [11, Eq. (2)], LMCO' [11, Eq. (17)], mTULA [33, Eqs. (6)-(7)], and HOLA [40, Eqs. (2)-(3)]. More precisely, we use aHOLLA, LMC, and LMCO' to sample from the standard normal distribution and the Gaussian mixture distribution given in (i) and (ii), and use aHOLA, mTULA, and HOLA to sample from the double-well potential distribution given in (iii). Denote by $\pi_\beta^{\text{sn}}, \pi_\beta^{\text{gm}}$, and π_β^{dw} the marginal distributions of (i), (ii) and (iii), respectively, with their densities given in (23). We compute the Wasserstein-1 distances between $\pi_\beta^{\text{sn}}, \pi_\beta^{\text{gm}}, \pi_\beta^{\text{dw}}$ and their corresponding distributions generated by the aforementioned numerical algorithms using 200 samples (collected from 200 independent Markov chains in each setting). For the numerical experiments, we set $\beta = 1, d = 100, \theta_0 = (2, 2, \dots, 2) \in \mathbb{R}^{100}$, $\lambda = \{0.025, 0.05, 0.075, 0.1, 0.25\}$, $n = 1000/\lambda$.

In Table 2, we present the first time (measured in number of iterations) the aforementioned Wasserstein distances fall below a precision level of 0.1, while we present in Table 3 the best accuracy achieved by each algorithm in Wasserstein-1 distance. We observe from Table 2 that, in most cases, the algorithms incorporating third order derivatives, i.e., aHOLLA, HOLA, and aHOLA, converge much faster than those utilizing only first and second order derivatives, i.e., LMC, LMCO', and mTULA. Moreover, as illustrated in Table 3, the approximated distributions generated by aHOLLA, HOLA, and aHOLA are much closer to the target distributions in Wasserstein-1 distance compared to LMC, LMCO', and mTULA, and the distances become smaller when λ decreases. These support our theoretical findings in Theorem 2.7 and Theorem 3.3 that aHOLLA and aHOLA converge with a higher rate of convergence in Wasserstein-1 distance. Furthermore, in Table 4, we report the running time required to complete the task in each setting,

while, in Table 5, we present the per-iteration cost for each algorithm. We highlight that the computation time required by aHOLLA and aHOLA to complete each task is comparable to that of LMC, as it is not computationally more expensive to compute the Hessian-vector product and the vector Laplacian in (6), (8), and (16) than computing the gradient. The per-iteration cost for both aHOLLA and aHOLA scales as $\mathcal{O}(d)$, which is the same as that of LMC, LMCO', and mTULA. We note that the per-iteration cost for HOLA is $\mathcal{O}(d^2)$, which is due to the design of its taming factor that involves the computation of the full Hessian matrix.

	Standard normal distribution				
	$\lambda = 0.25$	$\lambda = 0.1$	$\lambda = 0.075$	$\lambda = 0.05$	$\lambda = 0.025$
LMC	8	25	32	56	78
LMCO'	10	26	27	47	66
aHOLLA	11	26	40	45	151
	Gaussian-mixture distribution				
	$\lambda = 0.25$	$\lambda = 0.1$	$\lambda = 0.075$	$\lambda = 0.05$	$\lambda = 0.025$
LMC	14	42	103	72	207
LMCO'	28	40	67	104	122
aHOLLA	14	43	77	146	104
	Double-well potential distribution				
	$\lambda = 0.25$	$\lambda = 0.1$	$\lambda = 0.075$	$\lambda = 0.05$	$\lambda = 0.025$
mTULA	NA	NA	NA	NA	42
HOLA	NA	NA	NA	4365	28
aHOLA	NA	NA	NA	43	25

TABLE 2. (High-dimensional sampling) Number of iterations required for an algorithm to reach within 0.1 precision of the target distribution in Wasserstein-1 distance. “NA” implies the algorithm never reaches the 0.1 precision level within $1000/\lambda$ iterations.

	Standard normal distribution				
	$\lambda = 0.25$	$\lambda = 0.1$	$\lambda = 0.075$	$\lambda = 0.05$	$\lambda = 0.025$
LMC	0.0529	0.0449	0.0496	0.0444	0.0434
LMCO'	0.0533	0.0534	0.0419	0.0387	0.0438
aHOLLA	0.0484	0.0403	0.0491	0.0459	0.0382
	Gaussian-mixture distribution				
	$\lambda = 0.25$	$\lambda = 0.1$	$\lambda = 0.075$	$\lambda = 0.05$	$\lambda = 0.025$
LMC	0.051	0.0454	0.0445	0.0406	0.0432
LMCO'	0.051	0.0471	0.0465	0.0429	0.0436
aHOLLA	0.0457	0.0419	0.0457	0.0421	0.0397
	Double-well potential distribution				
	$\lambda = 0.25$	$\lambda = 0.1$	$\lambda = 0.075$	$\lambda = 0.05$	$\lambda = 0.025$
mTULA	0.3169	0.1747	0.1353	0.1039	0.05
HOLA	1.2993	0.381	0.2513	0.0964	0.0224
aHOLA	0.2049	0.1337	0.1115	0.0795	0.0383

TABLE 3. (High-dimensional sampling) Best accuracy achieved among $1000/\lambda$ iterations, measured in Wasserstein-1 distance between the approximated distribution generated by a numerical algorithm and the target distribution.

5.2. Binary logistic regression. We consider the problem of logistic regression [10] where we obtain a set of data points $\{x_i\}_{i=1}^N = \{(z_i, y_i)\}_{i=1}^N$ with $z_i \in \mathbb{R}^d$ denoting the feature vector and $y_i \in \{0, 1\}$ denoting the binary label. Our goal is to estimate the conditional distribution of Y_1 given Z_1 , which is equivalent to the estimation of the regression function $f(z) = \mathbb{P}(Y_1 = 1 | Z_1 = z)$, $z \in \mathbb{R}^d$. In the logistic regression model, $f(z)$ is approximated by $f_\theta(z) = e^{\langle \theta, z \rangle} / (1 + e^{\langle \theta, z \rangle})$. To estimate the parameter θ ,

	Standard normal distribution				
	$\lambda = 0.25$	$\lambda = 0.1$	$\lambda = 0.075$	$\lambda = 0.05$	$\lambda = 0.025$
LMC	6.38	14.33	19.01	28.52	56.88
LMCO'	5.91	14.61	19.39	29.05	58.04
aHOLLA	6.55	16.23	21.63	32.32	64.53
	Gaussian-mixture distribution				
	$\lambda = 0.25$	$\lambda = 0.1$	$\lambda = 0.075$	$\lambda = 0.05$	$\lambda = 0.025$
LMC	25.38	63.08	84.34	123.43	246.84
LMCO'	25.55	63.21	84.23	126.50	252.15
aHOLLA	29.33	71.36	96.87	142.64	285.16
	Double-well potential distribution				
	$\lambda = 0.25$	$\lambda = 0.1$	$\lambda = 0.075$	$\lambda = 0.05$	$\lambda = 0.025$
mTULA	18.27	46.58	60.77	92.99	185.56
HOLA	24.00	60.55	79.43	120.92	241.55
aHOLA	21.70	54.43	72.54	108.84	217.81

TABLE 4. (High-dimensional sampling) Running time (measured in seconds) for an algorithm to complete 200 runs with each run consisting of $1000/\lambda$ iterations.

LMC	LMCO'	aHOLLA	mTULA	HOLA	aHOLA
$\mathcal{O}(d)$	$\mathcal{O}(d)$	$\mathcal{O}(d)$	$\mathcal{O}(d)$	$\mathcal{O}(d^2)$	$\mathcal{O}(d)$

TABLE 5. (High-dimensional sampling) Per-iteration cost for each algorithm.

$d = 2$	$N = 100$	$N = 200$	$N = 300$	$N = 400$	$N = 500$
LMC	0.1261	0.0668	0.0361	0.0310	0.0194
LMCO'	0.1579	0.0668	0.0439	0.0316	0.0232
aHOLLA	0.1144	0.0942	0.0494	0.0327	0.0224
$d = 5$	$N = 100$	$N = 200$	$N = 300$	$N = 400$	$N = 500$
LMC	0.3834	0.2140	0.1355	0.1416	0.0938
LMCO'	0.3896	0.2214	0.1563	0.1148	0.0881
aHOLLA	0.4465	0.2121	0.1294	0.1142	0.0810
$d = 10$	$N = 100$	$N = 200$	$N = 300$	$N = 400$	$N = 500$
LMC	1.0737	0.5184	0.3786	0.2641	0.2413
LMCO'	1.0605	0.5249	0.4279	0.2533	0.2227
aHOLLA	1.1439	0.6004	0.3787	0.2976	0.2305

TABLE 6. (Binary logistic regression) MSE between the true value and the Bayesian posterior mean computed using 200 samples generated by an algorithm.

we adopt a Bayesian approach by sampling from a posterior distribution $\pi(\theta) = \pi_0(\theta) \prod_{i=1}^N p(y_i | z_i, \theta)$, where π_0 is a standard normal prior and $p(y_i | z_i, \theta)$ is the likelihood function given by $p(y_i | z_i, \theta) = (1 + e^{-\langle \theta, z_i \rangle})^{-y_i} (1 + e^{\langle \theta, z_i \rangle})^{y_i - 1}$. Thus, the posterior takes the form

$$\pi(d\theta) \propto \exp \left\{ -|\theta|^2/2 - \sum_{i=1}^N (1 - y_i) \langle \theta, z_i \rangle - \sum_{i=1}^N \log(1 + e^{-\langle \theta, z_i \rangle}) \right\} d\theta, \quad (24)$$

which satisfies the form of (1) with $\beta = 1$ and

$$U(\theta) = |\theta|^2/2 + \sum_{i=1}^N (1 - y_i) \langle \theta, z_i \rangle + \sum_{i=1}^N \log(1 + e^{-\langle \theta, z_i \rangle}), \quad (25)$$

for all $\theta \in \mathbb{R}^d$. We aim to use aHOLLA (15)-(17) to solve the logistic regression example described above, and we show, in the following result, that our theoretical findings can be applied.

$d = 2$	$N = 100$	$N = 200$	$N = 300$	$N = 400$	$N = 500$
LMC	19.55	22.85	26.72	30.30	33.80
LMCO'	52.28	62.82	72.69	81.51	91.09
aHOLLA	66.50	79.88	93.01	106.05	118.14
$d = 5$	$N = 100$	$N = 200$	$N = 300$	$N = 400$	$N = 500$
LMC	19.93	24.88	29.32	33.70	37.97
LMCO'	55.38	69.47	81.39	93.34	104.81
aHOLLA	69.98	88.84	104.65	120.30	136.19
$d = 10$	$N = 100$	$N = 200$	$N = 300$	$N = 400$	$N = 500$
LMC	21.89	27.76	33.80	39.72	45.28
LMCO'	60.19	76.62	92.72	108.89	124.09
aHOLLA	76.02	97.85	119.59	139.62	159.87

TABLE 7. (Binary logistic regression) Running time (measured in seconds) for an algorithm to complete 50 runs with each run consisting of 100000 iterations.

LMC	LMCO'	aHOLLA
$\mathcal{O}(Nd)$	$\mathcal{O}(Nd)$	$\mathcal{O}(Nd)$

TABLE 8. (Binary logistic regression) Per-iteration cost for each algorithm. N denotes the number of data points used for each iteration.

Proposition 5.2. *The function $U : \mathbb{R}^d \rightarrow \mathbb{R}$ given in (25) satisfies Assumptions 5 and 6.*

Proof. See Appendix A.2. □

For the numerical experiments, we set $\theta_0 = (2, 2, \dots, 2) \in \mathbb{R}^d$ to ensure that Assumption 4 is satisfied. Hence, Theorem 3.3 and Corollary 3.4 can be used to provide theoretical guarantees for aHOLLA to sample from π defined in (24). Furthermore, we set $\beta = 1$, $d = \{2, 5, 10\}$ and $N = \{100, 200, 300, 400, 500\}$. For each pair of d and N , and for each $i = 1, \dots, N$, we draw $z_i \in \mathbb{R}^d$ from a Rademacher distribution where each coordinate of z_i takes the value of 1 or -1 with probability $1/2$, and the corresponding y_i given z_i is drawn from a Bernoulli distribution with parameter $f_{\theta^*}(z_i)$, where θ^* is the true value set to be $(1, 1, \dots, 1) \in \mathbb{R}^d$. We generate samples from π using LMC [11, Eq. (2)], LMCO' [11, Eq. (17)], and aHOLLA (15)-(17). For each algorithm, we set the stepsize $\lambda = 0.01$ and the number of iterations $n = 100000$, then we run 50 independent Markov chains and compute the sample mean using the last 200 samples for each chain.

To compare the performance of LMC, LMCO', and aHOLLA, we follow the approach described in [10]. We compute the mean squared error (MSE) between the sample mean and θ^* averaged over 50 chains, and report the results in Table 6. We observe that the error increases when d increases while it decreases when N gets large. All three algorithms produce accurate and comparable results, indicating that they can be used to efficiently sample from the posterior distribution (24). Moreover, we present the running time and per-iteration cost for each algorithm in Tables 7 and 8, respectively. These results show that it takes less time for LMC to complete the experiment in each setting.

5.3. Conclusion. In this section, we illustrate the applicability of aHOLLA and aHOLA via two examples, namely, the high-dimensional sampling and the binary logistic regression. In the high-dimensional sampling example, aHOLLA and aHOLA exhibit superior empirical performance compared to the other algorithms in terms of the approximation accuracy measured in Wasserstein-1 distance. Crucially, the significant improvement in the accuracy is achieved by introducing only negligible (or small) amount of additional computation time. Thus, we recommend using our proposed algorithms for the problem of sampling from high-dimensional target distributions, especially those with highly non-linear potentials. Moreover, in the logistic regression example, we show that aHOLLA can be used to sample from a given posterior distribution, but it would take more time to complete the tasks compared to LMC.

6. PROOF OF MAIN RESULTS FOR AHOLA

In this section, we provide the proofs for Theorem 2.7 and Corollary 2.8. We first introduce auxiliary processes which we use throughout the convergence analysis. Then, we provide moment estimates for the newly introduced processes, which are followed by the detailed proofs for the main results. We postpone the proofs for the results presented in this section to Appendix A.3.

6.1. Auxiliary processes. Fix $\beta > 0$. Consider the Langevin SDE $(Z_t)_{t \geq 0}$ given by

$$Z_0 := \theta_0, \quad dZ_t = -h(Z_t) dt + \sqrt{2\beta^{-1}} dB_t, \quad (26)$$

where $(B_t)_{t \geq 0}$ is a d -dimensional Brownian motion on (Ω, \mathcal{F}, P) with its completed natural filtration denoted by $(\mathcal{F}_t)_{t \geq 0}$. Moreover, we assume that $(\mathcal{F}_t)_{t \geq 0}$ is independent of $\sigma(\theta_0)$. Under Assumptions 2 and 3, and by Remarks 2.4 and 2.5, we note that the Langevin SDE (26) admits a unique solution, which is adapted to $\mathcal{F}_t \vee \sigma(\theta_0)$, $t \geq 0$, due to [23, Theorem 1]. Its $2p$ -th moment estimate with $p \in \mathbb{N}$ is provided in Lemma A.1, which can be used to further deduce the $2p$ -th moment estimate of π_β (3).

For each $\lambda > 0$, recall $B_t^\lambda := B_{\lambda t}/\sqrt{\lambda}$, $t \geq 0$. Denote by $(\mathcal{F}_t^\lambda)_{t \geq 0}$ with $\mathcal{F}_t^\lambda := \mathcal{F}_{\lambda t}$, $t \geq 0$, its completed natural filtration, which is independent of $\sigma(\theta_0)$. Moreover, we denote by $Z_t^\lambda := Z_{\lambda t}$, $t \geq 0$, the time-changed version of Langevin SDE (26), which is given by

$$Z_0^\lambda := \theta_0, \quad dZ_t^\lambda = -\lambda h(Z_t^\lambda) dt + \sqrt{2\lambda\beta^{-1}} dB_t^\lambda. \quad (27)$$

Furthermore, we denote by $(\tilde{\theta}_t^\lambda)_{t \geq 0}$ the continuous-time interpolation of aHOLA (5)-(7) given by

$$\tilde{\theta}_0^\lambda := \theta_0, \quad d\tilde{\theta}_t^\lambda = \lambda \phi^\lambda(\tilde{\theta}_{[t]}^\lambda) dt + \sqrt{2\lambda\beta^{-1}} \psi^\lambda(\tilde{\theta}_{[t]}^\lambda) dB_t^\lambda, \quad (28)$$

where ϕ^λ and ψ^λ are defined in (6) and (7), respectively.

Remark 6.1. Similarly, denote by $(\bar{\theta}_t^\lambda)_{t \geq 0}$ the continuous-time interpolation of the order 1.5 scheme (9) given by

$$\begin{aligned} d\bar{\theta}_t^\lambda &= -\lambda h_\lambda(\bar{\theta}_{[t]}^\lambda) dt + \lambda^2 \int_{[t]}^t \left(H_\lambda(\bar{\theta}_{[s]}^\lambda) h_\lambda(\bar{\theta}_{[s]}^\lambda) - \beta^{-1} \Upsilon_\lambda(\bar{\theta}_{[s]}^\lambda) \right) ds dt \\ &\quad - \lambda \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t H_\lambda(\bar{\theta}_{[s]}^\lambda) dB_s^\lambda dt + \sqrt{2\lambda\beta^{-1}} dB_t^\lambda \end{aligned} \quad (29)$$

with $\bar{\theta}_0^\lambda := \theta_0$. We note that $\mathcal{L}(\bar{\theta}_n^\lambda) = \mathcal{L}(\theta_n^{\text{aHOLA}}) = \mathcal{L}(\theta_n^\lambda) = \mathcal{L}(\bar{\theta}_n^\lambda)$, for each $n \in \mathbb{N}_0$.

Finally, for any $s \geq 0$, consider the continuous-time process $(\zeta_t^{s,v,\lambda})_{t \geq s}$ defined by

$$\zeta_s^{s,v,\lambda} := v \in \mathbb{R}^d, \quad d\zeta_t^{s,v,\lambda} = -\lambda h(\zeta_t^{s,v,\lambda}) dt + \sqrt{2\lambda\beta^{-1}} dB_t^\lambda. \quad (30)$$

Definition 6.2. Fix $\lambda > 0$. Define $T \equiv T(\lambda) := \lfloor 1/\lambda \rfloor$. Then, for any $n \in \mathbb{N}_0$ and $t \geq nT$, define

$$\bar{\zeta}_t^{\lambda,n} := \zeta_t^{nT, \bar{\theta}_{nT}^\lambda, \lambda}.$$

6.2. Moment estimates. We first introduce the following Lyapunov functions: for each $p \in [2, \infty) \cap \mathbb{N}$, define $V_p(\theta) := (1 + |\theta|^2)^{p/2}$, for all $\theta \in \mathbb{R}^d$, as well as $v_p(w) := (1 + w^2)^{p/2}$, for all $w \geq 0$. We observe that V_p is twice continuously differentiable and satisfies:

$$\sup_{\theta \in \mathbb{R}^d} |\nabla V_p(\theta)|/V_p(\theta) < \infty, \quad \lim_{|\theta| \rightarrow \infty} \nabla V_p(\theta)/V_p(\theta) = 0. \quad (31)$$

Furthermore, we denote by $\mathcal{P}_{V_p}(\mathbb{R}^d)$ the set of probability measures $\mu \in \mathcal{P}(\mathbb{R}^d)$ which satisfies $\int_{\mathbb{R}^d} V_p(\theta) \mu(d\theta) < \infty$.

Next, we establish moment estimates for $(\tilde{\theta}_t^\lambda)_{t \geq 0}$ given in (28). The results with explicit constants are provided below. We note that for any $p \in [2, \infty) \cap \mathbb{N}$ and $t \geq 0$, we have that $\mathbb{E}[|\tilde{\theta}_t^\lambda|^{2p}] = \mathbb{E}[|\theta_t^\lambda|^{2p}]$.

Lemma 6.3. Let Assumptions 1, 2, and 3 hold. Then, we obtain the following estimates:

(i) For any $0 < \lambda \leq \lambda_{\max}$, $n \in \mathbb{N}_0$, and $t \in (n, n+1]$,

$$\mathbb{E} \left[|\tilde{\theta}_t^\lambda|^2 \right] \leq (1 - \lambda(t-n) a_D \kappa) (1 - \lambda a_D \kappa)^n \mathbb{E} [|\theta_0|^2] + c_0 (1 + 1/(a_D \kappa)),$$

where the constants c_0, κ are given explicitly in (75). In particular, the above inequality implies

$$\sup_{t \geq 0} \mathbb{E} \left[|\tilde{\theta}_t^\lambda|^2 \right] \leq \mathbb{E} [|\theta_0|^2] + c_0(1 + 1/(\mathfrak{a}_D \kappa)) < \infty.$$

(ii) For any $p \in [2, \infty) \cap \mathbb{N}$, $0 < \lambda \leq \lambda_{\max}$, $n \in \mathbb{N}_0$, and $t \in (n, n+1]$,

$$\mathbb{E} \left[|\tilde{\theta}_t^\lambda|^{2p} \right] \leq (1 - \lambda(t-n)\mathfrak{a}_D \kappa) (1 - \lambda \mathfrak{a}_D \kappa)^n \mathbb{E} [|\theta_0|^{2p}] + c_p (1 + 1/(\mathfrak{a}_D \kappa)),$$

where κ is given explicitly in (75) and c_p is given in (83). In particular, the above estimate implies

$$\sup_{t \geq 0} \mathbb{E} \left[|\tilde{\theta}_t^\lambda|^{2p} \right] \leq \mathbb{E} [|\theta_0|^{2p}] + c_p (1 + 1/(\mathfrak{a}_D \kappa)) < \infty.$$

Proof. See Appendix A.3. \square

We provide below a drift condition for V_p (defined in the beginning of Section 6.2), which is key to obtain the moment estimates of $(\bar{\zeta}_t^{\lambda, n})_{t \geq nT}$ defined in Definition 6.2.

Lemma 6.4. *Let Assumptions 1, 2, and 3 hold. Then, for any $p \in [2, \infty) \cap \mathbb{N}$, $\theta \in \mathbb{R}^d$, we obtain*

$$\Delta V_p(\theta)/\beta - \langle \nabla V_p(\theta), h(\theta) \rangle \leq -c_{V,1}(p)V_p(\theta) + c_{V,2}(p),$$

where $c_{V,1}(p) := \mathfrak{a}_D p/4$, $c_{V,2}(p) := (3/4)\mathfrak{a}_D p v_p(M_V(p))$ with

$$M_V(p) := (1/3 + 4\bar{\mathfrak{b}}_D/(3\mathfrak{a}_D) + 4d/(3\mathfrak{a}_D\beta) + 4(p-2)/(3\mathfrak{a}_D\beta))^{1/2}.$$

Proof. See [6, Lemma 3.5]. \square

By applying Lemma 6.3 and 6.4, we obtain the second and the fourth moment estimates for $(\bar{\zeta}_t^{\lambda, n})_{t \geq nT}$.

Lemma 6.5. *Let Assumptions 1, 2, and 3 hold. Then, for any $p \in \mathbb{N}$, $0 < \lambda \leq \lambda_{\max}$, $n \in \mathbb{N}_0$, and $t \geq nT$, we obtain*

$$\mathbb{E}[V_{2p}(\bar{\zeta}_t^{\lambda, n})] \leq 2^{p-1} e^{-\lambda \mathfrak{a}_D \min\{\kappa, 1/2\}t} \mathbb{E}[|\theta_0|^{2p}] + 2^{p-1} (c_p (1 + 1/(\mathfrak{a}_D \kappa)) + 1) + 3v_{2p}(M_V(2p)),$$

where κ and c_p are given in (75) and (83) (see also Lemma 6.3) and $M_V(2p)$ is given in Lemma 6.4.

Proof. See [33, Corollary 4.6]. \square

6.3. Proof of main results. In this section, we present key results used to obtain Theorem 2.7. We consider establishing a non-asymptotic error bound in Wasserstein-1 distance between the law of $\bar{\theta}_t^\lambda$ given in (29) and π_β , i.e., $W_1(\mathcal{L}(\bar{\theta}_t^\lambda), \pi_\beta)$, for any $n \in \mathbb{N}_0$ and $t \in (nT, (n+1)T]$. This and the fact that $W_1(\mathcal{L}(\theta_n^{\text{aHOLA}}), \pi_\beta) = W_1(\mathcal{L}(\bar{\theta}_t^\lambda), \pi_\beta)$ holds at any grid point yields the desired result. To this end, we split $W_1(\mathcal{L}(\bar{\theta}_t^\lambda), \pi_\beta)$ by using the law of $\bar{\zeta}_t^{\lambda, n}$ defined in Definition 6.2 and Z_t^λ given in (27) as follows:

$$W_1(\mathcal{L}(\bar{\theta}_t^\lambda), \pi_\beta) \leq W_1(\mathcal{L}(\bar{\theta}_t^\lambda), \mathcal{L}(\bar{\zeta}_t^{\lambda, n})) + W_1(\mathcal{L}(\bar{\zeta}_t^{\lambda, n}), \mathcal{L}(Z_t^\lambda)) + W_1(\mathcal{L}(Z_t^\lambda), \pi_\beta). \quad (32)$$

In the following lemma, we provide a non-asymptotic estimate for $W_2(\mathcal{L}(\bar{\theta}_t^\lambda), \mathcal{L}(\bar{\zeta}_t^{\lambda, n}))$, which can be used to upper bound the first term on the RHS of (32).

Lemma 6.6. *Let Assumptions 1, 2, and 3 hold. Then, for any $0 < \lambda \leq \lambda_{\max}$, $n \in \mathbb{N}_0$, and $t \in (nT, (n+1)T]$, we obtain*

$$W_2(\mathcal{L}(\bar{\theta}_t^\lambda), \mathcal{L}(\bar{\zeta}_t^{\lambda, n})) \leq \lambda^{1+q/2} \left(e^{-\mathfrak{a}_D \min\{\kappa, 1/2\}n/2} C_0 \mathbb{E}[|\theta_0|^{16(\rho+1)}] + C_1 \right)^{1/2},$$

where κ, C_0, C_1 are given explicitly in (75) and (107).

Proof. See Appendix A.3. \square

For the last two terms on the RHS of (32), we observe that they can be viewed as Wasserstein-1 distances between distributions of Langevin processes starting from different initial points. Therefore, to obtain their upper bounds, we introduce a semi-metric which allows us to establish a contraction result for the Langevin SDE (26) under our assumptions.

We consider the following semi-metric: for any $p \in [2, \infty) \cap \mathbb{N}$, $\mu, \nu \in \mathcal{P}_{V_p}(\mathbb{R}^d)$, let

$$w_{1,p}(\mu, \nu) := \inf_{\zeta \in \mathcal{C}(\mu, \nu)} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} [1 \wedge |\theta - \theta'|] (1 + V_p(\theta) + V_p(\theta')) \zeta(d\theta, d\theta'). \quad (33)$$

Then, we provide a result which states the contraction property of the Langevin SDE (26) in $w_{1,2}$.

Proposition 6.7. *Let Assumptions 1, 2, and 3 hold. Moreover, let $\theta'_0 \in \mathcal{L}^2$, and let $(Z'_t)_{t \geq 0}$ be the solution of SDE (26) whose starting point $Z'_0 := \theta'_0$ is assumed to be independent of $\mathcal{F}_\infty := \sigma(\bigcup_{t \geq 0} \mathcal{F}_t)$. Then, we obtain*

$$w_{1,2}(\mathcal{L}(Z_t), \mathcal{L}(Z'_t)) \leq \hat{c} e^{-\hat{c}t} w_{1,2}(\mathcal{L}(\theta_0), \mathcal{L}(\theta'_0)), \quad (34)$$

where the explicit expressions for \hat{c} , \hat{c} are given below.

The contraction constant \hat{c} is given by:

$$\hat{c} := \min\{\bar{\phi}, c_{V,1}(2), 4c_{V,2}(2)\epsilon c_{V,1}(2)\}/2,$$

where $c_{V,1}(2) := a_D/2$, $c_{V,2}(2) := 3a_D v_2(M_V(2))/2$ with $M_V(2) := (1/3 + 4\bar{b}_D/(3a_D) + 4d/(3a_D\beta))^{1/2}$, the constant $\bar{\phi}$ is given by

$$\bar{\phi} := \left(\sqrt{8\pi/(\beta L_{OS})} \hat{c}_0 \exp \left(\left(\hat{c}_0 \sqrt{\beta L_{OS}/8} + \sqrt{8/(\beta L_{OS})} \right)^2 \right) \right)^{-1},$$

and $\epsilon > 0$ is chosen such that

$$\epsilon \leq 1 \wedge \left(4c_{V,2}(2) \sqrt{2\beta\pi/L_{OS}} \int_0^{\hat{c}_1} \exp \left(\left(s \sqrt{\beta L_{OS}/8} + \sqrt{8/(\beta L_{OS})} \right)^2 \right) ds \right)^{-1}$$

with $\hat{c}_0 := 2(4c_{V,2}(2)(1 + c_{V,1}(2))/c_{V,1}(2) - 1)^{1/2}$ and $\hat{c}_1 := 2(2c_{V,2}(2)/c_{V,1}(2) - 1)^{1/2}$.

Moreover, the constant \hat{c} is given by:

$$\hat{c} := 2(1 + \hat{c}_0) \exp(\beta L_{OS} \hat{c}_0^2/8 + 2\hat{c}_0)/\epsilon.$$

Proof. We note that [16, Assumption 2.1] holds with $\kappa = L_{OS}$ due to Remark 2.5, [16, Assumption 2.2] holds with $V = V_2$ due to Remark 6.4, and [16, Assumptions 2.4 and 2.5] hold due to (31). Therefore, we can obtain (34) following the same arguments as in the proof of [6, Proposition 3.14] based on [16, Theorem 2.2, Corollary 2.3]. In addition, \hat{c} , \hat{c} can be obtained following the arguments in the proof of [27, Proposition 4.6]. \square

By using the above result and $W_1 \leq w_{1,2}$ (see [27, Lemma A.3]), we can establish a non-asymptotic error bound for the second term on the RHS of (32). The explicit statement is given below.

Lemma 6.8. *Let Assumptions 1, 2, and 3 hold. Then, for any $0 < \lambda \leq \lambda_{\max}$, $n \in \mathbb{N}_0$, and $t \in (nT, (n+1)T]$, we obtain*

$$W_1(\mathcal{L}(\bar{\zeta}_t^{\lambda,n}), \mathcal{L}(Z_t^\lambda)) \leq \lambda^{1+q/2} \left(e^{-\min\{\hat{c}, a_D\kappa, a_D/2\}n/4} C_2 \mathbb{E}[|\theta_0|^{16(\rho+1)}] + C_3 \right),$$

where

$$\begin{aligned} C_2 &:= \hat{c} \left(1 + \frac{4}{\min\{\hat{c}, a_D\kappa, a_D/2\}} \right) e^{\min\{\hat{c}, a_D\kappa, a_D/2\}/4} (C_0 + 2^{8\rho+11}), \\ C_3 &:= 2(\hat{c}/\hat{c}) e^{\hat{c}/2} (C_1 + 15 + 2^{8\rho+11} (c_{8(\rho+1)}(1 + 1/(a_D\kappa)) + 1) + 18v_{16(\rho+1)}(M_V(16(\rho+1)))) \end{aligned} \quad (35)$$

with \hat{c} , \hat{c} given in Proposition 6.7, C_0, C_1 given in (107) (see also Lemma 6.6), $\kappa, c_{8(\rho+1)}$ given in Lemma 6.3, and $M_V(16(\rho+1))$ given in Lemma 6.5.

Proof. See [27, Lemma 4.7]. \square

To obtain an upper bound for the last term on the RHS of (32), we observe that π_β is the invariant measure of the Langevin SDE (27). Thus, by applying Proposition 6.7, we have that

$$W_1(\mathcal{L}(Z_t^\lambda), \pi_\beta) \leq \hat{c} e^{-\hat{c}\lambda t} w_{1,2}(\mathcal{L}(\theta_0), \pi_\beta) \leq \hat{c} e^{-\hat{c}\lambda t} \left[1 + \mathbb{E}[V_2(\theta_0)] + \int_{\mathbb{R}^d} V_2(\theta) \pi_\beta(d\theta) \right]. \quad (36)$$

By using Lemma 6.6, 6.8 and (36), we can obtain an upper bound for each $W_2(\mathcal{L}(\theta_n^{\text{HOLA}}), \pi_\beta)$, $n \in \mathbb{N}_0$, as stated in Theorem 2.7.

Proof of Theorem 2.7. Substituting the results in Lemma 6.6, 6.8 and (36) into (32), for any $0 < \lambda \leq \lambda_{\max}$, $n \in \mathbb{N}_0$, and $t \in (nT, (n+1)T]$, we have that

$$\begin{aligned} W_1(\mathcal{L}(\bar{\theta}_t^\lambda), \pi_\beta) &\leq \lambda^{1+q/2} \left(e^{-a_D \min\{\kappa, 1/2\}n/2} C_0 \mathbb{E}[|\theta_0|^{16(\rho+1)}] + C_1 \right)^{1/2} \\ &\quad + \lambda^{1+q/2} \left(e^{-\min\{\hat{c}, a_D\kappa, a_D/2\}n/4} C_2 \mathbb{E}[|\theta_0|^{16(\rho+1)}] + C_3 \right) \end{aligned}$$

$$\begin{aligned}
& + \hat{c}e^{-\hat{c}\lambda t} \left[1 + \mathbb{E}[V_2(\theta_0)] + \int_{\mathbb{R}^d} V_2(\theta)\pi_\beta(d\theta) \right] \\
& \leq C_1 e^{-C_0(n+1)} (\mathbb{E}[|\theta_0|^{16(\rho+1)}] + 1) + C_2 \lambda^{1+q/2},
\end{aligned}$$

where

$$\begin{aligned}
C_0 & := \min\{\hat{c}, a_D \kappa, a_D/2\}/4, \\
C_1 & := e^{\min\{\hat{c}, a_D \kappa, a_D/2\}/4} \left[C_0^{1/2} + C_2 + \hat{c} \left(3 + \int_{\mathbb{R}^d} V_2(\theta)\pi_\beta(d\theta) \right) \right], \\
C_2 & := C_1^{1/2} + C_3
\end{aligned} \tag{37}$$

with \hat{c}, \hat{c} given in Proposition 6.7, κ given in Lemma 6.3, C_0, C_1 given in (107) (see also Lemma 6.6), C_2, C_3 given in (35) (see also Lemma 6.8). The above result implies that, for each $n \in \mathbb{N}_0$,

$$W_1(\mathcal{L}(\bar{\theta}_{nT}^\lambda), \pi_\beta) \leq C_1 e^{-C_0 n} (\mathbb{E}[|\theta_0|^{16(\rho+1)}] + 1) + C_2 \lambda^{1+q/2},$$

which further yields, by setting nT to n on the LHS and n to n/T on the RHS, that

$$W_1(\mathcal{L}(\bar{\theta}_n^\lambda), \pi_\beta) = W_1(\mathcal{L}(\theta_n^\lambda), \pi_\beta) = W_1(\mathcal{L}(\theta_n^{\text{HOLA}}), \pi_\beta) \leq C_1 e^{-C_0 \lambda n} (\mathbb{E}[|\theta_0|^{16(\rho+1)}] + 1) + C_2 \lambda^{1+q/2},$$

where the inequality holds due to $n\lambda \leq n/T$. This completes the proof. \square

By using similar arguments as in the proof of Theorem 2.7, we can obtain the upper bound for $W_2(\mathcal{L}(\theta_n^{\text{HOLA}}), \pi_\beta)$, $n \in \mathbb{N}_0$, as stated in Corollary 2.8.

Proof of Corollary 2.8. To establish a non-asymptotic error bound for $W_2(\mathcal{L}(\theta_n^{\text{HOLA}}), \pi_\beta)$, we consider the following splitting: for any $0 < \lambda \leq \lambda_{\max}$, $n \in \mathbb{N}_0$, and $t \in (nT, (n+1)T]$,

$$W_2(\mathcal{L}(\bar{\theta}_t^\lambda), \pi_\beta) \leq W_2(\mathcal{L}(\bar{\theta}_t^\lambda), \mathcal{L}(\bar{\zeta}_t^{\lambda, n})) + W_2(\mathcal{L}(\bar{\zeta}_t^{\lambda, n}), \mathcal{L}(Z_t^\lambda)) + W_2(\mathcal{L}(Z_t^\lambda), \pi_\beta). \tag{38}$$

An upper bound for the first term on the RHS of (38) is provided in Lemma 6.6. To establish an estimate for the second term on the RHS of (38), we use $W_2 \leq \sqrt{2w_{1,2}}$ (see [27, Lemma A.3] for the proof) and follow the same arguments as that in the proof of [27, Lemma 4.7]. Consequently, for any $0 < \lambda \leq \lambda_{\max}$, $n \in \mathbb{N}_0$, and $t \in (nT, (n+1)T]$, we obtain that,

$$W_2(\mathcal{L}(\bar{\zeta}_t^{\lambda, n}), \mathcal{L}(Z_t^\lambda)) \leq \lambda^{1/2+q/4} \left(e^{-\min\{\hat{c}, a_D \kappa, a_D/2\}n/8} C_4 \mathbb{E}^{1/2}[|\theta_0|^{16(\rho+1)}] + C_5 \right), \tag{39}$$

where

$$\begin{aligned}
C_4 & := \sqrt{\hat{c}} \left(1 + \frac{8}{\min\{\hat{c}, a_D \kappa, a_D/2\}} \right) e^{\min\{\hat{c}, a_D \kappa, a_D/2\}/8} \left(\sqrt{2} C_0^{1/2} + 2^{4\rho+5} \right), \\
C_5 & := 4(\sqrt{\hat{c}}/\hat{c}) e^{\hat{c}/4} \left(\sqrt{2} C_1^{1/2} + 3\sqrt{2} + 2^{4\rho+5} (c_{8(\rho+1)} (1 + 1/(a_D \kappa)) + 1)^{1/2} \right. \\
& \quad \left. + \sqrt{6} v_{16(\rho+1)}^{1/2} (M_V(16(\rho+1))) \right)
\end{aligned} \tag{40}$$

with \hat{c}, \hat{c} given in Proposition 6.7, C_0, C_1 given in (107) (see also Lemma 6.6), $\kappa, c_{8(\rho+1)}$ given in Lemma 6.3, and $M_V(16(\rho+1))$ given in Lemma 6.5. An upper bound for the last term on the RHS of (38) can be obtained by using $W_2 \leq \sqrt{2w_{1,2}}$ and Proposition 6.7:

$$\begin{aligned}
W_2(\mathcal{L}(Z_t^\lambda), \pi_\beta) & \leq \sqrt{2\hat{c}} e^{-\hat{c}\lambda t/2} w_{1,2}^{1/2}(\mathcal{L}(\theta_0), \pi_\beta) \\
& \leq \sqrt{2\hat{c}} e^{-\hat{c}\lambda t/2} \left[1 + \mathbb{E}[V_2(\theta_0)] + \int_{\mathbb{R}^d} V_2(\theta)\pi_\beta(d\theta) \right]^{1/2}.
\end{aligned} \tag{41}$$

Applying the results in Lemma 6.6, (39), (41) to (38) yields, for any $0 < \lambda \leq \lambda_{\max}$, $n \in \mathbb{N}_0$, and $t \in (nT, (n+1)T]$, that

$$\begin{aligned}
W_2(\mathcal{L}(\bar{\theta}_t^\lambda), \pi_\beta) & \leq \lambda^{1+q/2} \left(e^{-a_D \min\{\kappa, 1/2\}n/2} C_0 \mathbb{E}[|\theta_0|^{16(\rho+1)}] + C_1 \right)^{1/2} \\
& \quad + \lambda^{1/2+q/4} \left(e^{-\min\{\hat{c}, a_D \kappa, a_D/2\}n/8} C_4 \mathbb{E}^{1/2}[|\theta_0|^{16(\rho+1)}] + C_5 \right) \\
& \quad + \sqrt{2\hat{c}} e^{-\hat{c}\lambda t/2} \left[1 + \mathbb{E}[V_2(\theta_0)] + \int_{\mathbb{R}^d} V_2(\theta)\pi_\beta(d\theta) \right]^{1/2} \\
& \leq C_4 e^{-C_3(n+1)} (\mathbb{E}[|\theta_0|^{16(\rho+1)}] + 1)^{1/2} + C_5 \lambda^{1/2+q/4},
\end{aligned}$$

where

$$\begin{aligned}
C_3 &:= \min\{\hat{c}, a_D \kappa, a_D/2\}/8, \\
C_4 &:= e^{\min\{\hat{c}, a_D \kappa, a_D/2\}/8} \left[C_0^{1/2} + C_4 + \sqrt{2\hat{c}} \left(3 + \int_{\mathbb{R}^d} V_2(\theta) \pi_\beta(d\theta) \right)^{1/2} \right], \\
C_5 &:= C_1^{1/2} + C_5
\end{aligned} \tag{42}$$

with \hat{c}, \hat{c} given in Proposition 6.7, κ given in Lemma 6.3, C_0, C_1 given in (107) (see also Lemma 6.6), C_4, C_5 given in (40). This further implies that, for each $n \in \mathbb{N}_0$,

$$\begin{aligned}
W_2(\mathcal{L}(\bar{\theta}_n^\lambda), \pi_\beta) &= W_2(\mathcal{L}(\theta_n^\lambda), \pi_\beta) = W_2(\mathcal{L}(\theta_n^{\text{aHOLA}}), \pi_\beta) \\
&\leq C_4 e^{-C_3 \lambda n} (\mathbb{E}[|\theta_0|^{16(\rho+1)}] + 1)^{1/2} + C_5 \lambda^{1/2+q/4},
\end{aligned}$$

which completes the proof. □

APPENDIX A. PROOF OF AUXILIARY RESULTS

A.1. Proof of auxiliary results in Section 2.

Proof of statements in Remark 2.2. We provide detailed proofs for inequalities (10)-(12), the other inequalities can be obtained by using similar arguments. By Assumption 2, for any $\theta, \bar{\theta} \in \mathbb{R}^d$, $i = 1, \dots, d$, we have that

$$\begin{aligned} |\nabla^2 h^{(i)}(\theta)| &\leq |\nabla^2 h^{(i)}(\theta) - \nabla^2 h^{(i)}(0)| + |\nabla^2 h^{(i)}(0)| \\ &\leq L(1 + |\theta|)^{\rho-2} |\theta|^q + |\nabla^2 h^{(i)}(0)| \\ &\leq K_0(1 + |\theta|)^{\rho+q-2}, \end{aligned}$$

where $K_0 := 2^{1-q} \max\{L, |\nabla^2 h^{(1)}(0)|, \dots, |\nabla^2 h^{(d)}(0)|\}$. Moreover, fix $\theta, \bar{\theta} \in \mathbb{R}^d$, denote by $g(t) := \nabla h^{(i)}(t\theta + (1-t)\bar{\theta})$, $t \in [0, 1]$. Then, by using the above inequality, we obtain that

$$\begin{aligned} |\nabla h^{(i)}(\theta) - \nabla h^{(i)}(\bar{\theta})| &= \left| \int_0^1 \nabla^2 h^{(i)}(t\theta + (1-t)\bar{\theta})(\theta - \bar{\theta}) dt \right| \\ &\leq \int_0^1 K_0(1 + |t\theta + (1-t)\bar{\theta}|)^{\rho+q-2} dt |\theta - \bar{\theta}| \\ &\leq K_0(1 + |\theta| + |\bar{\theta}|)^{\rho+q-2} |\theta - \bar{\theta}|. \end{aligned}$$

In addition, for any $\theta, \bar{\theta} \in \mathbb{R}^d$, by Assumption 2, we have that

$$\begin{aligned} |\Upsilon(\theta) - \Upsilon(\bar{\theta})| &= \left(\sum_{i=1}^d \left(\sum_{j=1}^d (\partial_{\theta^{(j)}}^2 h^{(i)}(\theta) - \partial_{\theta^{(j)}}^2 h^{(i)}(\bar{\theta})) \right)^2 \right)^{1/2} \\ &\leq \left(d \sum_{i=1}^d \sum_{j=1}^d \left(\partial_{\theta^{(j)}}^2 h^{(i)}(\theta) - \partial_{\theta^{(j)}}^2 h^{(i)}(\bar{\theta}) \right)^2 \right)^{1/2} \\ &\leq \left(d \sum_{i=1}^d |\nabla^2 h^{(i)}(\theta) - \nabla^2 h^{(i)}(\bar{\theta})|_{\mathbb{F}}^2 \right)^{1/2} \\ &\leq \left(d^2 \sum_{i=1}^d |\nabla^2 h^{(i)}(\theta) - \nabla^2 h^{(i)}(\bar{\theta})|^2 \right)^{1/2} \\ &\leq d^{3/2} L(1 + |\theta| + |\bar{\theta}|)^{\rho-2} |\theta - \bar{\theta}|^q, \end{aligned}$$

which completes the proof. \square

Proof of statements in Remark 2.4. By Assumption 3, for any $\theta \in \mathbb{R}^d$, we have that

$$\langle \theta, h(\theta) - h(0) \rangle \geq a|\theta|^2 |\theta|^r - b|\theta|^2 |\theta|^{\bar{r}},$$

which implies that

$$\begin{aligned} \langle \theta, h(\theta) \rangle &\geq a|\theta|^{r+2} - b|\theta|^{\bar{r}+2} + \langle \theta, h(0) \rangle \\ &\geq a|\theta|^{r+2} - b|\theta|^{\bar{r}+2} - |\langle \theta, h(0) \rangle| \\ &\geq (a/2)|\theta|^{r+2} + ((a/4)|\theta|^{r+2} - b|\theta|^{\bar{r}+2}) + ((a/4)|\theta|^{r+2} - a|\theta|^2/2) - |h(0)|^2/(2a), \end{aligned} \quad (43)$$

where the last inequality holds due to the fact that $yz \leq \varepsilon y^2/2 + z^2/(2\varepsilon)$ for any $y, z \geq 0$ and $\varepsilon > 0$. Denote by $R_D := \max\{(4b/a)^{1/(r-\bar{r})}, 2^{1/r}\} > 1$. We observe that for any $|\theta| > R_D$,

$$(a/4)|\theta|^{r+2} - b|\theta|^{\bar{r}+2} > 0, \quad (a/4)|\theta|^{r+2} - a|\theta|^2/2 > 0,$$

and thus (43) becomes

$$\langle \theta, h(\theta) \rangle > (a/2)|\theta|^{r+2} - |h(0)|^2/(2a). \quad (44)$$

Moreover, for any $|\theta| \leq R_D$, (43) becomes

$$\langle \theta, h(\theta) \rangle \geq (a/2)|\theta|^{r+2} - (b + a/2)R_D^{\bar{r}+2} - |h(0)|^2/(2a). \quad (45)$$

Combining (44) and (45) yields

$$\langle \theta, h(\theta) \rangle \geq a_D |\theta|^{r+2} - b_D, \quad (46)$$

where $a_D := a/2$ and $b_D := (a/2 + b)R_D^{\bar{r}+2} + |h(0)|^2/(2a)$ with $R_D := \max\{(4b/a)^{1/(r-\bar{r})}, 2^{1/r}\}$. Moreover, by (46), for any $\theta \in \mathbb{R}^d$, we observe that,

$$\langle \theta, h(\theta) \rangle \geq a_D |\theta|^2 - \bar{b}_D,$$

where $\bar{b}_D := a_D + b_D$. Indeed, for $|\theta| > 1$, it holds that

$$\langle \theta, h(\theta) \rangle \geq a_D |\theta|^{r+2} - b_D > a_D |\theta|^2 - b_D > a_D |\theta|^2 - \bar{b}_D,$$

while for $|\theta| \leq 1$, we have that

$$\langle \theta, h(\theta) \rangle \geq a_D |\theta|^{r+2} - b_D \geq a_D - (a_D + b_D) \geq a_D |\theta|^2 - \bar{b}_D,$$

which completes the proof. \square

Proof of statements in Remark 2.5. For any $\theta, \bar{\theta} \in \mathbb{R}^d$, we observe that the result holds trivially when $\theta = \bar{\theta}$, and thus we consider only the case where $\theta \neq \bar{\theta}$. Denote by $R_{OS} := (b/a)^{1/(r-\bar{r})}$. By Assumption 3, for any $|\theta|, |\bar{\theta}| > R_{OS}$, we have that

$$\langle \theta - \bar{\theta}, h(\theta) - h(\bar{\theta}) \rangle > 0. \quad (47)$$

Furthermore, we note that by Remark 2.2,

$$|h(\theta) - h(\bar{\theta})| \leq \sqrt{d}K_1(1 + |\theta| + |\bar{\theta}|)^{\rho+q-1}|\theta - \bar{\theta}|.$$

Thus, for any $|\theta|, |\bar{\theta}| \leq R_{OS}$, we have that

$$-\langle \theta - \bar{\theta}, h(\theta) - h(\bar{\theta}) \rangle \leq |\theta - \bar{\theta}| |h(\theta) - h(\bar{\theta})| \leq L_{OS} |\theta - \bar{\theta}|^2, \quad (48)$$

where $L_{OS} := \sqrt{d}K_1(1 + 2R_{OS})^{\rho+q-1} > 0$. In addition, for $|\theta| \leq R_{OS}$, $|\bar{\theta}| > R_{OS}$, we consider the following two cases:

- (i) For $|\theta| = R_{OS}$, $|\bar{\theta}| > R_{OS}$, we obtain (47).
- (ii) For any $|\theta| < R_{OS}$, $|\bar{\theta}| > R_{OS}$, there is a unique $\hat{\theta} \in \mathbb{R}^d$ with $|\hat{\theta}| = |R_{OS}|$ such that

$$\theta - \hat{\theta} = c_{\theta, \bar{\theta}}(\theta - \bar{\theta}), \quad \hat{\theta} - \bar{\theta} = (1 - c_{\theta, \bar{\theta}})(\theta - \bar{\theta}),$$

where $c_{\theta, \bar{\theta}} \in (0, 1)$. Then, we obtain that

$$\begin{aligned} \langle \theta - \bar{\theta}, h(\theta) - h(\bar{\theta}) \rangle &= \langle \theta - \bar{\theta}, h(\theta) - h(\hat{\theta}) \rangle + \langle \theta - \hat{\theta}, h(\hat{\theta}) - h(\bar{\theta}) \rangle + \langle \hat{\theta} - \bar{\theta}, h(\hat{\theta}) - h(\bar{\theta}) \rangle \\ &= \langle \theta - \bar{\theta}, h(\theta) - h(\hat{\theta}) \rangle + (c_{\theta, \bar{\theta}}/(1 - c_{\theta, \bar{\theta}})) \langle \hat{\theta} - \bar{\theta}, h(\hat{\theta}) - h(\bar{\theta}) \rangle \\ &> (1/c_{\theta, \bar{\theta}}) \langle \theta - \hat{\theta}, h(\theta) - h(\hat{\theta}) \rangle \end{aligned} \quad (49)$$

$$\geq -(L_{OS}/c_{\theta, \bar{\theta}}) |\theta - \hat{\theta}|^2 \quad (50)$$

$$\begin{aligned} &= -L_{OS} c_{\theta, \bar{\theta}} |\theta - \hat{\theta}|^2 \\ &\geq -L_{OS} |\theta - \hat{\theta}|^2, \end{aligned} \quad (51)$$

where (49) holds due to (47), (50) holds due to (48), and (51) holds due to $c_{\theta, \bar{\theta}} \in (0, 1)$.

Thus, for $|\theta| \leq R_{OS}$, $|\bar{\theta}| > R_{OS}$, combining the two cases yield

$$\langle \theta - \bar{\theta}, h(\theta) - h(\bar{\theta}) \rangle \geq -L_{OS} |\theta - \bar{\theta}|^2. \quad (52)$$

Finally, for $|\theta| > R_{OS}$, $|\bar{\theta}| \leq R_{OS}$, we obtain (52) by applying the same arguments above. Combining (47), (48), and (52) yields the desired result. \square

A.2. Proof of auxiliary results in Section 5.

Proof of Proposition 5.1. We first show that the potentials given in (20) and (21) satisfy Assumptions 5 and 6.

(i) For U given in (20), we have that, for any $\theta \in \mathbb{R}^d$, $i = 1, \dots, d$,

$$h(\theta) = \theta, \quad H(\theta) = I_d, \quad \nabla^2 h^{(i)}(\theta) = 0.$$

Then, we observe that Assumption 5 holds with $\bar{L}_1 = 0$, $q = 1$, $\bar{L}_2 = 0$, and $\bar{L}_3 = 1$. Moreover, Assumption 6 holds with $\bar{a} = 1$, $\bar{b} = 0$ since

$$\langle \theta, h(\theta) \rangle = |\theta|^2.$$

(ii) For U given in (21), we have that, for any $\theta \in \mathbb{R}^d$, $i = 1, \dots, d$,

$$\begin{aligned} h(\theta) &= \theta - \hat{a} + \frac{2\hat{a}}{1 + \exp(2\langle \theta, \hat{a} \rangle)}, \\ H(\theta) &= I_d - \frac{4\hat{a}\hat{a}^\top \exp(2\langle \theta, \hat{a} \rangle)}{(1 + \exp(2\langle \theta, \hat{a} \rangle))^2}, \\ \nabla^2 h^{(i)}(\theta) &= \frac{8\hat{a}^{(i)}\hat{a}\hat{a}^\top (\exp(4\langle \theta, \hat{a} \rangle) - \exp(2\langle \theta, \hat{a} \rangle))}{(1 + \exp(2\langle \theta, \hat{a} \rangle))^3}. \end{aligned}$$

To show that Assumption 5 holds, we consider the following calculations. For any $\theta, \bar{\theta} \in \mathbb{R}^d$, we have that

$$\begin{aligned} &|h(\theta) - h(\bar{\theta})| \\ &= \left| \theta - \hat{a} + \frac{2\hat{a}}{1 + \exp(2\langle \theta, \hat{a} \rangle)} - \left(\bar{\theta} - \hat{a} + \frac{2\hat{a}}{1 + \exp(2\langle \bar{\theta}, \hat{a} \rangle)} \right) \right| \\ &\leq |\theta - \bar{\theta}| + 2|\hat{a}| \frac{|\exp(2\langle \bar{\theta}, \hat{a} \rangle) - \exp(2\langle \theta, \hat{a} \rangle)|}{(1 + \exp(2\langle \theta, \hat{a} \rangle))(1 + \exp(2\langle \bar{\theta}, \hat{a} \rangle))} \\ &= |\theta - \bar{\theta}| + 2|\hat{a}| \frac{\max\{\exp(2\langle \theta, \hat{a} \rangle), \exp(2\langle \bar{\theta}, \hat{a} \rangle)\} |1 - \exp(-2|\langle \theta, \hat{a} \rangle - \langle \bar{\theta}, \hat{a} \rangle||)}{(1 + \exp(2\langle \theta, \hat{a} \rangle))(1 + \exp(2\langle \bar{\theta}, \hat{a} \rangle))} \quad (53) \\ &\leq |\theta - \bar{\theta}| + 4|\hat{a}| |\langle \theta, \hat{a} \rangle - \langle \bar{\theta}, \hat{a} \rangle| \\ &\leq (1 + 4|\hat{a}|^2) |\theta - \bar{\theta}|, \end{aligned}$$

where the second last inequality holds due to $1 - e^{-x} \leq x$, for all $x \in \mathbb{R}$, and the last inequality holds due to Cauchy-Schwarz inequality. Similarly, we have that, for any $\theta, \bar{\theta} \in \mathbb{R}^d$,

$$\begin{aligned} &|H(\theta) - H(\bar{\theta})| \\ &= 4|\hat{a}|^2 \frac{|\exp(2\langle \bar{\theta}, \hat{a} \rangle)(1 + \exp(2\langle \theta, \hat{a} \rangle))^2 - \exp(2\langle \theta, \hat{a} \rangle)(1 + \exp(2\langle \bar{\theta}, \hat{a} \rangle))^2|}{(1 + \exp(2\langle \theta, \hat{a} \rangle))^2(1 + \exp(2\langle \bar{\theta}, \hat{a} \rangle))^2} \\ &\leq 4|\hat{a}|^2 \frac{(1 + \exp(2\langle \bar{\theta} + \theta, \hat{a} \rangle)) |\exp(2\langle \bar{\theta}, \hat{a} \rangle) - \exp(2\langle \theta, \hat{a} \rangle)|}{(1 + \exp(2\langle \theta, \hat{a} \rangle))^2(1 + \exp(2\langle \bar{\theta}, \hat{a} \rangle))^2} \\ &\leq 8|\hat{a}|^3 |\theta - \bar{\theta}|, \end{aligned}$$

where the last inequality holds due to the calculations in (53). Furthermore, we note that, for any $\theta, \bar{\theta} \in \mathbb{R}^d$,

$$\begin{aligned} &|\nabla^2 h^{(i)}(\theta) - \nabla^2 h^{(i)}(\bar{\theta})| \\ &\leq 8|\hat{a}|^3 \left| \frac{\exp(4\langle \theta, \hat{a} \rangle) - \exp(2\langle \theta, \hat{a} \rangle)}{(1 + \exp(2\langle \theta, \hat{a} \rangle))^3} - \frac{\exp(4\langle \bar{\theta}, \hat{a} \rangle) - \exp(2\langle \bar{\theta}, \hat{a} \rangle)}{(1 + \exp(2\langle \bar{\theta}, \hat{a} \rangle))^3} \right| \\ &\leq 8|\hat{a}|^3 \frac{(1 + \exp(2\langle \bar{\theta} + \theta, \hat{a} \rangle)) |\exp(4\langle \bar{\theta}, \hat{a} \rangle) - \exp(4\langle \theta, \hat{a} \rangle)|}{(1 + \exp(2\langle \theta, \hat{a} \rangle))^3(1 + \exp(2\langle \bar{\theta}, \hat{a} \rangle))^3} \\ &\quad + 8|\hat{a}|^3 \frac{(1 + 6\exp(2\langle \bar{\theta} + \theta, \hat{a} \rangle) + \exp(4\langle \bar{\theta} + \theta, \hat{a} \rangle)) |\exp(2\langle \bar{\theta}, \hat{a} \rangle) - \exp(2\langle \theta, \hat{a} \rangle)|}{(1 + \exp(2\langle \theta, \hat{a} \rangle))^3(1 + \exp(2\langle \bar{\theta}, \hat{a} \rangle))^3} \\ &\leq 56|\hat{a}|^4 |\theta - \bar{\theta}|, \end{aligned}$$

where the last inequality holds due to the calculations in (53). Consequently, Assumption 5 holds with $\bar{L}_1 = 56|\hat{a}|^4$, $q = 1$, $\bar{L}_2 = 8|\hat{a}|^3$, and $\bar{L}_3 = 1 + 4|\hat{a}|^2$. Now, we show Assumption 6 holds with $\bar{a} = 1/2$, $\bar{b} = 1/2$. Indeed, we have, for any $\theta \in \mathbb{R}^d$, that

$$\langle \theta, h(\theta) \rangle = |\theta|^2 + \frac{1 - \exp(2\langle \theta, \hat{a} \rangle)}{1 + \exp(2\langle \theta, \hat{a} \rangle)} \langle \theta, \hat{a} \rangle \geq |\theta|^2 - \frac{1}{2}(|\theta|^2 + |\hat{a}|^2) = \frac{1}{2}(|\theta|^2 - |\hat{a}|^2).$$

Finally, we show that the potential (22) satisfies Assumptions 2 and 3.

(iii) For U given in (22), by Remark 2.3, we have that Assumption 2 is satisfied with $\rho = 2$, $L = 6$, $q = 1$, $K_h = 2$, and $K_H = 3$. In addition, Assumption 3 is satisfied with $a = 1/2$, $b = 1$, $r = 2$, and $\bar{r} = 0$. Indeed, for any $\theta \in \mathbb{R}^d$, by noticing

$$h(\theta) = \theta(|\theta|^2 - 1),$$

we obtain that

$$\begin{aligned} \langle \theta - \bar{\theta}, h(\theta) - h(\bar{\theta}) \rangle &= \langle \theta - \bar{\theta}, \theta|\theta|^2 - \theta - (\bar{\theta}|\bar{\theta}|^2 - \bar{\theta}) \rangle \\ &= \frac{1}{2} \langle \theta - \bar{\theta}, \theta|\theta|^2 - \theta|\bar{\theta}|^2 + \theta|\bar{\theta}|^2 - \bar{\theta}|\bar{\theta}|^2 \rangle \\ &\quad + \frac{1}{2} \langle \theta - \bar{\theta}, \theta|\theta|^2 - \bar{\theta}|\theta|^2 + \bar{\theta}|\theta|^2 - \bar{\theta}|\bar{\theta}|^2 \rangle - |\theta - \bar{\theta}|^2 \\ &= \frac{1}{2}(|\theta|^2 + |\bar{\theta}|^2)|\theta - \bar{\theta}|^2 + \frac{1}{2}(|\theta|^2 - |\bar{\theta}|^2)^2 - |\theta - \bar{\theta}|^2 \\ &\geq \frac{1}{2}(|\theta|^2 + |\bar{\theta}|^2)|\theta - \bar{\theta}|^2 - |\theta - \bar{\theta}|^2. \end{aligned}$$

This completes the proof. \square

Proof of Proposition 5.2. For U defined in (25), we note, for any $\theta \in \mathbb{R}^d$, $l = 1, \dots, d$, that

$$\begin{aligned} h(\theta) &= \theta + \sum_{i=1}^N (1 - y_i) z_i - \sum_{i=1}^N \frac{z_i}{1 + e^{\langle \theta, z_i \rangle}}, \quad H(\theta) = I_d + \sum_{i=1}^N \frac{z_i z_i^\top e^{\langle \theta, z_i \rangle}}{(1 + e^{\langle \theta, z_i \rangle})^2}, \\ \nabla^2 h^{(l)}(\theta) &= \sum_{i=1}^N \frac{z_i^{(l)} z_i z_i^\top e^{\langle \theta, z_i \rangle} (1 - e^{\langle \theta, z_i \rangle})}{(1 + e^{\langle \theta, z_i \rangle})^3}. \end{aligned}$$

We first show that Assumption 5 holds. To this end, denote by

$$g_1(t) := (e^t - e^{2t})/(1 + e^t)^3, \quad g_2(t) := e^t/(1 + e^t)^2, \quad g_3(t) = 1/(1 + e^t),$$

for all $t \in \mathbb{R}$. We have that

$$\begin{aligned} \sup_{t \in \mathbb{R}} |g_3'(t)| &= \sup_{t \in \mathbb{R}} |g_2(t)| \leq 1, \quad \sup_{t \in \mathbb{R}} |g_2'(t)| = \sup_{t \in \mathbb{R}} |g_1(t)| \leq 1, \\ \sup_{t \in \mathbb{R}} |g_1'(t)| &= \sup_{t \in \mathbb{R}} |e^t(1 - 4e^t + e^{2t})/(1 + e^t)^4| \leq 1, \end{aligned}$$

which implies, for all $l = 1, \dots, d$, $\theta, \bar{\theta} \in \mathbb{R}^d$, that

$$\begin{aligned} |\nabla^2 h^{(l)}(\theta) - \nabla^2 h^{(l)}(\bar{\theta})| &= \left| \sum_{i=1}^N z_i^{(l)} z_i z_i^\top (g_1(\langle \theta, z_i \rangle) - g_1(\langle \bar{\theta}, z_i \rangle)) \right| \leq \sum_{i=1}^N |z_i|^4 |\theta - \bar{\theta}|, \\ |H(\theta) - H(\bar{\theta})| &= \left| \sum_{i=1}^N z_i z_i^\top (g_2(\langle \theta, z_i \rangle) - g_2(\langle \bar{\theta}, z_i \rangle)) \right| \leq \sum_{i=1}^N |z_i|^3 |\theta - \bar{\theta}|, \\ |h(\theta) - h(\bar{\theta})| &= \left| \theta - \bar{\theta} - \sum_{i=1}^N z_i (g_3(\langle \theta, z_i \rangle) - g_3(\langle \bar{\theta}, z_i \rangle)) \right| \leq (1 + \sum_{i=1}^N |z_i|^2) |\theta - \bar{\theta}|. \end{aligned}$$

Thus, Assumption 5 holds with $\bar{L}_1 = \sum_{i=1}^N |z_i|^4$, $\bar{L}_2 = \sum_{i=1}^N |z_i|^3$, $\bar{L}_3 = (1 + \sum_{i=1}^N |z_i|^2)$.

Moreover, to show that Assumption 6 holds, we note, for any $\theta \in \mathbb{R}^d$, that

$$\langle \theta, h(\theta) \rangle = |\theta|^2 + \sum_{i=1}^N (1 - y_i) \langle z_i, \theta \rangle - \sum_{i=1}^N \frac{\langle z_i, \theta \rangle}{1 + e^{\langle \theta, z_i \rangle}} \geq |\theta|^2/2 - \left(\sum_{i=1}^N |1 - y_i| |z_i| \right)^2 / 2 - N.$$

This implies that Assumption 6 holds with $\bar{a} = 1/2$ and $\bar{b} = \left(\sum_{i=1}^N |1 - y_i| |z_i|\right)^2 / 2 + N$. \square

A.3. Proof of auxiliary results in Section 6.

Lemma A.1. *Let Assumptions 1, 2, and 3 hold. Then, for any $p \in \mathbb{N}, t \geq 0$, we obtain that*

$$\mathbb{E}[|Z_t^\lambda|^{2p}] \leq e^{-\lambda p a_D t} \mathbb{E}[|\theta_0|^{2p}] + 2(\bar{b}_D + \beta^{-1}(d + 2(p-1))) M_0^{2p-2} / a_D < \infty,$$

where $M_0 := (2(\bar{b}_D + \beta^{-1}(d + 2(p-1))) / a_D)^{1/2}$.

Proof. See [27, Lemma A.1]. \square

Proof of Lemma 6.3-(i). For any $0 < \lambda \leq \lambda_{\max} \leq 1$ with λ_{\max} given in (14), $t \in (n, n+1]$, $n \in \mathbb{N}_0$, we define

$$\Delta_{n,t}^\lambda := \tilde{\theta}_n^\lambda + \lambda \phi^\lambda(\tilde{\theta}_n^\lambda)(t-n), \quad \Xi_{n,t}^\lambda := \sqrt{2\lambda\beta^{-1}} \psi^\lambda(\tilde{\theta}_n^\lambda)(B_t^\lambda - B_n^\lambda), \quad (54)$$

where, for all $\theta \in \mathbb{R}^d$,

$$\phi^\lambda(\theta) := -h_\lambda(\theta) + (\lambda/2) (H_\lambda(\theta)h_\lambda(\theta) - \beta^{-1}\Upsilon_\lambda(\theta)), \quad (55)$$

and

$$\psi^\lambda(\theta) := \sqrt{I_d - \lambda H_\lambda(\theta) + (\lambda^2/3)(H_\lambda(\theta))^2}. \quad (56)$$

By using (28) together with (54) – (56), and by noticing $\mathbb{E}[\langle \Delta_{n,t}^\lambda, \Xi_{n,t}^\lambda \rangle | \tilde{\theta}_n^\lambda] = 0$, we obtain that

$$\mathbb{E}[|\tilde{\theta}_t^\lambda|^2 | \tilde{\theta}_n^\lambda] = |\Delta_{n,t}^\lambda|^2 + \mathbb{E}[|\Xi_{n,t}^\lambda|^2 | \tilde{\theta}_n^\lambda]. \quad (57)$$

Further calculations yield the following upper bound for the second term on the RHS of (57):

$$\begin{aligned} \mathbb{E}[|\Xi_{n,t}^\lambda|^2 | \tilde{\theta}_n^\lambda] &= 2\lambda\beta^{-1} \mathbb{E}\left[\left\langle \psi^\lambda(\tilde{\theta}_n^\lambda)(B_t^\lambda - B_n^\lambda), \psi^\lambda(\tilde{\theta}_n^\lambda)(B_t^\lambda - B_n^\lambda) \right\rangle \middle| \tilde{\theta}_n^\lambda\right] \\ &= 2\lambda\beta^{-1} \mathbb{E}\left[\left\langle B_t^\lambda - B_n^\lambda, \left(I_d - \lambda H_\lambda(\tilde{\theta}_n^\lambda) + (\lambda^2/3)(H_\lambda(\tilde{\theta}_n^\lambda))^2\right) (B_t^\lambda - B_n^\lambda) \right\rangle \middle| \tilde{\theta}_n^\lambda\right] \\ &\leq 2\lambda\beta^{-1} \left(d(t-n) + \lambda \sum_{i=1}^d |H_\lambda^{(i,i)}(\tilde{\theta}_n^\lambda)|(t-n) + (\lambda^2/3) |H_\lambda(\tilde{\theta}_n^\lambda)|_{\mathbb{F}}^2(t-n) \right) \\ &\leq 2\lambda\beta^{-1}(t-n) \left(d + \lambda\sqrt{d} |H_\lambda(\tilde{\theta}_n^\lambda)|_{\mathbb{F}} + (\lambda^2/3) |H_\lambda(\tilde{\theta}_n^\lambda)|_{\mathbb{F}}^2 \right). \end{aligned} \quad (58)$$

By Remark 2.2, we have that, for all $\theta \in \mathbb{R}^d$,

$$|H(\theta)|_{\mathbb{F}} = \left(\sum_{i=1}^d |\nabla h^{(i)}(\theta)|^2 \right)^{1/2} \leq \left(\sum_{i=1}^d \mathbf{K}_1^2 (1 + |\theta|)^{2(\rho+q-1)} \right)^{1/2} = \sqrt{d} \mathbf{K}_1 (1 + |\theta|)^{\rho+q-1},$$

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

$$\begin{aligned} |H_\lambda(\theta)|_{\mathbb{F}} &= \frac{|H(\theta)|_{\mathbb{F}}}{(1 + \lambda^{3/2} |\theta|^{3(\rho+q-1)})^{1/3}} \leq \frac{4^{1/3} |H(\theta)|_{\mathbb{F}}}{1 + \lambda^{1/2} |\theta|^{(\rho+q-1)}} \leq \frac{4^{1/3} \lambda^{-1/2} |H(\theta)|_{\mathbb{F}}}{1 + |\theta|^{(\rho+q-1)}} \\ &\leq \frac{4^{1/3} \lambda^{-1/2} \sqrt{d} \mathbf{K}_1 (1 + |\theta|)^{\rho+q-1}}{1 + |\theta|^{(\rho+q-1)}} \leq 2^{\rho+q-1} \lambda^{-1/2} \sqrt{d} \mathbf{K}_1. \end{aligned} \quad (59)$$

Substituting (59) into (58) yields

$$\begin{aligned} \mathbb{E}[|\Xi_{n,t}^\lambda|^2 | \tilde{\theta}_n^\lambda] &\leq 2\lambda\beta^{-1}(t-n) \left(d + 2^{\rho+q-1} d \mathbf{K}_1 + 2^{2(\rho+q-1)} d \mathbf{K}_1^2 \right) \\ &\leq 2^{2(\rho+q)} \lambda \beta^{-1} (t-n) d (1 + \mathbf{K}_1)^2, \end{aligned} \quad (60)$$

where we use $1 + 2^{\rho+q-1} + 2^{2(\rho+q-1)} \leq 2^{2\rho+2q-3} + 2^{2\rho+2q-3} + 2^{2(\rho+q-1)} \leq 2^{2(\rho+q)-1}$ with $\rho \in [2, \infty) \cap \mathbb{N}$ and $q \in (0, 1]$ to obtain the last inequality. To upper bound the first term on the RHS of (57), we use (54) to obtain

$$|\Delta_{n,t}^\lambda|^2 = |\tilde{\theta}_n^\lambda|^2 + 2\lambda(t-n) \left\langle \tilde{\theta}_n^\lambda, \phi^\lambda(\tilde{\theta}_n^\lambda) \right\rangle + \lambda^2(t-n)^2 |\phi^\lambda(\tilde{\theta}_n^\lambda)|^2. \quad (61)$$

By using (55), the second term on the RHS of (61) can be estimated as follows:

$$\begin{aligned}
\langle \tilde{\theta}_n^\lambda, \phi^\lambda(\tilde{\theta}_n^\lambda) \rangle &= -\frac{\langle \tilde{\theta}_n^\lambda, h(\tilde{\theta}_n^\lambda) \rangle}{\left(1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}\right)^{1/3}} + \frac{(\lambda/2) \langle \tilde{\theta}_n^\lambda, H(\tilde{\theta}_n^\lambda) h(\tilde{\theta}_n^\lambda) \rangle}{\left(1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}\right)^{2/3}} \\
&\quad - \frac{(\lambda/2) \beta^{-1} \langle \tilde{\theta}_n^\lambda, \Upsilon(\tilde{\theta}_n^\lambda) \rangle}{\left(1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}\right)^{1/3}} \\
&\leq -\frac{\mathbf{a}_D |\tilde{\theta}_n^\lambda|^{\rho+q+1} - \mathbf{b}_D}{\left(1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}\right)^{1/3}} + \frac{2\lambda K_h K_H \left(1 + |\tilde{\theta}_n^\lambda|^{2(\rho+q)}\right)}{\left(1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}\right)^{2/3}} \\
&\quad + \frac{\lambda \beta^{-1} 2^{\rho+q-3} \mathbf{K}_{3,d} \left(1 + |\tilde{\theta}_n^\lambda|^{\rho+q-1}\right)}{\left(1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}\right)^{1/3}} \\
&\leq -\frac{\mathbf{a}_D |\tilde{\theta}_n^\lambda|^{\rho+q+1}}{\left(1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}\right)^{1/3}} + \frac{2\lambda K_h K_H |\tilde{\theta}_n^\lambda|^{2(\rho+q)}}{\left(1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}\right)^{2/3}} \\
&\quad + \mathbf{b}_D + 2K_h K_H + \beta^{-1} 2^{\rho+q-2} \mathbf{K}_{3,d},
\end{aligned} \tag{62}$$

where the first inequality holds due to Remark 2.4 with $r = \rho + q - 1$ and the fact that: for all $\theta \in \mathbb{R}^d$,

$$\begin{aligned}
|\theta| |H(\theta)| |h(\theta)| &\leq K_H K_h |\theta| (1 + |\theta|^{\rho+q-1}) (1 + |\theta|^{\rho+q}) \\
&= K_H K_h (|\theta| + |\theta|^{\rho+q} + |\theta|^{\rho+q+1} + |\theta|^{2(\rho+q)}) \leq 4K_h K_H (1 + |\theta|^{2(\rho+q)}), \\
|\theta| |\Upsilon(\theta)| &\leq \mathbf{K}_{3,d} |\theta| (1 + |\theta|)^{\rho+q-2} \leq \mathbf{K}_{3,d} (1 + |\theta|)^{\rho+q-1} \leq 2^{\rho+q-2} \mathbf{K}_{3,d} (1 + |\theta|^{\rho+q-1}),
\end{aligned}$$

where we recall from our assumptions that $\rho \in [2, \infty) \cap \mathbb{N}$ and $q \in (0, 1]$, and where the last inequality holds since, for all $\theta \in \mathbb{R}^d$,

$$\left(1 + \lambda^{3/2} |\theta|^{3(\rho+q-1)}\right)^{1/3} \geq 4^{-1/3} (1 + \lambda^{1/2} |\theta|^{\rho+q-1}) \geq 4^{-1/3} \lambda^{1/2} (1 + |\theta|^{\rho+q-1}).$$

Similarly, the third term on the RHS of (61) can be upper bounded as follows:

$$\begin{aligned}
|\phi^\lambda(\tilde{\theta}_n^\lambda)|^2 &= | -h_\lambda(\tilde{\theta}_n^\lambda) + (\lambda/2) H_\lambda(\tilde{\theta}_n^\lambda) h_\lambda(\tilde{\theta}_n^\lambda) - (\lambda/2) \beta^{-1} \Upsilon_\lambda(\tilde{\theta}_n^\lambda) |^2 \\
&= |h_\lambda(\tilde{\theta}_n^\lambda)|^2 + (\lambda^2/4) |H_\lambda(\tilde{\theta}_n^\lambda) h_\lambda(\tilde{\theta}_n^\lambda)|^2 + (\lambda^2/4) \beta^{-2} |\Upsilon_\lambda(\tilde{\theta}_n^\lambda)|^2 \\
&\quad - \lambda \langle h_\lambda(\tilde{\theta}_n^\lambda), H_\lambda(\tilde{\theta}_n^\lambda) h_\lambda(\tilde{\theta}_n^\lambda) \rangle + \lambda \beta^{-1} \langle h_\lambda(\tilde{\theta}_n^\lambda), \Upsilon_\lambda(\tilde{\theta}_n^\lambda) \rangle \\
&\quad - (\lambda^2/2) \beta^{-1} \langle H_\lambda(\tilde{\theta}_n^\lambda) h_\lambda(\tilde{\theta}_n^\lambda), \Upsilon_\lambda(\tilde{\theta}_n^\lambda) \rangle.
\end{aligned} \tag{63}$$

By using Assumption 2, Remark 2.2, and the inequalities $(u + w)^v \geq 2^{v-1} (u^v + w^v)$, $(u + w)^\nu \leq 2^{\nu-1} (u^\nu + w^\nu)$, $u, w \geq 0$, $0 < v \leq 1$, $\nu \geq 1$, $|\theta|^k \leq 1 + |\theta|^l$, $\theta \in \mathbb{R}^d$, $k \leq l$, we obtain

$$\begin{aligned}
|h_\lambda(\tilde{\theta}_n^\lambda)|^2 &\leq \frac{K_h^2 \left(1 + |\tilde{\theta}_n^\lambda|^{\rho+q}\right)^2}{\left(1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}\right)^{2/3}} \leq \frac{2K_h^2 \left(1 + |\tilde{\theta}_n^\lambda|^{2(\rho+q)}\right)}{\left(1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}\right)^{2/3}} \\
&\leq \frac{2K_h^2 |\tilde{\theta}_n^\lambda|^{2(\rho+q)}}{\left(1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}\right)^{2/3}} + 2K_h^2, \\
(\lambda^2/4) |H_\lambda(\tilde{\theta}_n^\lambda) h_\lambda(\tilde{\theta}_n^\lambda)|^2 &\leq \frac{(\lambda^2/4) K_H^2 K_h^2 \left(1 + |\tilde{\theta}_n^\lambda|^{\rho+q-1}\right)^2 \left(1 + |\tilde{\theta}_n^\lambda|^{\rho+q}\right)^2}{\left(1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}\right)^{4/3}}
\end{aligned} \tag{64}$$

$$\begin{aligned}
&\leq \frac{\lambda^2 K_H^2 K_h^2 \left(1 + |\tilde{\theta}_n^\lambda|^{2(\rho+q-1)} + |\tilde{\theta}_n^\lambda|^{2(\rho+q)} + |\tilde{\theta}_n^\lambda|^{4(\rho+q)-2}\right)}{\left(1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}\right)^{4/3}} \\
&\leq \frac{3\lambda^2 K_H^2 K_h^2 \left(1 + |\tilde{\theta}_n^\lambda|^{4(\rho+q)-2}\right)}{\left(1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}\right)^{4/3}} \\
&\leq \frac{3\lambda^2 K_H^2 K_h^2 |\tilde{\theta}_n^\lambda|^{4(\rho+q)-2}}{\left(1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}\right)^{4/3}} + 3K_H^2 K_h^2, \tag{65}
\end{aligned}$$

$$\begin{aligned}
(\lambda^2/4)\beta^{-2} |\Upsilon_\lambda(\tilde{\theta}_n^\lambda)|^2 &\leq \frac{(\lambda^2/4)\beta^{-2} \mathbf{K}_{3,d}^2 \left(1 + |\tilde{\theta}_n^\lambda|\right)^{2(\rho+q-2)}}{\left(1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}\right)^{2/3}} \\
&\leq \frac{2^{2(\rho+q)-7} \lambda^2 \beta^{-2} \mathbf{K}_{3,d}^2 \left(1 + |\tilde{\theta}_n^\lambda|^{2(\rho+q-2)}\right)}{\left(1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}\right)^{2/3}} \\
&\leq \frac{2^{2(\rho+q)-6} \lambda^2 \beta^{-2} \mathbf{K}_{3,d}^2 \left(1 + |\tilde{\theta}_n^\lambda|^{2(\rho+q-2)}\right)}{1 + \lambda |\tilde{\theta}_n^\lambda|^{2(\rho+q-1)}} \\
&\leq 2^{2(\rho+q)-6} \beta^{-2} \mathbf{K}_{3,d}^2, \tag{66}
\end{aligned}$$

$$\begin{aligned}
-\lambda \left\langle h_\lambda(\tilde{\theta}_n^\lambda), H_\lambda(\tilde{\theta}_n^\lambda) h_\lambda(\tilde{\theta}_n^\lambda) \right\rangle &\leq \frac{\lambda K_h^2 K_H \left(1 + |\tilde{\theta}_n^\lambda|^{\rho+q}\right)^2 \left(1 + |\tilde{\theta}_n^\lambda|^{\rho+q-1}\right)}{1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}} \\
&\leq \frac{2\lambda K_h^2 K_H \left(1 + |\tilde{\theta}_n^\lambda|^{2(\rho+q)} + |\tilde{\theta}_n^\lambda|^{\rho+q-1} + |\tilde{\theta}_n^\lambda|^{3(\rho+q)-1}\right)}{1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}} \\
&\leq \frac{6\lambda K_h^2 K_H \left(1 + |\tilde{\theta}_n^\lambda|^{3(\rho+q)-1}\right)}{1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}} \\
&\leq \frac{6\lambda K_h^2 K_H |\tilde{\theta}_n^\lambda|^{3(\rho+q)-1}}{1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}} + 6K_h^2 K_H, \tag{67}
\end{aligned}$$

$$\begin{aligned}
\lambda\beta^{-1} \left\langle h_\lambda(\tilde{\theta}_n^\lambda), \Upsilon_\lambda(\tilde{\theta}_n^\lambda) \right\rangle &\leq \frac{\lambda\beta^{-1} K_h \mathbf{K}_{3,d} \left(1 + |\tilde{\theta}_n^\lambda|^{\rho+q}\right) \left(1 + |\tilde{\theta}_n^\lambda|\right)^{\rho+q-2}}{\left(1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}\right)^{2/3}} \\
&\leq \frac{\lambda\beta^{-1} K_h \mathbf{K}_{3,d} 2^{\rho+q-3} \left(1 + |\tilde{\theta}_n^\lambda|^{\rho+q} + |\tilde{\theta}_n^\lambda|^{\rho+q-2} + |\tilde{\theta}_n^\lambda|^{2(\rho+q-1)}\right)}{\left(1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}\right)^{2/3}} \\
&\leq \frac{3\lambda\beta^{-1} K_h \mathbf{K}_{3,d} 2^{\rho+q-2} \left(1 + |\tilde{\theta}_n^\lambda|^{2(\rho+q-1)}\right)}{1 + \lambda |\tilde{\theta}_n^\lambda|^{2(\rho+q-1)}} \\
&\leq \beta^{-1} K_h \mathbf{K}_{3,d} 2^{\rho+q}, \tag{68}
\end{aligned}$$

$$\begin{aligned}
&-(\lambda^2/2)\beta^{-1} \left\langle H_\lambda(\tilde{\theta}_n^\lambda) h_\lambda(\tilde{\theta}_n^\lambda), \Upsilon_\lambda(\tilde{\theta}_n^\lambda) \right\rangle \\
&\leq \frac{(\lambda^2/2)\beta^{-1} K_H K_h \mathbf{K}_{3,d} \left(1 + |\tilde{\theta}_n^\lambda|^{\rho+q-1}\right) \left(1 + |\tilde{\theta}_n^\lambda|^{\rho+q}\right) \left(1 + |\tilde{\theta}_n^\lambda|\right)^{\rho+q-2}}{1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}}
\end{aligned}$$

$$\begin{aligned}
&\leq \frac{\lambda^2 \beta^{-1} K_H K_h \mathbf{K}_{3,d} 2^{\rho+q-4} \left(1 + |\tilde{\theta}_n^\lambda|^{\rho+q-1} + |\tilde{\theta}_n^\lambda|^{\rho+q} + |\tilde{\theta}_n^\lambda|^{2(\rho+q)-1}\right) \left(1 + |\tilde{\theta}_n^\lambda|^{\rho+q-2}\right)}{1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}} \\
&\leq \frac{7\lambda^2 \beta^{-1} K_H K_h \mathbf{K}_{3,d} 2^{\rho+q-4} \left(1 + |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}\right)}{1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}} \leq \beta^{-1} K_H K_h \mathbf{K}_{3,d} 2^{\rho+q-1}. \tag{69}
\end{aligned}$$

Substituting (64) – (69) into (63) yields

$$\begin{aligned}
|\phi^\lambda(\tilde{\theta}_n^\lambda)|^2 &\leq \frac{2K_h^2 |\tilde{\theta}_n^\lambda|^{2(\rho+q)}}{\left(1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}\right)^{2/3}} + 2K_h^2 + \frac{3\lambda^2 K_H^2 K_h^2 |\tilde{\theta}_n^\lambda|^{4(\rho+q)-2}}{\left(1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}\right)^{4/3}} + 3K_H^2 K_h^2 \\
&\quad + \frac{6\lambda K_H^2 K_h |\tilde{\theta}_n^\lambda|^{3(\rho+q)-1}}{1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}} + 6K_h^2 K_H + 2^{2(\rho+q)-6} \beta^{-2} \mathbf{K}_{3,d}^2 + \beta^{-1} K_h \mathbf{K}_{3,d} 2^{\rho+q} \\
&\quad + \beta^{-1} K_H K_h \mathbf{K}_{3,d} 2^{\rho+q-1}. \tag{70}
\end{aligned}$$

Combining the results in (62) and (70), we obtain the following upper bound for (61):

$$\begin{aligned}
|\Delta_{n,t}^\lambda|^2 &\leq |\tilde{\theta}_n^\lambda|^2 - \frac{2\lambda(t-n)\mathbf{a}_D |\tilde{\theta}_n^\lambda|^{\rho+q+1}}{\left(1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}\right)^{1/3}} + \frac{4\lambda^2(t-n)K_h K_H |\tilde{\theta}_n^\lambda|^{2(\rho+q)}}{\left(1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}\right)^{2/3}} \\
&\quad + \frac{\lambda^2(t-n)2K_h^2 |\tilde{\theta}_n^\lambda|^{2(\rho+q)}}{\left(1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}\right)^{2/3}} + \frac{\lambda^4(t-n)3K_H^2 K_h^2 |\tilde{\theta}_n^\lambda|^{4(\rho+q)-2}}{\left(1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}\right)^{4/3}} \\
&\quad + \frac{\lambda^3(t-n)6K_h^2 K_H |\tilde{\theta}_n^\lambda|^{3(\rho+q)-1}}{1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}} + \lambda(t-n) \left(2\mathbf{b}_D + 4K_h K_H \right. \\
&\quad \left. + \beta^{-1} 2^{\rho+q-1} \mathbf{K}_{3,d} + 2K_h^2 + 3K_H^2 K_h^2 + 6K_h^2 K_H \right. \\
&\quad \left. + 2^{2(\rho+q)-6} \beta^{-2} \mathbf{K}_{3,d}^2 + \beta^{-1} K_h \mathbf{K}_{3,d} 2^{\rho+q} + \beta^{-1} K_H K_h \mathbf{K}_{3,d} 2^{\rho+q-1}\right) \\
&= \left(1 - \frac{\lambda(t-n)\mathbf{a}_D |\tilde{\theta}_n^\lambda|^{\rho+q-1}}{\left(1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}\right)^{1/3}}\right) |\tilde{\theta}_n^\lambda|^2 - \lambda(t-n) \mathfrak{J}_1^\lambda(\tilde{\theta}_n^\lambda) + \lambda(t-n) \mathbf{c}_1, \tag{71}
\end{aligned}$$

where for any $\theta \in \mathbb{R}^d$

$$\begin{aligned}
\mathfrak{J}_1^\lambda(\theta) &:= \frac{\mathbf{a}_D |\theta|^{\rho+q+1}}{\left(1 + \lambda^{3/2} |\theta|^{3(\rho+q-1)}\right)^{1/3}} - \frac{\lambda(4K_h K_H + 2K_h^2) |\theta|^{2(\rho+q)}}{\left(1 + \lambda^{3/2} |\theta|^{3(\rho+q-1)}\right)^{2/3}} \\
&\quad - \frac{\lambda^3 3K_H^2 K_h^2 |\theta|^{4(\rho+q)-2}}{\left(1 + \lambda^{3/2} |\theta|^{3(\rho+q-1)}\right)^{4/3}} - \frac{\lambda^2 6K_h^2 K_H |\tilde{\theta}_n^\lambda|^{3(\rho+q)-1}}{1 + \lambda^{3/2} |\tilde{\theta}_n^\lambda|^{3(\rho+q-1)}},
\end{aligned}$$

and where

$$\begin{aligned}
\mathbf{c}_1 &:= \left(2\mathbf{b}_D + 4K_h K_H + \beta^{-1} 2^{\rho+q-1} \mathbf{K}_{3,d} + 2K_h^2 + 3K_H^2 K_h^2 + 6K_h^2 K_H \right. \\
&\quad \left. + 2^{2(\rho+q)-6} \beta^{-2} \mathbf{K}_{3,d}^2 + \beta^{-1} K_h \mathbf{K}_{3,d} 2^{\rho+q} + \beta^{-1} K_H K_h \mathbf{K}_{3,d} 2^{\rho+q-1}\right).
\end{aligned}$$

We note that, for $0 < \lambda \leq \lambda_{\max} \leq \min\{(19\mathbf{a}_D/240K_h K_H)^2, (19\mathbf{a}_D/240K_h^2)^2, (\mathbf{a}_D/120K_h^2 K_H^2)^{2/3}, \mathbf{a}_D/(480K_h^2 K_H)\}$, and for all $\theta \in \mathbb{R}^d$,

$$\mathfrak{J}_1^\lambda(\theta) \geq 0. \tag{72}$$

Indeed, by using the expression of \mathfrak{J}_1^λ , we have that, for all $\theta \in \mathbb{R}^d$,

$$\begin{aligned}
\mathfrak{J}_1^\lambda(\theta) &= \frac{(38\mathbf{a}_D/60 + 19\mathbf{a}_D/60) |\theta|^{\rho+q+1} \left(1 + \lambda^{3/2} |\theta|^{3(\rho+q-1)}\right)^{1/3} - \lambda(4K_h K_H + 2K_h^2) |\theta|^{2(\rho+q)}}{\left(1 + \lambda^{3/2} |\theta|^{3(\rho+q-1)}\right)^{2/3}} \\
&\quad + \frac{(\mathbf{a}_D/40) |\theta|^{\rho+q+1} \left(1 + \lambda^{3/2} |\theta|^{3(\rho+q-1)}\right) - \lambda^3 3K_H^2 K_h^2 |\theta|^{4(\rho+q)-2}}{\left(1 + \lambda^{3/2} |\theta|^{3(\rho+q-1)}\right)^{4/3}}
\end{aligned}$$

$$\begin{aligned}
& + \frac{(\mathbf{a}_D/40)|\theta|^{\rho+q+1} (1 + \lambda^{3/2}|\theta|^{3(\rho+q-1)})^{2/3} - \lambda^2 6K_h^2 K_H |\theta|^{3(\rho+q)-1}}{1 + \lambda^{3/2}|\theta|^{3(\rho+q-1)}} \\
& \geq \frac{(19\mathbf{a}_D/60 + 19\mathbf{a}_D/120)\lambda^{1/2}|\theta|^{2(\rho+q)} - \lambda(4K_h K_H + 2K_h^2)|\theta|^{2(\rho+q)}}{(1 + \lambda^{3/2}|\theta|^{3(\rho+q-1)})^{2/3}} \\
& + \frac{(\mathbf{a}_D/40)\lambda^{3/2}|\theta|^{4(\rho+q)-2} - \lambda^3 3K_H^2 K_h^2 |\theta|^{4(\rho+q)-2}}{(1 + \lambda^{3/2}|\theta|^{3(\rho+q-1)})^{4/3}} \\
& + \frac{(\mathbf{a}_D/80)\lambda|\theta|^{3(\rho+q)-1} - \lambda^2 6K_h^2 K_H |\theta|^{3(\rho+q)-1}}{1 + \lambda^{3/2}|\theta|^{3(\rho+q-1)}} \\
& \geq 0,
\end{aligned}$$

where the last inequality holds due to $0 < \lambda \leq \lambda_{\max}$. Substituting (72) into (71) yields

$$|\Delta_{n,t}^\lambda|^2 \leq \left(1 - \frac{\lambda(t-n)\mathbf{a}_D |\tilde{\theta}_n^\lambda|^{\rho+q-1}}{(1 + \lambda^{3/2}|\tilde{\theta}_n^\lambda|^{3(\rho+q-1)})^{1/3}} \right) |\tilde{\theta}_n^\lambda|^2 + \lambda(t-n)\mathbf{c}_1.$$

In addition, since $f(s) = s/(1 + \lambda^{3/2}s^3)^{1/3}$ is non-decreasing for all $s \geq 0$, we obtain that, for all $|\theta| > M_1 > 0$,

$$\frac{|\theta|^{\rho+q-1}}{(1 + \lambda^{3/2}|\theta|^{3(\rho+q-1)})^{1/3}} \geq \frac{M_1^{\rho+q-1}}{(1 + \lambda^{3/2}M_1^{3(\rho+q-1)})^{1/3}} \geq \frac{M_1^{\rho+q-1}}{(1 + M_1^{3(\rho+q-1)})^{1/3}} =: \kappa.$$

Denote by $S_{n,M_1} := \{\omega \in \Omega : |\tilde{\theta}_n^\lambda(\omega)| > M_1\}$. The above inequality further implies that,

$$|\Delta_{n,t}^\lambda|^2 \mathbb{1}_{S_{n,M_1}} \leq (1 - \lambda(t-n)\mathbf{a}_D\kappa) |\tilde{\theta}_n^\lambda|^2 \mathbb{1}_{S_{n,M_1}} + \lambda(t-n)\mathbf{c}_1 \mathbb{1}_{S_{n,M_1}}.$$

Similarly, we have that

$$|\Delta_{n,t}^\lambda|^2 \mathbb{1}_{S_{n,M_1}^c} \leq (1 - \lambda(t-n)\mathbf{a}_D\kappa) |\tilde{\theta}_n^\lambda|^2 \mathbb{1}_{S_{n,M_1}^c} + \lambda(t-n)(\mathbf{a}_D\kappa M_1^2 + \mathbf{c}_1) \mathbb{1}_{S_{n,M_1}^c}.$$

Combining the two cases yields

$$|\Delta_{n,t}^\lambda|^2 \leq (1 - \lambda(t-n)\mathbf{a}_D\kappa) |\tilde{\theta}_n^\lambda|^2 + \lambda(t-n)(\mathbf{a}_D\kappa M_1^2 + \mathbf{c}_1). \quad (73)$$

Finally, by substituting (60) and (73) into (57), we obtain

$$\mathbb{E} \left[|\tilde{\theta}_t^\lambda|^2 \mid \tilde{\theta}_n^\lambda \right] \leq (1 - \lambda(t-n)\mathbf{a}_D\kappa) |\tilde{\theta}_n^\lambda|^2 + \lambda(t-n)\mathbf{c}_0, \quad (74)$$

where

$$\begin{aligned}
\kappa & := M_1^{\rho+q-1} / \left(1 + M_1^{3(\rho+q-1)} \right)^{1/3} \quad (\text{with } M_1 > 0), \\
\mathbf{c}_0 & := \mathbf{a}_D\kappa M_1^2 + \mathbf{c}_1 + 2^{2(\rho+q)}\beta^{-1}d(1 + \mathbf{K}_1)^2, \\
\mathbf{c}_1 & := \left(2\mathbf{b}_D + 4K_h K_H + \beta^{-1}2^{\rho+q-1}\mathbf{K}_{3,d} + 2K_h^2 + 3K_H^2 K_h^2 + 6K_h^2 K_H \right. \\
& \quad \left. + 2^{2(\rho+q)-6}\beta^{-2}\mathbf{K}_{3,d}^2 + \beta^{-1}K_h \mathbf{K}_{3,d} 2^{\rho+q} + \beta^{-1}K_H K_h \mathbf{K}_{3,d} 2^{\rho+q-1} \right).
\end{aligned} \quad (75)$$

We observe that, for $0 < \lambda \leq \lambda_{\max} \leq 1/\mathbf{a}_D$,

$$1 > 1 - \lambda(t-n)\mathbf{a}_D\kappa > 1 - \lambda\mathbf{a}_D \geq 0,$$

then, by induction, (74) implies, for $t \in (n, n+1]$, $n \in \mathbb{N}_0$, $0 < \lambda \leq \lambda_{\max} \leq 1$, that,

$$\begin{aligned}
\mathbb{E} \left[|\tilde{\theta}_t^\lambda|^2 \right] & \leq (1 - \lambda(t-n)\mathbf{a}_D\kappa) \mathbb{E} \left[|\tilde{\theta}_n^\lambda|^2 \right] + \lambda(t-n)\mathbf{c}_0 \\
& \leq (1 - \lambda(t-n)\mathbf{a}_D\kappa) (1 - \lambda\mathbf{a}_D\kappa) \mathbb{E} \left[|\tilde{\theta}_{n-1}^\lambda|^2 \right] + \mathbf{c}_0 + \lambda\mathbf{c}_0 \\
& \leq (1 - \lambda(t-n)\mathbf{a}_D\kappa) (1 - \lambda\mathbf{a}_D\kappa)^2 \mathbb{E} \left[|\tilde{\theta}_{n-2}^\lambda|^2 \right] + \mathbf{c}_0 + \lambda\mathbf{c}_0 (1 + (1 - \lambda\mathbf{a}_D\kappa)) \\
& \leq \dots \\
& \leq (1 - \lambda(t-n)\mathbf{a}_D\kappa) (1 - \lambda\mathbf{a}_D\kappa)^n \mathbb{E} \left[|\theta_0|^2 \right] + \mathbf{c}_0 (1 + 1/(\mathbf{a}_D\kappa)),
\end{aligned}$$

which completes the proof. \square

Proof of Lemma 6.3-(ii). For any $p \in [2, \infty) \cap \mathbb{N}$, $0 < \lambda \leq \lambda_{\max} \leq 1$ with λ_{\max} given in (14), $t \in (n, n+1]$, $n \in \mathbb{N}_0$, by using the same arguments as in the proof of [27, Lemma 4.2-(ii)] up to the inequality before [27, Eq. (134)] and by using (28) with (54), we obtain that

$$\begin{aligned} \mathbb{E} \left[|\tilde{\theta}_t^\lambda|^{2p} \left| \tilde{\theta}_n^\lambda \right. \right] &\leq |\Delta_{n,t}^\lambda|^{2p} + 2^{2p-3} p(2p-1) |\Delta_{n,t}^\lambda|^{2p-2} \mathbb{E} \left[|\Xi_{n,t}^\lambda|^2 \left| \tilde{\theta}_n^\lambda \right. \right] \\ &\quad + 2^{2p-3} p(2p-1) \mathbb{E} \left[|\Xi_{n,t}^\lambda|^{2p} \left| \tilde{\theta}_n^\lambda \right. \right]. \end{aligned} \quad (76)$$

Then, by using (54) and (56), we can obtain an upper estimate for the last term in (76) as follows:

$$\begin{aligned} \mathbb{E} \left[|\Xi_{n,t}^\lambda|^{2p} \left| \tilde{\theta}_n^\lambda \right. \right] &= (2\lambda\beta^{-1})^p \mathbb{E} \left[\left\langle (B_t^\lambda - B_n^\lambda), \left(\psi^\lambda(\tilde{\theta}_n^\lambda) \right)^2 (B_t^\lambda - B_n^\lambda) \right\rangle^p \left| \tilde{\theta}_n^\lambda \right. \right] \\ &\leq (2\lambda\beta^{-1})^p \left| \text{Id} - \lambda H_\lambda(\tilde{\theta}_n^\lambda) + (\lambda^2/3)(H_\lambda(\tilde{\theta}_n^\lambda))^2 \right|^p \mathbb{E} \left[|B_t^\lambda - B_n^\lambda|^{2p} \right] \\ &\leq (2\lambda\beta^{-1} dp(2p-1)(t-n))^{p-1} 3^{p-1} \left(1 + \lambda^p |H_\lambda(\tilde{\theta}_n^\lambda)|_{\mathbb{F}}^p + (\lambda^2/3)^p |H_\lambda(\tilde{\theta}_n^\lambda)|_{\mathbb{F}}^{2p} \right) \\ &\leq \lambda(t-n)(2\beta^{-1} dp(2p-1))^{p-1} 3^{p-1} \left(1 + \lambda^p |H_\lambda(\tilde{\theta}_n^\lambda)|_{\mathbb{F}}^p \right)^2 \\ &\leq \lambda(t-n)(2\beta^{-1} dp(2p-1))^{p-1} 3^{p-1} \left(1 + 2^{p(\rho+q-1)} d^{p/2} \mathbf{K}_1^p \right)^2, \end{aligned} \quad (77)$$

where the last inequality holds due to (59). Substituting (60) and (77) into (76) yields

$$\begin{aligned} \mathbb{E} \left[|\tilde{\theta}_t^\lambda|^{2p} \left| \tilde{\theta}_n^\lambda \right. \right] &\leq |\Delta_{n,t}^\lambda|^{2p} + \lambda(t-n) 2^{2(p+\rho+q)-3} p(2p-1) \beta^{-1} d(1 + \mathbf{K}_1)^2 |\Delta_{n,t}^\lambda|^{2p-2} \\ &\quad + \lambda(t-n) c_\Xi(p), \end{aligned} \quad (78)$$

where $c_\Xi(p) := 2^{2p-4} (\beta^{-1} d)^p (2p(2p-1))^{p-1} 3^{p-1} \left(1 + 2^{p(\rho+q-1)} d^{p/2} \mathbf{K}_1^p \right)^2$. Next, we apply (73) to obtain

$$\begin{aligned} |\Delta_{n,t}^\lambda|^{2p} &\leq \left((1 - \lambda(t-n) \mathbf{a}_D \kappa) |\tilde{\theta}_n^\lambda|^2 + \lambda(t-n) (\mathbf{a}_D \kappa \mathbf{M}_1^2 + c_1) \right)^p \\ &\leq (1 + \lambda(t-n) \mathbf{a}_D \kappa / 2)^{p-1} (1 - \lambda(t-n) \mathbf{a}_D \kappa)^p |\tilde{\theta}_n^\lambda|^{2p} \\ &\quad + (1 + 2/(\lambda(t-n) \mathbf{a}_D \kappa))^{p-1} (\lambda(t-n))^p (\mathbf{a}_D \kappa \mathbf{M}_1^2 + c_1)^p \\ &\leq (1 - \lambda(t-n) \mathbf{a}_D \kappa / 2)^{p-1} (1 - \lambda(t-n) \mathbf{a}_D \kappa) |\tilde{\theta}_n^\lambda|^{2p} \\ &\quad + \lambda(t-n) (1 + 2/(\mathbf{a}_D \kappa))^{p-1} (\mathbf{a}_D \kappa \mathbf{M}_1^2 + c_1)^p \\ &= \bar{c}_{n,t}^\lambda(p) |\tilde{\theta}_n^\lambda|^{2p} + \tilde{c}_{n,t}^\lambda(p), \end{aligned} \quad (79)$$

where the second inequality holds due to $(u+v)^p \leq (1+\varepsilon)^{p-1} u^p + (1+\varepsilon^{-1})^{p-1} v^p$, $u, v \geq 0$, $\varepsilon > 0$ with $\varepsilon = \lambda(t-n) \mathbf{a}_D \kappa / 2$, and where

$$\begin{aligned} \bar{c}_{n,t}^\lambda(p) &:= (1 - \lambda(t-n) \mathbf{a}_D \kappa / 2)^{p-1} (1 - \lambda(t-n) \mathbf{a}_D \kappa), \\ \tilde{c}_{n,t}^\lambda(p) &:= \lambda(t-n) (1 + 2/(\mathbf{a}_D \kappa))^{p-1} (\mathbf{a}_D \kappa \mathbf{M}_1^2 + c_1)^p. \end{aligned}$$

In addition, we observe that by (79),

$$|\Delta_{n,t}^\lambda|^{2p-2} \leq \bar{c}_{n,t}^\lambda(p-1) |\tilde{\theta}_n^\lambda|^{2p-2} + \tilde{c}_{n,t}^\lambda(p-1), \quad (80)$$

and, in particular, when $p = 2$, (80) yields $|\Delta_{n,t}^\lambda|^2 \leq \bar{c}_{n,t}^\lambda(1) |\tilde{\theta}_n^\lambda|^2 + \tilde{c}_{n,t}^\lambda(1)$ which is exactly the upper bound (73). By substituting (79) and (80) into (78), we have that

$$\begin{aligned} \mathbb{E} \left[|\tilde{\theta}_t^\lambda|^{2p} \left| \tilde{\theta}_n^\lambda \right. \right] &\leq \bar{c}_{n,t}^\lambda(p) |\tilde{\theta}_n^\lambda|^{2p} + \tilde{c}_{n,t}^\lambda(p) + \lambda(t-n) c_\Xi(p) \\ &\quad + \lambda(t-n) 2^{2(p+\rho+q)-3} p(2p-1) \beta^{-1} d(1 + \mathbf{K}_1)^2 \\ &\quad \times \left(\bar{c}_{n,t}^\lambda(p-1) |\tilde{\theta}_n^\lambda|^{2p-2} + \tilde{c}_{n,t}^\lambda(p-1) \right). \end{aligned} \quad (81)$$

Denote by $\mathbf{M}_2(p) := (2^{2(p+\rho+q)-1} p(2p-1) \beta^{-1} d(1 + \mathbf{K}_1)^2 / (\mathbf{a}_D \kappa))^{1/2}$. For all $|\theta| > \mathbf{M}_2(p)$, we have that

$$(\lambda(t-n) \mathbf{a}_D \kappa / 4) |\theta|^{2p} > (\lambda(t-n) 2^{2(p+\rho+q)-1} p(2p-1) \beta^{-1} d(1 + \mathbf{K}_1)^2 / 4) |\theta|^{2p-2}.$$

Denote by $S_{n, M_2(p)} := \{\omega \in \Omega : |\tilde{\theta}_n^\lambda(\omega)| > M_2(p)\}$. By using the above inequality, (81) can be further bounded as follows:

$$\begin{aligned}
& \mathbb{E} \left[|\tilde{\theta}_t^\lambda|^{2p} \mathbb{1}_{S_{n, M_2(p)}} \left| \tilde{\theta}_n^\lambda \right| \right] \\
& \leq (1 - \lambda(t - n) a_D \kappa / 4) \bar{c}_{n, t}^\lambda (p - 1) |\tilde{\theta}_n^\lambda|^{2p} \mathbb{1}_{S_{n, M_2(p)}} + \left(\bar{c}_{n, t}^\lambda (p) + \lambda(t - n) c_\Xi(p) \right) \mathbb{1}_{S_{n, M_2(p)}} \\
& \quad + \lambda(t - n) 2^{2(p + \rho + q) - 3} p (2p - 1) \beta^{-1} d (1 + K_1)^2 \bar{c}_{n, t}^\lambda (p - 1) \mathbb{1}_{S_{n, M_2(p)}} \\
& \quad - (\lambda(t - n) a_D \kappa / 4) \bar{c}_{n, t}^\lambda (p - 1) |\tilde{\theta}_n^\lambda|^{2p} \mathbb{1}_{S_{n, M_2(p)}} \\
& \quad + (\lambda(t - n) 2^{2(p + \rho + q) - 1} p (2p - 1) \beta^{-1} d (1 + K_1)^2 / 4) \bar{c}_{n, t}^\lambda (p - 1) |\tilde{\theta}_n^\lambda|^{2p - 2} \mathbb{1}_{S_{n, M_2(p)}} \\
& \leq (1 - \lambda(t - n) a_D \kappa) |\tilde{\theta}_n^\lambda|^{2p} \mathbb{1}_{S_{n, M_2(p)}} + \lambda(t - n) \left[(1 + 2 / (a_D \kappa))^{p - 1} (a_D \kappa M_1^2 + c_1)^p + c_\Xi(p) \right. \\
& \quad \left. + 2^{2(p + \rho + q) - 3} p (2p - 1) \beta^{-1} d (1 + K_1)^2 \right. \\
& \quad \left. \times \left((1 + 2 / (a_D \kappa))^{p - 2} (a_D \kappa M_1^2 + c_1)^{p - 1} + M_2(p)^{2p - 2} \right) \right] \mathbb{1}_{S_{n, M_2(p)}} \\
& \leq (1 - \lambda(t - n) a_D \kappa) |\tilde{\theta}_n^\lambda|^{2p} \mathbb{1}_{S_{n, M_2(p)}} + \lambda(t - n) c_p \mathbb{1}_{S_{n, M_2(p)}}, \tag{82}
\end{aligned}$$

where

$$\begin{aligned}
c_p & := (1 + 2 / (a_D \kappa))^{p - 1} (a_D \kappa M_1^2 + c_1)^p + c_\Xi(p) + 2^{2(p + \rho + q) - 3} p (2p - 1) \beta^{-1} d (1 + K_1)^2 \\
& \quad \times \left((1 + 2 / (a_D \kappa))^{p - 2} (a_D \kappa M_1^2 + c_1)^{p - 1} + M_2(p)^{2p - 2} \right), \tag{83} \\
c_\Xi(p) & := 2^{2p - 4} (\beta^{-1} d)^p (2p (2p - 1))^{p + 1} 3^{p - 1} \left(1 + 2^{p(\rho + q - 1)} d^{p/2} K_1^p \right)^2, \\
M_2(p) & := (2^{2(p + \rho + q) - 1} p (2p - 1) \beta^{-1} d (1 + K_1)^2 / (a_D \kappa))^{1/2}
\end{aligned}$$

with κ and c_1 given in (75). Similarly, by using (81), we have that

$$\mathbb{E} \left[|\tilde{\theta}_t^\lambda|^{2p} \mathbb{1}_{S_{n, M_2(p)}^c} \left| \tilde{\theta}_n^\lambda \right| \right] \leq (1 - \lambda(t - n) a_D \kappa) |\tilde{\theta}_n^\lambda|^{2p} \mathbb{1}_{S_{n, M_2(p)}^c} + \lambda(t - n) c_p \mathbb{1}_{S_{n, M_2(p)}^c}. \tag{84}$$

Combing (82) and (84) yields the desired result. \square

Lemma A.2. *Let Assumptions 1, 2, and 3 hold. Then, for any $0 < \lambda \leq \lambda_{\max}$, $p > 0$, $t \geq 0$, we obtain*

$$\mathbb{E} \left[|\bar{\theta}_t^\lambda - \bar{\theta}_{[t]}^\lambda|^{2p} \right] \leq \lambda^p \left(e^{-\lambda a_D \kappa [t]} \bar{C}_{A0, p} \mathbb{E}[|\theta_0|^{4[p](\rho + 1)}] + \tilde{C}_{A0, p} \right), \tag{85}$$

$$\mathbb{E} \left[|\bar{\zeta}_t^{\lambda, n} - \bar{\zeta}_{[t]}^{\lambda, n}|^{2p} \right] \leq \lambda^p \left(e^{-\lambda a_D \min\{\kappa, 1/2\} [t]} \bar{C}_{A1, p} \mathbb{E}[|\theta_0|^{2[p](\rho + 1)}] + \tilde{C}_{A1, p} \right), \tag{86}$$

where $\bar{C}_{A0, p}$ and $\tilde{C}_{A0, p}$ are given in (87), and $\bar{C}_{A1, p}$ and $\tilde{C}_{A1, p}$ are given in (88).

Proof. To show that (85) holds, we use the definition of $(\bar{\theta}_t^\lambda)_{t \geq 0}$ given in (29) and obtain that, for any $t \geq 0$,

$$\begin{aligned}
& \mathbb{E} \left[|\bar{\theta}_t^\lambda - \bar{\theta}_{[t]}^\lambda|^{2p} \right] \\
& = \mathbb{E} \left[\left| -\lambda \int_{[t]}^t h_\lambda(\bar{\theta}_{[s]}^\lambda) ds + \lambda^2 \int_{[t]}^t \int_{[s]}^s \left(H_\lambda(\bar{\theta}_{[r]}^\lambda) h_\lambda(\bar{\theta}_{[r]}^\lambda) - \beta^{-1} \Upsilon_\lambda(\bar{\theta}_{[r]}^\lambda) \right) dr ds \right. \right. \\
& \quad \left. \left. - \lambda \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t \int_{[s]}^s H_\lambda(\bar{\theta}_{[r]}^\lambda) dB_r^\lambda ds + \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t dB_s^\lambda \right|^{2p} \right] \\
& \leq 5^{2p} \left(\lambda^{2p} \mathbb{E} \left[|h_\lambda(\bar{\theta}_{[t]}^\lambda)|^{2p} \right] + \lambda^{4p} \mathbb{E} \left[|H_\lambda(\bar{\theta}_{[r]}^\lambda)|^{2p} |h_\lambda(\bar{\theta}_{[t]}^\lambda)|^{2p} \right] \right. \\
& \quad \left. + \lambda^{4p} \beta^{-2p} \mathbb{E} \left[|\Upsilon_\lambda(\bar{\theta}_{[t]}^\lambda)|^{2p} \right] + \lambda^p (2\beta^{-1})^p \mathbb{E} \left[\left| \int_{[t]}^t dB_s^\lambda \right|^{2p} \right] \right)
\end{aligned}$$

$$\begin{aligned}
& + \lambda^{3p}(2\beta^{-1})^p \mathbb{E} \left[\left| H_\lambda(\bar{\theta}_{[t]}^\lambda) \right|^{2p} \left| \int_{[t]}^t \int_{[s]}^s dB_r^\lambda ds \right|^{2p} \right] \\
& \leq 5^{2p} \lambda^p \left(K_h^{2p} \mathbb{E} \left[(1 + |\bar{\theta}_{[t]}^\lambda|^{\rho+q})^{2p} \right] + \beta^{-2p} \mathbf{K}_{3,d}^{2p} \mathbb{E} \left[(1 + |\bar{\theta}_{[t]}^\lambda|)^{2p(\rho+q-2)} \right] \right. \\
& \quad + K_H^{2p} K_h^{2p} \mathbb{E} \left[(1 + |\bar{\theta}_{[t]}^\lambda|^{\rho+q-1})^{2p} (1 + |\bar{\theta}_{[t]}^\lambda|^{\rho+q})^{2p} \right] + (2\beta^{-1}(p+1)(d+2p))^p \\
& \quad \left. + (\beta^{-1}(4p+2)(d+4p))^p K_H^{2p} \left(\mathbb{E} \left[(1 + |\bar{\theta}_{[t]}^\lambda|^{\rho+q-1})^{4p} \right] \right)^{1/2} \right) \\
& \leq \lambda^p 5^{2p} \left(K_h^{2p} + \beta^{-2p} \mathbf{K}_{3,d}^{2p} + K_H^{2p} K_h^{2p} + (\beta^{-1}(4p+2)(d+4p))^p (1 + K_H^{2p}) \right) \\
& \quad \times 2^{4[p](\rho+1)} \left(\mathbb{E} \left[|\bar{\theta}_{[t]}^\lambda|^{4[p](\rho+1)} \right] + 1 \right) \\
& \leq \lambda^p \left(e^{-\lambda a_D \kappa [t]} \bar{\mathbf{C}}_{\mathbf{A}0,p} \mathbb{E}[|\theta_0|^{4[p](\rho+1)}] + \tilde{\mathbf{C}}_{\mathbf{A}0,p} \right),
\end{aligned}$$

where the first inequality holds due to $(\sum_{l=1}^v u_l)^w \leq v^w \sum_{l=1}^v u_l^w$, $v \in \mathbb{N}$, $u_l \geq 0$, $w > 0$, the second inequality holds due to Remark 2.2, Cauchy-Schwarz inequality, and the following inequality:

$$\mathbb{E} \left[\left| \int_{[t]}^t dB_s^\lambda \right|^{2p} \right] \leq \max\{d^p, (p(d+2p-2))^p\} \leq ((p+1)(d+2p))^p,$$

the fourth inequality holds due to Lemma 6.3, and where

$$\begin{aligned}
\bar{\mathbf{C}}_{\mathbf{A}0,p} & := 5^{2p} 2^{4[p](\rho+1)} \left(K_h^{2p} + \beta^{-2p} \mathbf{K}_{3,d}^{2p} + K_H^{2p} K_h^{2p} + (\beta^{-1}(4p+2)(d+4p))^p (1 + K_H^{2p}) \right), \\
\tilde{\mathbf{C}}_{\mathbf{A}0,p} & := \bar{\mathbf{C}}_{\mathbf{A}0,p} (c_{2[p](\rho+1)} (1 + 1/(a_D \kappa)) + 1).
\end{aligned} \tag{87}$$

The inequality (86) can be obtained by using similar arguments. More precisely, by using Definition 6.2 with (30), we obtain that, for any $t \geq 0$,

$$\begin{aligned}
\mathbb{E} \left[|\bar{\zeta}_t^{\lambda,n} - \bar{\zeta}_{[t]}^{\lambda,n}|^{2p} \right] & = \mathbb{E} \left[\left| -\lambda \int_{[t]}^t h(\bar{\zeta}_s^{\lambda,n}) ds + \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t dB_s^\lambda \right|^{2p} \right] \\
& \leq 2^{2p} \left(\lambda^{2p} \mathbb{E} \left[\int_{[t]}^t |h(\bar{\zeta}_s^{\lambda,n})|^{2p} ds \right] + \lambda^p (2\beta^{-1})^p \mathbb{E} \left[\left| \int_{[t]}^t dB_s^\lambda \right|^{2p} \right] \right) \\
& \leq 2^{2p} \lambda^p \left(K_h^{2p} \int_{[t]}^t \mathbb{E} \left[(1 + |\bar{\zeta}_s^{\lambda,n}|^{\rho+q})^{2p} \right] ds + (2\beta^{-1}(p+1)(d+2p))^p \right) \\
& \leq \lambda^p \left(2^{2p} K_h^{2p} \int_{[t]}^t \mathbb{E} \left[V_{2[p](\rho+1)}(\bar{\zeta}_s^{\lambda,n}) \right] ds + 2^{2p} (2\beta^{-1}(p+1)(d+2p))^p \right) \\
& \leq \lambda^p \left(e^{-\lambda a_D \min\{\kappa, 1/2\} [t]} \bar{\mathbf{C}}_{\mathbf{A}1,p} \mathbb{E}[|\theta_0|^{2[p](\rho+1)}] + \tilde{\mathbf{C}}_{\mathbf{A}1,p} \right),
\end{aligned}$$

where the last inequality holds due to Lemma 6.5 and where

$$\begin{aligned}
\bar{\mathbf{C}}_{\mathbf{A}1,p} & := 2^{2p+[p](\rho+1)-1} K_h^{2p}, \\
\tilde{\mathbf{C}}_{\mathbf{A}1,p} & := \bar{\mathbf{C}}_{\mathbf{A}1,p} (c_{[p](\rho+1)} (1 + 1/(a_D \kappa)) + 1) + 2^{2p} K_h^{2p} 3v_{2[p](\rho+1)} (M_V(2[p](\rho+1))) \\
& \quad + 2^{2p} (2\beta^{-1}(p+1)(d+2p))^p.
\end{aligned} \tag{88}$$

This completes the proof. \square

Lemma A.3. *Let Assumptions 1, 2, and 3 hold. Then, for any $0 < \lambda \leq \lambda_{\max}$, $n \in \mathbb{N}_0$, $t \geq nT$, we obtain the following inequalities:*

$$\mathbb{E} \left[\left| -\lambda \int_{[t]}^t \left(H(\bar{\theta}_s^\lambda) - H_\lambda(\bar{\theta}_{[s]}^\lambda) \right) h_\lambda(\bar{\theta}_{[s]}^\lambda) ds \right|^2 \right] \leq \lambda^3 \left(e^{-a_D \kappa n/2} \bar{\mathbf{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + \tilde{\mathbf{C}}_{\mathbf{A}2} \right),$$

$$\begin{aligned}
& \mathbb{E} \left[\left| \lambda^2 \int_{[t]}^t H(\bar{\theta}_s^\lambda) \int_{[s]}^s H_\lambda(\bar{\theta}_{[r]}^\lambda) h_\lambda(\bar{\theta}_{[r]}^\lambda) dr ds \right|^2 \right] \leq \lambda^3 \left(e^{-a_D \kappa n/2} \bar{\mathcal{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + \tilde{\mathcal{C}}_{\mathbf{A}2} \right), \\
& \mathbb{E} \left[\left| -\lambda^2 \beta^{-1} \int_{[t]}^t H(\bar{\theta}_s^\lambda) \int_{[s]}^s \Upsilon_\lambda(\bar{\theta}_{[r]}^\lambda) dr ds \right|^2 \right] \leq \lambda^3 \left(e^{-a_D \kappa n/2} \bar{\mathcal{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + \tilde{\mathcal{C}}_{\mathbf{A}2} \right), \\
& \mathbb{E} \left[\left| -\lambda \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t H(\bar{\theta}_s^\lambda) \int_{[s]}^s H_\lambda(\bar{\theta}_{[r]}^\lambda) dB_r^\lambda ds \right|^2 \right] \leq \lambda^3 \left(e^{-a_D \kappa n/2} \bar{\mathcal{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + \tilde{\mathcal{C}}_{\mathbf{A}2} \right), \\
& \mathbb{E} \left[\left| \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t \left(H(\bar{\theta}_s^\lambda) - H_\lambda(\bar{\theta}_{[s]}^\lambda) \right) dB_s^\lambda \right|^2 \right] \leq \lambda^2 \left(e^{-a_D \kappa n/2} \bar{\mathcal{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + \tilde{\mathcal{C}}_{\mathbf{A}2} \right), \\
& \mathbb{E} \left[\left| \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t \left(H(\bar{\zeta}_s^{\lambda,n}) - H(\bar{\zeta}_{[s]}^{\lambda,n}) \right) dB_s^\lambda \right|^2 \right] \\
& \qquad \qquad \qquad \leq \lambda^2 \left(e^{-a_D \min\{\kappa, 1/2\} n/2} \bar{\mathcal{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + \tilde{\mathcal{C}}_{\mathbf{A}2} \right), \\
& \mathbb{E} \left[\left| \lambda \beta^{-1} \int_{[t]}^t \left(\Upsilon(\bar{\theta}_s^\lambda) - \Upsilon_\lambda(\bar{\theta}_{[s]}^\lambda) \right) ds \right|^2 \right] \leq \lambda^{2+q} \left(e^{-a_D \kappa n/2} \bar{\mathcal{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + \tilde{\mathcal{C}}_{\mathbf{A}2} \right), \\
& \mathbb{E} \left[\left| -\lambda \int_{[t]}^t H(\bar{\theta}_s^\lambda) h_\lambda(\bar{\theta}_{[s]}^\lambda) ds \right|^2 + \left| \lambda \beta^{-1} \int_{[t]}^t \Upsilon(\bar{\theta}_s^\lambda) ds \right|^2 \right] \\
& \qquad \qquad \qquad \leq \lambda^2 \left(e^{-a_D \kappa n/2} \bar{\mathcal{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + \tilde{\mathcal{C}}_{\mathbf{A}2} \right), \\
& \mathbb{E} \left[\left| -\lambda \int_{[t]}^t H(\bar{\zeta}_s^{\lambda,n}) h(\bar{\zeta}_s^{\lambda,n}) ds \right|^2 + \left| \lambda \beta^{-1} \int_{[t]}^t \Upsilon(\bar{\zeta}_s^{\lambda,n}) ds \right|^2 \right] \\
& \qquad \qquad \qquad \leq \lambda^2 \left(e^{-a_D \min\{\kappa, 1/2\} n/2} \bar{\mathcal{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + \tilde{\mathcal{C}}_{\mathbf{A}2} \right),
\end{aligned}$$

where $\bar{\mathcal{C}}_{\mathbf{A}2}$ and $\tilde{\mathcal{C}}_{\mathbf{A}2}$ are given in (90).

Proof. We note that, by using (4), for any $f : \mathbb{R}^d \rightarrow \mathbb{R}^{i \times j}$, $i, j \in \mathbb{N}$, $\theta \in \mathbb{R}^d$,

$$|f(\theta) - f_\lambda(\theta)| = \left| \frac{((1 + \lambda^{3/2} |\theta|^{3(\rho+q-1)})^{1/3} - 1) f(\theta)}{(1 + \lambda^{3/2} |\theta|^{3(\rho+q-1)})^{1/3}} \right| \leq \lambda^{3/2} |\theta|^{3(\rho+q-1)} |f(\theta)|. \quad (89)$$

The inequalities can be obtained by using the following arguments:

(i) To show that the first inequality holds, by using Remark 2.2 and (89), we have that

$$\begin{aligned}
& \mathbb{E} \left[\left| -\lambda \int_{[t]}^t \left(H(\bar{\theta}_s^\lambda) - H_\lambda(\bar{\theta}_{[s]}^\lambda) \right) h_\lambda(\bar{\theta}_{[s]}^\lambda) ds \right|^2 \right] \\
& \leq \lambda^2 \int_{[t]}^t \mathbb{E} \left[|H(\bar{\theta}_s^\lambda) - H_\lambda(\bar{\theta}_{[s]}^\lambda)|^2 |h_\lambda(\bar{\theta}_{[s]}^\lambda)|^2 \right] ds \\
& \leq 2\lambda^2 \int_{[t]}^t \mathbb{E} \left[|H(\bar{\theta}_s^\lambda) - H(\bar{\theta}_{[s]}^\lambda)|_{\mathbb{F}}^2 |h_\lambda(\bar{\theta}_{[s]}^\lambda)|^2 \right] ds \\
& \quad + 2\lambda^2 \int_{[t]}^t \mathbb{E} \left[|H(\bar{\theta}_{[s]}^\lambda) - H_\lambda(\bar{\theta}_{[s]}^\lambda)|^2 |h_\lambda(\bar{\theta}_{[s]}^\lambda)|^2 \right] ds \\
& \leq 2\lambda^2 \int_{[t]}^t \mathbb{E} \left[dK_0^2 (1 + |\bar{\theta}_s^\lambda| + |\bar{\theta}_{[s]}^\lambda|)^{2(\rho+q-2)} |\bar{\theta}_s^\lambda - \bar{\theta}_{[s]}^\lambda|^2 K_h^2 (1 + |\bar{\theta}_{[s]}^\lambda|^{\rho+q})^2 \right] ds
\end{aligned}$$

$$\begin{aligned}
& + 2\lambda^2 \int_{[t]}^t \mathbb{E} \left[\lambda^3 |\bar{\theta}_{[s]}^\lambda|^{6(\rho+q-1)} K_H^2 (1 + |\bar{\theta}_{[s]}^\lambda|^{\rho+q-1})^2 K_h^2 (1 + |\bar{\theta}_{[s]}^\lambda|^{\rho+q})^2 \right] ds \\
& \leq 2\lambda^2 dK_0^2 K_h^2 \int_{[t]}^t 3^{8(\rho+1)} \left(\mathbb{E} \left[1 + |\bar{\theta}_s^\lambda|^{16(\rho+1)} + |\bar{\theta}_{[s]}^\lambda|^{16(\rho+1)} \right] \right)^{1/2} \\
& \quad \times \left(\mathbb{E} \left[|\bar{\theta}_s^\lambda - \bar{\theta}_{[s]}^\lambda|^4 \right] \right)^{1/2} ds \\
& \quad + \lambda^5 K_H^2 K_h^2 \int_{[t]}^t 2^{16(\rho+1)} \mathbb{E} \left[1 + |\bar{\theta}_{[s]}^\lambda|^{16(\rho+1)} \right] ds \\
& \leq \lambda^2 3^{8(\rho+1)+1} dK_0^2 K_h^2 \left(2e^{-\lambda a_D \kappa [t]} \mathbb{E} \left[|\theta_0|^{16(\rho+1)} \right] + 2c_{8(\rho+1)} (1 + 1/(a_D \kappa)) + 1 \right)^{1/2} \\
& \quad \times \lambda \left(e^{-\lambda a_D \kappa [t]} \bar{C}_{\mathbf{A}0,2} \mathbb{E} [|\theta_0|^{8(\rho+1)}] + \tilde{C}_{\mathbf{A}0,2} \right)^{1/2} \\
& \quad + \lambda^5 2^{16(\rho+1)} K_H^2 K_h^2 \left(e^{-\lambda a_D \kappa [t]} \mathbb{E} \left[|\theta_0|^{16(\rho+1)} \right] + c_{8(\rho+1)} (1 + 1/(a_D \kappa)) + 1 \right) \\
& \leq \lambda^3 3^{8(\rho+1)+1} dK_0^2 K_h^2 \left(e^{-\lambda a_D \kappa [t]} (2 + \bar{C}_{\mathbf{A}0,2}) \mathbb{E} \left[|\theta_0|^{16(\rho+1)} \right] \right. \\
& \quad \left. + 2c_{8(\rho+1)} (1 + 1/(a_D \kappa)) + 2\tilde{C}_{\mathbf{A}0,2} + 1 \right) \\
& \quad + \lambda^3 2^{16(\rho+1)} K_H^2 K_h^2 \left(e^{-\lambda a_D \kappa [t]} \mathbb{E} \left[|\theta_0|^{16(\rho+1)} \right] + c_{8(\rho+1)} (1 + 1/(a_D \kappa)) + 1 \right) \\
& \leq \lambda^3 \left(e^{-a_D \kappa n/2} \bar{C}_{\mathbf{A}2} \mathbb{E} \left[|\theta_0|^{16(\rho+1)} \right] + \tilde{C}_{\mathbf{A}2} \right),
\end{aligned}$$

where the fifth inequality holds by applying Lemma 6.3 and A.2, the second last inequality holds by using $\bar{C}_{\mathbf{A}0,2} < \tilde{C}_{\mathbf{A}0,2}$, the last inequality holds due to $\lambda [t] \geq \lambda nT \geq n/2$, and where κ is given in (75),

$$\begin{aligned}
\bar{C}_{\mathbf{A}2} & := 3^{24(\rho+1)} d^3 (1 + K_0)^2 (1 + K_h)^2 (1 + K_H)^4 (1 + \beta^{-1})^2 (1 + L)^2 (1 + K_{3,d})^2 \\
& \quad \times (2 + \max\{\bar{C}_{\mathbf{A}0,2}, \bar{C}_{\mathbf{A}0,2q}, \bar{C}_{\mathbf{A}1,2}, \bar{C}_{\mathbf{A}0,2+2q}\}) \\
\tilde{C}_{\mathbf{A}2} & := 3^{24(\rho+1)} d^3 (1 + K_0)^2 (1 + K_h)^2 (1 + K_H)^4 (1 + \beta^{-1})^2 (1 + L)^2 (1 + K_{3,d})^2 \\
& \quad \times \left(2c_{8(\rho+1)} (1 + 1/(a_D \kappa)) + 2 \max\{\tilde{C}_{\mathbf{A}0,2}, \tilde{C}_{\mathbf{A}0,2q}, \tilde{C}_{\mathbf{A}1,2}, \tilde{C}_{\mathbf{A}0,2+2q}\} + 1 \right. \\
& \quad \left. + v_{16(\rho+1)} (M_V(16(\rho+1))) \right)
\end{aligned} \tag{90}$$

with $\bar{C}_{\mathbf{A}0,p}$, $\tilde{C}_{\mathbf{A}0,p}$, $\bar{C}_{\mathbf{A}1,p}$, $\tilde{C}_{\mathbf{A}1,p}$, $p > 0$, given in (87) and (88), and c_p , $M_V(p)$, $p \in [2, \infty) \cap \mathbb{N}$ given in (83) (see also Lemma 6.3) and Lemma 6.4.

(ii) To establish the second inequality, we apply Lemma 6.3 to obtain

$$\begin{aligned}
& \mathbb{E} \left[\left| \lambda^2 \int_{[t]}^t H(\bar{\theta}_s^\lambda) \int_{[s]}^s H_\lambda(\bar{\theta}_{[r]}^\lambda) h_\lambda(\bar{\theta}_{[r]}^\lambda) dr ds \right|^2 \right] \\
& \leq \lambda^4 \int_{[t]}^t \mathbb{E} \left[|H(\bar{\theta}_s^\lambda)|^2 |H_\lambda(\bar{\theta}_{[s]}^\lambda)|^2 |h_\lambda(\bar{\theta}_{[s]}^\lambda)|^2 \right] ds \\
& \leq \lambda^4 \int_{[t]}^t \mathbb{E} \left[K_H^4 (1 + |\bar{\theta}_s^\lambda|^{\rho+q-1})^2 (1 + |\bar{\theta}_{[s]}^\lambda|^{\rho+q-1})^2 K_h^2 (1 + |\bar{\theta}_{[s]}^\lambda|^{\rho+q})^2 \right] ds \\
& \leq \lambda^4 K_H^4 K_h^2 \int_{[t]}^t 3^{16(\rho+1)} \mathbb{E} \left[1 + |\bar{\theta}_s^\lambda|^{16(\rho+1)} + |\bar{\theta}_{[s]}^\lambda|^{16(\rho+1)} \right] ds \\
& \leq \lambda^3 3^{16(\rho+1)} K_H^4 K_h^2 \left(2e^{-\lambda a_D \kappa [t]} \mathbb{E} \left[|\theta_0|^{16(\rho+1)} \right] + 2c_{8(\rho+1)} (1 + 1/(a_D \kappa)) + 1 \right) \\
& \leq \lambda^3 \left(e^{-a_D \kappa n/2} \bar{C}_{\mathbf{A}2} \mathbb{E} [|\theta_0|^{16(\rho+1)}] + \tilde{C}_{\mathbf{A}2} \right),
\end{aligned}$$

where $\bar{C}_{\mathbf{A}2}$ and $\tilde{C}_{\mathbf{A}2}$ are given in (90).

(iii) To obtain the third inequality, we apply Lemma 6.3 and write

$$\begin{aligned}
& \mathbb{E} \left[\left| -\lambda^2 \beta^{-1} \int_{[t]}^t H(\bar{\theta}_s^\lambda) \int_{[s]}^s \Upsilon_\lambda(\bar{\theta}_{[r]}^\lambda) dr ds \right|^2 \right] \\
& \leq \lambda^4 \beta^{-2} \int_{[t]}^t \mathbb{E} \left[|H(\bar{\theta}_s^\lambda)|^2 |\Upsilon_\lambda(\bar{\theta}_{[s]}^\lambda)|^2 \right] ds \\
& \leq \lambda^4 \beta^{-2} \int_{[t]}^t \mathbb{E} \left[K_H^2 (1 + |\bar{\theta}_s^\lambda|^{\rho+q-1})^2 K_{3,d}^2 (1 + |\bar{\theta}_{[s]}^\lambda|)^{2(\rho+q-2)} \right] ds \\
& \leq \lambda^4 \beta^{-2} K_H^2 K_{3,d}^2 \int_{[t]}^t 3^{16(\rho+1)} \mathbb{E} \left[1 + |\bar{\theta}_s^\lambda|^{16(\rho+1)} + |\bar{\theta}_{[s]}^\lambda|^{16(\rho+1)} \right] ds \\
& \leq \lambda^3 3^{16(\rho+1)} \beta^{-2} K_H^2 K_{3,d}^2 \left(2e^{-\lambda a_D \kappa [t]} \mathbb{E} [|\theta_0|^{16(\rho+1)}] + 2c_{8(\rho+1)} (1 + 1/(a_D \kappa)) + 1 \right) \\
& \leq \lambda^3 \left(e^{-a_D \kappa n/2} \bar{C}_{A2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + \tilde{C}_{A2} \right),
\end{aligned}$$

where \bar{C}_{A2} and \tilde{C}_{A2} are given in (90).

(iv) To obtain the fourth inequality, we use Cauchy-Schwarz inequality and Lemma 6.3:

$$\begin{aligned}
& \mathbb{E} \left[\left| -\lambda \sqrt{2\lambda \beta^{-1}} \int_{[t]}^t H(\bar{\theta}_s^\lambda) \int_{[s]}^s H_\lambda(\bar{\theta}_{[r]}^\lambda) dB_r^\lambda ds \right|^2 \right] \\
& \leq 2\lambda^3 \beta^{-1} \int_{[t]}^t \mathbb{E} \left[|H(\bar{\theta}_s^\lambda)|^2 |H_\lambda(\bar{\theta}_{[s]}^\lambda)|^2 \left| \int_{[s]}^s dB_r^\lambda \right|^2 \right] ds \\
& \leq 2\lambda^3 \beta^{-1} \int_{[t]}^t \mathbb{E} \left[K_H^4 (1 + |\bar{\theta}_s^\lambda|^{\rho+q-1})^2 (1 + |\bar{\theta}_{[s]}^\lambda|^{\rho+q-1})^2 \left| \int_{[s]}^s dB_r^\lambda \right|^2 \right] ds \\
& \leq 2\lambda^3 \beta^{-1} K_H^4 \int_{[t]}^t 3^{8(\rho+1)} \left(\mathbb{E} \left[1 + |\bar{\theta}_s^\lambda|^{16(\rho+1)} + |\bar{\theta}_{[s]}^\lambda|^{16(\rho+1)} \right] \right)^{1/2} \\
& \quad \times \left(\mathbb{E} \left[\left| \int_{[s]}^s dB_r^\lambda \right|^4 \right] \right)^{1/2} ds \\
& \leq \lambda^3 \beta^{-1} K_H^4 3^{8(\rho+1)+2} (d+4) \left(2e^{-\lambda a_D \kappa [t]} \mathbb{E} [|\theta_0|^{16(\rho+1)}] + 2c_{8(\rho+1)} (1 + 1/(a_D \kappa)) + 1 \right) \\
& \leq \lambda^3 \left(e^{-a_D \kappa n/2} \bar{C}_{A2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + \tilde{C}_{A2} \right),
\end{aligned}$$

where \bar{C}_{A2} and \tilde{C}_{A2} are given in (90).

(v) To obtain the fifth inequality, we apply Remark 2.2, (89), and Cauchy-Schwarz inequality:

$$\begin{aligned}
& \mathbb{E} \left[\left| \sqrt{2\lambda \beta^{-1}} \int_{[t]}^t \left(H(\bar{\theta}_s^\lambda) - H_\lambda(\bar{\theta}_{[s]}^\lambda) \right) dB_s^\lambda \right|^2 \right] \\
& = 2\lambda \beta^{-1} \int_{[t]}^t \mathbb{E} \left[|H(\bar{\theta}_s^\lambda) - H_\lambda(\bar{\theta}_{[s]}^\lambda)|_F^2 \right] ds \\
& \leq 4\lambda \beta^{-1} \int_{[t]}^t \mathbb{E} \left[|H(\bar{\theta}_s^\lambda) - H(\bar{\theta}_{[s]}^\lambda)|_F^2 \right] ds + 4\lambda \beta^{-1} \int_{[t]}^t \mathbb{E} \left[|H(\bar{\theta}_{[s]}^\lambda) - H_\lambda(\bar{\theta}_{[s]}^\lambda)|_F^2 \right] ds \\
& \leq 4\lambda \beta^{-1} \int_{[t]}^t \mathbb{E} \left[dK_0^2 (1 + |\bar{\theta}_s^\lambda| + |\bar{\theta}_{[s]}^\lambda|)^{2(\rho+q-2)} |\bar{\theta}_s^\lambda - \bar{\theta}_{[s]}^\lambda|^2 \right] ds \\
& \quad + 4\lambda \beta^{-1} \int_{[t]}^t \mathbb{E} \left[\lambda^3 |\bar{\theta}_{[s]}^\lambda|^{6(\rho+q-1)} dK_H^2 (1 + |\bar{\theta}_{[s]}^\lambda|^{\rho+q-1})^2 \right] ds
\end{aligned}$$

$$\begin{aligned}
&\leq 4\lambda\beta^{-1}d\mathcal{K}_0^2 \int_{[t]}^t 3^{8(\rho+1)} \left(\mathbb{E} \left[1 + |\bar{\theta}_s^\lambda|^{16(\rho+1)} + |\bar{\theta}_{[s]}^\lambda|^{16(\rho+1)} \right] \right)^{1/2} \\
&\quad \times \left(\mathbb{E} \left[|\bar{\theta}_s^\lambda - \bar{\theta}_{[s]}^\lambda|^4 \right] \right)^{1/2} ds \\
&\quad + 4\lambda^4\beta^{-1}dK_H^2 \int_{[t]}^t 2^{16(\rho+1)} \mathbb{E} \left[1 + |\bar{\theta}_{[s]}^\lambda|^{16(\rho+1)} \right] ds \\
&\leq \lambda\beta^{-1}d\mathcal{K}_0^2 3^{8(\rho+1)+2} \left(2e^{-\lambda a_D \kappa [t]} \mathbb{E} \left[|\theta_0|^{16(\rho+1)} \right] + 2c_{8(\rho+1)} (1 + 1/(a_D \kappa)) + 1 \right)^{1/2} \\
&\quad \times \lambda \left(e^{-\lambda a_D \kappa [t]} \bar{\mathcal{C}}_{\mathbf{A}0,2} \mathbb{E} [|\theta_0|^{8(\rho+1)}] + \tilde{\mathcal{C}}_{\mathbf{A}0,2} \right)^{1/2} \\
&\quad + \lambda^4\beta^{-1}dK_H^2 2^{16(\rho+1)+2} \left(e^{-\lambda a_D \kappa [t]} \mathbb{E} \left[|\theta_0|^{16(\rho+1)} \right] + c_{8(\rho+1)} (1 + 1/(a_D \kappa)) + 1 \right) \\
&\leq \lambda^2\beta^{-1}d\mathcal{K}_0^2 3^{8(\rho+1)+2} \left(e^{-\lambda a_D \kappa [t]} (2 + \bar{\mathcal{C}}_{\mathbf{A}0,2}) \mathbb{E} \left[|\theta_0|^{16(\rho+1)} \right] \right. \\
&\quad \left. + 2c_{8(\rho+1)} (1 + 1/(a_D \kappa)) + 2\tilde{\mathcal{C}}_{\mathbf{A}0,2} + 1 \right) \\
&\quad + \lambda^2\beta^{-1}dK_H^2 2^{16(\rho+1)+2} \left(e^{-\lambda a_D \kappa [t]} \mathbb{E} \left[|\theta_0|^{16(\rho+1)} \right] + c_{8(\rho+1)} (1 + 1/(a_D \kappa)) + 1 \right) \\
&\leq \lambda^2 \left(e^{-a_D \kappa n/2} \bar{\mathcal{C}}_{\mathbf{A}2} \mathbb{E} [|\theta_0|^{16(\rho+1)}] + \tilde{\mathcal{C}}_{\mathbf{A}2} \right),
\end{aligned}$$

where the fourth inequality holds by applying Lemma 6.3 and A.2, and where $\bar{\mathcal{C}}_{\mathbf{A}2}$ and $\tilde{\mathcal{C}}_{\mathbf{A}2}$ are given in (90).

(vi) To obtain the sixth inequality, we use the same arguments as in (v):

$$\begin{aligned}
&\mathbb{E} \left[\left| \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t \left(H(\bar{\zeta}_s^{\lambda,n}) - H(\bar{\zeta}_{[s]}^{\lambda,n}) \right) dB_s^\lambda \right|^2 \right] \\
&= 2\lambda\beta^{-1} \int_{[t]}^t \mathbb{E} \left[|H(\bar{\zeta}_s^{\lambda,n}) - H(\bar{\zeta}_{[s]}^{\lambda,n})|_F^2 \right] ds \\
&\leq 2\lambda\beta^{-1} \int_{[t]}^t \mathbb{E} \left[d\mathcal{K}_0^2 (1 + |\bar{\zeta}_s^{\lambda,n}| + |\bar{\zeta}_{[s]}^{\lambda,n}|)^{2(\rho+q-2)} |\bar{\zeta}_s^{\lambda,n} - \bar{\zeta}_{[s]}^{\lambda,n}|^2 \right] ds \\
&\leq 2\lambda\beta^{-1}d\mathcal{K}_0^2 \int_{[t]}^t 3^{8(\rho+1)} \left(\mathbb{E} \left[1 + |\bar{\zeta}_s^{\lambda,n}|^{16(\rho+1)} + |\bar{\zeta}_{[s]}^{\lambda,n}|^{16(\rho+1)} \right] \right)^{1/2} \\
&\quad \times \left(\mathbb{E} \left[|\bar{\zeta}_s^{\lambda,n} - \bar{\zeta}_{[s]}^{\lambda,n}|^4 \right] \right)^{1/2} ds \\
&\leq \lambda\beta^{-1}d\mathcal{K}_0^2 3^{8(\rho+1)+1} \left(2^{8(\rho+1)} e^{-\lambda a_D \min\{\kappa, 1/2\} [t]} \mathbb{E} [|\theta_0|^{16(\rho+1)}] \right. \\
&\quad \left. + 2^{8(\rho+1)} (c_{8(\rho+1)} (1 + 1/(a_D \kappa)) + 1) + 6v_{16(\rho+1)} (\mathbf{M}_V(16(\rho+1))) \right)^{1/2} \\
&\quad \times \lambda \left(e^{-\lambda a_D \min\{\kappa, 1/2\} [t]} \bar{\mathcal{C}}_{\mathbf{A}1,2} \mathbb{E} [|\theta_0|^{4(\rho+1)}] + \tilde{\mathcal{C}}_{\mathbf{A}1,2} \right)^{1/2} \\
&\leq \lambda^2\beta^{-1}d\mathcal{K}_0^2 3^{16(\rho+1)} \left(e^{-\lambda a_D \min\{\kappa, 1/2\} [t]} (1 + \bar{\mathcal{C}}_{\mathbf{A}1,2}) \mathbb{E} [|\theta_0|^{16(\rho+1)}] \right. \\
&\quad \left. + c_{8(\rho+1)} (1 + 1/(a_D \kappa)) + 1 + 2\tilde{\mathcal{C}}_{\mathbf{A}1,2} + v_{16(\rho+1)} (\mathbf{M}_V(16(\rho+1))) \right) \\
&\leq \lambda^2 \left(e^{-a_D \min\{\kappa, 1/2\} n/2} \bar{\mathcal{C}}_{\mathbf{A}2} \mathbb{E} [|\theta_0|^{16(\rho+1)}] + \tilde{\mathcal{C}}_{\mathbf{A}2} \right),
\end{aligned}$$

where the third inequality holds due to Lemma 6.5 and A.2, the fourth inequality holds due to $\bar{\mathcal{C}}_{\mathbf{A}1,2} < \tilde{\mathcal{C}}_{\mathbf{A}1,2}$, and where $\bar{\mathcal{C}}_{\mathbf{A}2}$ and $\tilde{\mathcal{C}}_{\mathbf{A}2}$ are given in (90).

(vii) To establish the seventh inequality, we apply Remark 2.2, (89), and Cauchy-Schwarz inequality to obtain

$$\mathbb{E} \left[\left| \lambda\beta^{-1} \int_{[t]}^t \left(\Upsilon(\bar{\theta}_s^\lambda) - \Upsilon_\lambda(\bar{\theta}_{[s]}^\lambda) \right) ds \right|^2 \right]$$

$$\begin{aligned}
&\leq \lambda^2 \beta^{-2} \int_{[t]}^t \mathbb{E} \left[|\Upsilon(\bar{\theta}_s^\lambda) - \Upsilon_\lambda(\bar{\theta}_{[s]}^\lambda)|^2 \right] ds \\
&\leq 2\lambda^2 \beta^{-2} \int_{[t]}^t \mathbb{E} \left[|\Upsilon(\bar{\theta}_s^\lambda) - \Upsilon(\bar{\theta}_{[s]}^\lambda)|^2 \right] ds + 2\lambda^2 \beta^{-2} \int_{[t]}^t \mathbb{E} \left[|\Upsilon(\bar{\theta}_{[s]}^\lambda) - \Upsilon_\lambda(\bar{\theta}_{[s]}^\lambda)|^2 \right] ds \\
&\leq 2\lambda^2 \beta^{-2} \int_{[t]}^t \mathbb{E} \left[d^3 L^2 (1 + |\bar{\theta}_s^\lambda| + |\bar{\theta}_{[s]}^\lambda|)^{2(\rho-2)} |\bar{\theta}_s^\lambda - \bar{\theta}_{[s]}^\lambda|^{2q} \right] ds \\
&\quad + 2\lambda^2 \beta^{-2} \int_{[t]}^t \mathbb{E} \left[\lambda^3 |\bar{\theta}_{[s]}^\lambda|^{6(\rho+q-1)} \mathbf{K}_{3,d}^2 (1 + |\bar{\theta}_{[s]}^\lambda|)^{2(\rho+q-2)} \right] ds \\
&\leq 2\lambda^2 \beta^{-2} d^3 L^2 \int_{[t]}^t 3^{8(\rho+1)} \left(\mathbb{E} \left[1 + |\bar{\theta}_s^\lambda|^{16(\rho+1)} + |\bar{\theta}_{[s]}^\lambda|^{16(\rho+1)} \right] \right)^{1/2} \\
&\quad \times \left(\mathbb{E} \left[|\bar{\theta}_s^\lambda - \bar{\theta}_{[s]}^\lambda|^{4q} \right] \right)^{1/2} ds \\
&\quad + 2\lambda^5 \beta^{-2} \mathbf{K}_{3,d}^2 \int_{[t]}^t 2^{16(\rho+1)} \mathbb{E} \left[1 + |\bar{\theta}_{[s]}^\lambda|^{16(\rho+1)} \right] ds \\
&\leq \lambda^2 \beta^{-2} d^3 L^2 3^{8(\rho+1)+1} \left(2e^{-\lambda \mathbf{a}_D \kappa [t]} \mathbb{E} \left[|\theta_0|^{16(\rho+1)} \right] + 2\mathbf{c}_{8(\rho+1)} (1 + 1/(\mathbf{a}_D \kappa)) + 1 \right)^{1/2} \\
&\quad \times \lambda^q \left(e^{-\lambda \mathbf{a}_D \kappa [t]} \bar{\mathbf{C}}_{\mathbf{A}0,2q} \mathbb{E} \left[|\theta_0|^{4[2q](\rho+1)} \right] + \tilde{\mathbf{C}}_{\mathbf{A}0,2q} \right)^{1/2} \\
&\quad + \lambda^5 \beta^{-2} \mathbf{K}_{3,d}^2 2^{16(\rho+1)+1} \left(e^{-\lambda \mathbf{a}_D \kappa [t]} \mathbb{E} \left[|\theta_0|^{16(\rho+1)} \right] + \mathbf{c}_{8(\rho+1)} (1 + 1/(\mathbf{a}_D \kappa)) + 1 \right) \\
&\leq \lambda^{2+q} \beta^{-2} d^3 L^2 3^{8(\rho+1)+1} \\
&\quad \times \left(e^{-\lambda \mathbf{a}_D \kappa [t]} (2 + \bar{\mathbf{C}}_{\mathbf{A}0,2q}) \mathbb{E} \left[|\theta_0|^{16(\rho+1)} \right] + 2\mathbf{c}_{8(\rho+1)} (1 + 1/(\mathbf{a}_D \kappa)) + 2\tilde{\mathbf{C}}_{\mathbf{A}0,2q} + 1 \right) \\
&\quad + \lambda^{2+q} \beta^{-2} \mathbf{K}_{3,d}^2 2^{16(\rho+1)+1} \left(e^{-\lambda \mathbf{a}_D \kappa [t]} \mathbb{E} \left[|\theta_0|^{16(\rho+1)} \right] + \mathbf{c}_{8(\rho+1)} (1 + 1/(\mathbf{a}_D \kappa)) + 1 \right) \\
&\leq \lambda^{2+q} \left(e^{-\mathbf{a}_D \kappa n/2} \bar{\mathbf{C}}_{\mathbf{A}2} \mathbb{E} \left[|\theta_0|^{16(\rho+1)} \right] + \tilde{\mathbf{C}}_{\mathbf{A}2} \right),
\end{aligned}$$

where the fourth inequality holds due to Lemma 6.3 and A.2, and $\bar{\mathbf{C}}_{\mathbf{A}2}$, $\tilde{\mathbf{C}}_{\mathbf{A}2}$ are given in (90).
(viii) To obtain the eighth inequality, we apply Remark 2.2 and write

$$\begin{aligned}
&\mathbb{E} \left[\left| -\lambda \int_{[t]}^t H(\bar{\theta}_s^\lambda) h_\lambda(\bar{\theta}_{[s]}^\lambda) ds \right|^2 + \left| \lambda \beta^{-1} \int_{[t]}^t \Upsilon(\bar{\theta}_s^\lambda) ds \right|^2 \right] \\
&\leq \lambda^2 \int_{[t]}^t \mathbb{E} \left[|H(\bar{\theta}_s^\lambda)|^2 |h_\lambda(\bar{\theta}_{[s]}^\lambda)|^2 \right] ds + \lambda^2 \beta^{-2} \int_{[t]}^t \mathbb{E} \left[|\Upsilon(\bar{\theta}_s^\lambda)|^2 \right] ds \\
&\leq \lambda^2 \int_{[t]}^t \mathbb{E} \left[K_H^2 (1 + |\bar{\theta}_s^\lambda|^{\rho+q-1})^2 K_h^2 (1 + |\bar{\theta}_{[s]}^\lambda|^{\rho+q})^2 \right] ds \\
&\quad + \lambda^2 \beta^{-2} \int_{[t]}^t \mathbb{E} \left[\mathbf{K}_{3,d}^2 (1 + |\bar{\theta}_s^\lambda|)^{2(\rho+q-2)} \right] ds \\
&\leq \lambda^2 (1 + K_H)^2 (1 + K_h)^2 (1 + \beta^{-1})^2 (1 + \mathbf{K}_{3,d})^2 3^{16(\rho+1)} \\
&\quad \times \int_{[t]}^t \mathbb{E} \left[1 + |\bar{\theta}_s^\lambda|^{16(\rho+1)} + |\bar{\theta}_{[s]}^\lambda|^{16(\rho+1)} \right] ds \\
&\leq \lambda^2 (1 + K_H)^2 (1 + K_h)^2 (1 + \beta^{-1})^2 (1 + \mathbf{K}_{3,d})^2 3^{16(\rho+1)} \\
&\quad \times \left(2e^{-\lambda \mathbf{a}_D \kappa [t]} \mathbb{E} \left[|\theta_0|^{16(\rho+1)} \right] + 2\mathbf{c}_{8(\rho+1)} (1 + 1/(\mathbf{a}_D \kappa)) + 1 \right) \\
&\leq \lambda^2 \left(e^{-\mathbf{a}_D \kappa n/2} \bar{\mathbf{C}}_{\mathbf{A}2} \mathbb{E} \left[|\theta_0|^{16(\rho+1)} \right] + \tilde{\mathbf{C}}_{\mathbf{A}2} \right),
\end{aligned}$$

where the fourth inequality holds due to Lemma 6.3, and where $\bar{\mathbf{C}}_{\mathbf{A}2}$, $\tilde{\mathbf{C}}_{\mathbf{A}2}$ are given in (90).

(ix) To establish the last inequality, we follow the arguments in (viii):

$$\begin{aligned}
& \mathbb{E} \left[\left| -\lambda \int_{[t]}^t H(\bar{\zeta}_s^{\lambda,n}) h(\bar{\zeta}_s^{\lambda,n}) ds \right|^2 + \left| \lambda \beta^{-1} \int_{[t]}^t \Upsilon(\bar{\zeta}_s^{\lambda,n}) ds \right|^2 \right] \\
& \leq \lambda^2 (1 + K_H)^2 (1 + K_h)^2 (1 + \beta^{-1})^2 (1 + \mathbf{K}_{3,d})^2 \mathfrak{Z}^{16(\rho+1)} \\
& \quad \times \int_{[t]}^t \mathbb{E} \left[1 + |\bar{\zeta}_s^{\lambda,n}|^{16(\rho+1)} + |\bar{\zeta}_{[s]}^{\lambda,n}|^{16(\rho+1)} \right] ds \\
& \leq \lambda^2 (1 + K_H)^2 (1 + K_h)^2 (1 + \beta^{-1})^2 (1 + \mathbf{K}_{3,d})^2 \mathfrak{Z}^{16(\rho+1)} \\
& \quad \times \left(2^{8(\rho+1)} e^{-\lambda \mathbf{a}_D \min\{\kappa, 1/2\} [t]} \mathbb{E}[|\theta_0|^{16(\rho+1)}] \right. \\
& \quad \left. + 2^{8(\rho+1)} (\mathbf{c}_{8(\rho+1)} (1 + 1/(\mathbf{a}_D \kappa)) + 1) + 6 \mathbf{v}_{16(\rho+1)} (\mathbf{M}_V(16(\rho+1))) \right) \\
& \leq \lambda^2 \left(e^{-\mathbf{a}_D \min\{\kappa, 1/2\} n/2} \bar{\mathbf{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + \tilde{\mathbf{C}}_{\mathbf{A}2} \right),
\end{aligned}$$

where the second inequality holds due to Lemma 6.5, and where $\bar{\mathbf{C}}_{\mathbf{A}2}$ and $\tilde{\mathbf{C}}_{\mathbf{A}2}$ are given in (90).

This completes the proof. \square

Corollary A.4. *Let Assumptions 1, 2, and 3 hold. Then, for any $0 < \lambda \leq \lambda_{\max}$, $n \in \mathbb{N}_0$, $t \geq nT$, we obtain the following inequalities:*

$$\begin{aligned}
& \mathbb{E} \left[\left| h(\bar{\theta}_t^\lambda) - h(\bar{\theta}_{[t]}^\lambda) + \lambda \int_{[t]}^t \left(H_\lambda(\bar{\theta}_{[s]}^\lambda) h_\lambda(\bar{\theta}_{[s]}^\lambda) - \beta^{-1} \Upsilon_\lambda(\bar{\theta}_{[s]}^\lambda) \right) ds \right. \right. \\
& \quad \left. \left. - \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t H_\lambda(\bar{\theta}_{[s]}^\lambda) dB_s^\lambda \right|^2 \right] \leq \lambda^2 \left(e^{-\mathbf{a}_D \kappa n/2} 36 \bar{\mathbf{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + 36 \tilde{\mathbf{C}}_{\mathbf{A}2} \right), \\
& \mathbb{E} \left[\left| h(\bar{\zeta}_t^{\lambda,n}) - h(\bar{\zeta}_{[t]}^{\lambda,n}) - \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t H(\bar{\zeta}_{[s]}^{\lambda,n}) dB_s^\lambda - \left(h(\bar{\theta}_t^\lambda) - h(\bar{\theta}_{[t]}^\lambda) \right. \right. \right. \\
& \quad \left. \left. - \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t H(\bar{\theta}_{[s]}^\lambda) dB_s^\lambda \right|^2 \right] \leq \lambda^2 \left(e^{-\mathbf{a}_D \min\{\kappa, 1/2\} n/2} 72 \bar{\mathbf{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + 72 \tilde{\mathbf{C}}_{\mathbf{A}2} \right),
\end{aligned}$$

where $\bar{\mathbf{C}}_{\mathbf{A}2}$ and $\tilde{\mathbf{C}}_{\mathbf{A}2}$ are given in (90).

Proof. For any $t \geq nT$, by applying Itô's formula to $h(\bar{\theta}_t^\lambda)$, we obtain, almost surely

$$\begin{aligned}
h(\bar{\theta}_t^\lambda) - h(\bar{\theta}_{[t]}^\lambda) &= -\lambda \int_{[t]}^t H(\bar{\theta}_s^\lambda) h_\lambda(\bar{\theta}_{[s]}^\lambda) ds + \lambda^2 \int_{[t]}^t H(\bar{\theta}_s^\lambda) \int_{[s]}^s H_\lambda(\bar{\theta}_{[r]}^\lambda) h_\lambda(\bar{\theta}_{[r]}^\lambda) dr ds \\
& \quad - \lambda^2 \beta^{-1} \int_{[t]}^t H(\bar{\theta}_s^\lambda) \int_{[s]}^s \Upsilon_\lambda(\bar{\theta}_{[r]}^\lambda) dr ds \\
& \quad - \lambda \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t H(\bar{\theta}_s^\lambda) \int_{[s]}^s H_\lambda(\bar{\theta}_{[r]}^\lambda) dB_r^\lambda ds \\
& \quad + \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t H(\bar{\theta}_s^\lambda) dB_s^\lambda + \lambda \beta^{-1} \int_{[t]}^t \Upsilon(\bar{\theta}_s^\lambda) ds.
\end{aligned} \tag{91}$$

Similarly, applying Itô's formula to $h(\bar{\zeta}_t^{\lambda,n})$ yields, almost surely

$$\begin{aligned}
h(\bar{\zeta}_t^{\lambda,n}) - h(\bar{\zeta}_{[t]}^{\lambda,n}) &= -\lambda \int_{[t]}^t H(\bar{\zeta}_s^{\lambda,n}) h(\bar{\zeta}_s^{\lambda,n}) ds + \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t H(\bar{\zeta}_s^{\lambda,n}) dB_s^\lambda \\
& \quad + \lambda \beta^{-1} \int_{[t]}^t \Upsilon(\bar{\zeta}_s^{\lambda,n}) ds.
\end{aligned} \tag{92}$$

(i) To obtain the first inequality, by using (91) with $(\sum_{l=1}^v u_l)^2 \leq v \sum_{l=1}^v u_l^2$, $v \in \mathbb{N}$, $u_l \geq 0$, we obtain that

$$\begin{aligned}
& \mathbb{E} \left[\left| h(\bar{\theta}_t^\lambda) - h(\bar{\theta}_{[t]}^\lambda) + \lambda \int_{[t]}^t \left(H_\lambda(\bar{\theta}_{[s]}^\lambda) h_\lambda(\bar{\theta}_{[s]}^\lambda) - \beta^{-1} \Upsilon_\lambda(\bar{\theta}_{[s]}^\lambda) \right) ds \right. \right. \\
& \quad \left. \left. - \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t H_\lambda(\bar{\theta}_{[s]}^\lambda) dB_s^\lambda \right|^2 \right] \\
& \leq 6\mathbb{E} \left[\left| -\lambda \int_{[t]}^t \left(H(\bar{\theta}_s^\lambda) - H_\lambda(\bar{\theta}_{[s]}^\lambda) \right) h_\lambda(\bar{\theta}_{[s]}^\lambda) ds \right|^2 \right] \\
& \quad + 6\mathbb{E} \left[\left| \lambda^2 \int_{[t]}^t H(\bar{\theta}_s^\lambda) \int_{[s]}^s H_\lambda(\bar{\theta}_{[r]}^\lambda) h_\lambda(\bar{\theta}_{[r]}^\lambda) dr ds \right|^2 \right] \\
& \quad + 6\mathbb{E} \left[\left| -\lambda^2 \beta^{-1} \int_{[t]}^t H(\bar{\theta}_s^\lambda) \int_{[s]}^s \Upsilon_\lambda(\bar{\theta}_{[r]}^\lambda) dr ds \right|^2 \right] \\
& \quad + 6\mathbb{E} \left[\left| -\lambda \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t H(\bar{\theta}_s^\lambda) \int_{[s]}^s H_\lambda(\bar{\theta}_{[r]}^\lambda) dB_r^\lambda ds \right|^2 \right] \\
& \quad + 6\mathbb{E} \left[\left| \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t \left(H(\bar{\theta}_s^\lambda) - H_\lambda(\bar{\theta}_{[s]}^\lambda) \right) dB_s^\lambda \right|^2 \right] \\
& \quad + 6\mathbb{E} \left[\left| \lambda \beta^{-1} \int_{[t]}^t \left(\Upsilon(\bar{\theta}_s^\lambda) - \Upsilon_\lambda(\bar{\theta}_{[s]}^\lambda) \right) ds \right|^2 \right] \\
& \leq \lambda^2 \left(e^{-\mathfrak{a}_0 \kappa n / 2} 36 \bar{\mathbf{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + 36 \tilde{\mathbf{C}}_{\mathbf{A}2} \right),
\end{aligned}$$

where the last inequality holds due to Lemma A.3.

(ii) To establish the second inequality, we use (91) and (92) to obtain

$$\begin{aligned}
& \mathbb{E} \left[\left| h(\bar{\zeta}_t^{\lambda,n}) - h(\bar{\zeta}_{[t]}^{\lambda,n}) - \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t H(\bar{\zeta}_{[s]}^{\lambda,n}) dB_s^\lambda \right. \right. \\
& \quad \left. \left. - \left(h(\bar{\theta}_t^\lambda) - h(\bar{\theta}_{[t]}^\lambda) - \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t H(\bar{\theta}_{[s]}^\lambda) dB_s^\lambda \right) \right|^2 \right] \\
& \leq 2\mathbb{E} \left[\left| h(\bar{\zeta}_t^{\lambda,n}) - h(\bar{\zeta}_{[t]}^{\lambda,n}) - \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t H(\bar{\zeta}_{[s]}^{\lambda,n}) dB_s^\lambda \right|^2 \right] \\
& \quad + 2\mathbb{E} \left[\left| h(\bar{\theta}_t^\lambda) - h(\bar{\theta}_{[t]}^\lambda) - \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t H(\bar{\theta}_{[s]}^\lambda) dB_s^\lambda \right|^2 \right] \\
& \leq 6\mathbb{E} \left[\left| -\lambda \int_{[t]}^t H(\bar{\zeta}_s^{\lambda,n}) h(\bar{\zeta}_s^{\lambda,n}) ds \right|^2 \right] + 6\mathbb{E} \left[\left| \lambda \beta^{-1} \int_{[t]}^t \Upsilon(\bar{\zeta}_s^{\lambda,n}) ds \right|^2 \right] \\
& \quad + 6\mathbb{E} \left[\left| \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t \left(H(\bar{\zeta}_s^{\lambda,n}) - H(\bar{\zeta}_{[s]}^{\lambda,n}) \right) dB_s^\lambda \right|^2 \right] \\
& \quad + 12\mathbb{E} \left[\left| -\lambda \int_{[t]}^t H(\bar{\theta}_s^\lambda) h_\lambda(\bar{\theta}_{[s]}^\lambda) ds \right|^2 \right] + 12\mathbb{E} \left[\left| \lambda \beta^{-1} \int_{[t]}^t \Upsilon(\bar{\theta}_s^\lambda) ds \right|^2 \right]
\end{aligned}$$

$$\begin{aligned}
& + 12\mathbb{E} \left[\left| \lambda^2 \int_{[t]}^t H(\bar{\theta}_s^\lambda) \int_{[s]}^s H_\lambda(\bar{\theta}_{[r]}^\lambda) h_\lambda(\bar{\theta}_{[r]}^\lambda) dr ds \right|^2 \right] \\
& + 12\mathbb{E} \left[\left| -\lambda^2 \beta^{-1} \int_{[t]}^t H(\bar{\theta}_s^\lambda) \int_{[s]}^s \Upsilon_\lambda(\bar{\theta}_{[r]}^\lambda) dr ds \right|^2 \right] \\
& + 12\mathbb{E} \left[\left| -\lambda \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t H(\bar{\theta}_s^\lambda) \int_{[s]}^s H_\lambda(\bar{\theta}_{[r]}^\lambda) dB_r^\lambda ds \right|^2 \right] \\
& + 12\mathbb{E} \left[\left| \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t \left(H(\bar{\theta}_s^\lambda) - H(\bar{\theta}_{[s]}^\lambda) \right) dB_s^\lambda \right|^2 \right] \\
& \leq 12\lambda^2 \left(e^{-\text{aD} \min\{\kappa, 1/2\} n/2} \bar{\mathcal{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + \tilde{\mathcal{C}}_{\mathbf{A}2} \right) \\
& \quad + 60\lambda^2 \left(e^{-\text{aD} \min\{\kappa, 1/2\} n/2} \bar{\mathcal{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + \tilde{\mathcal{C}}_{\mathbf{A}2} \right) \\
& \leq \lambda^2 \left(e^{-\text{aD} \min\{\kappa, 1/2\} n/2} 72\bar{\mathcal{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + 72\tilde{\mathcal{C}}_{\mathbf{A}2} \right),
\end{aligned}$$

where the second last inequality holds by using Lemma A.3 with the fact that

$$\begin{aligned}
& \mathbb{E} \left[\left| \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t \left(H(\bar{\theta}_s^\lambda) - H(\bar{\theta}_{[s]}^\lambda) \right) dB_s^\lambda \right|^2 \right] \\
& = 2\lambda\beta^{-1} \int_{[t]}^t \mathbb{E} \left[|H(\bar{\theta}_s^\lambda) - H(\bar{\theta}_{[s]}^\lambda)|_{\mathbb{F}}^2 \right] ds \\
& \leq \lambda^2 \left(e^{-\text{aD} \min\{\kappa, 1/2\} n/2} \bar{\mathcal{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + \tilde{\mathcal{C}}_{\mathbf{A}2} \right)
\end{aligned}$$

as indicated in the calculations in (v) of Lemma A.3.

This completes the proof. \square

Definition A.5. Define $\mathfrak{M} = (\mathfrak{M}^{(i,j)})_{i,j=1,\dots,d} : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ by setting, for every $i, j = 1, \dots, d$,

$$\mathfrak{M}^{(i,j)}(\theta, \bar{\theta}) = \langle \nabla H^{(i,j)}(\bar{\theta}), \theta - \bar{\theta} \rangle, \quad \theta, \bar{\theta} \in \mathbb{R}^d.$$

Lemma A.6. Let Assumptions 1, 2, and 3 hold. Then, for any $0 < \lambda \leq \lambda_{\max}$, $n \in \mathbb{N}_0$, $t \geq nT$, we obtain the following inequalities:

$$\begin{aligned}
& \mathbb{E} \left[\left| \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t \left(H(\bar{\theta}_s^\lambda) - H(\bar{\theta}_{[s]}^\lambda) - \mathfrak{M}(\bar{\theta}_s^\lambda, \bar{\theta}_{[s]}^\lambda) \right) dB_s^\lambda \right|^2 \right] \\
& \leq \lambda^{2+q} \left(e^{-\text{aD} \kappa n/2} \bar{\mathcal{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + \tilde{\mathcal{C}}_{\mathbf{A}2} \right),
\end{aligned} \tag{93}$$

$$\mathbb{E} \left[\left| \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t \mathfrak{M}(\bar{\theta}_s^\lambda, \bar{\theta}_{[s]}^\lambda) dB_s^\lambda \right|^2 \right] \leq \lambda^2 \left(e^{-\text{aD} \kappa n/2} \bar{\mathcal{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + \tilde{\mathcal{C}}_{\mathbf{A}2} \right), \tag{94}$$

$$\begin{aligned}
& \mathbb{E} \left[2\lambda\beta^{-1} \left\langle \int_{[t]}^t \int_{[s]}^s \left(H(\bar{\zeta}_{[r]}^{\lambda,n}) - H(\bar{\theta}_{[r]}^\lambda) \right) dB_r^\lambda ds, \int_{[t]}^t \mathfrak{M}(\bar{\theta}_s^\lambda, \bar{\theta}_{[s]}^\lambda) dB_s^\lambda \right\rangle \right] \\
& \leq \lambda^2 \left(e^{-\text{aD} \min\{\kappa, 1/2\} n/2} 10\bar{\mathcal{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + 10\tilde{\mathcal{C}}_{\mathbf{A}2} \right),
\end{aligned} \tag{95}$$

where $\bar{\mathcal{C}}_{\mathbf{A}2}$ and $\tilde{\mathcal{C}}_{\mathbf{A}2}$ are given in (90).

Proof. To show the inequalities hold, we follow the arguments below:

(i) To establish the first inequality (93), by using Definition A.5, we observe that, for fixed $\theta, \bar{\theta} \in \mathbb{R}^d$,

$$|H(\theta) - H(\bar{\theta}) - \mathfrak{M}(\theta, \bar{\theta})|_{\mathbb{F}}^2$$

$$\begin{aligned}
&= \sum_{i,j=1}^d |H^{(i,j)}(\theta) - H^{(i,j)}(\bar{\theta}) - \mathfrak{M}^{(i,j)}(\theta, \bar{\theta})|^2 \\
&= \sum_{i,j=1}^d \left| \int_0^1 \langle \nabla H^{(i,j)}(\nu\theta + (1-\nu)\bar{\theta}), \theta - \bar{\theta} \rangle d\nu - \langle \nabla H^{(i,j)}(\bar{\theta}), \theta - \bar{\theta} \rangle \right|^2 \\
&\leq \int_0^1 \sum_{i,j=1}^d |\nabla H^{(i,j)}(\nu\theta + (1-\nu)\bar{\theta}) - \nabla H^{(i,j)}(\bar{\theta})|^2 d\nu |\theta - \bar{\theta}|^2 \\
&\leq \int_0^1 \sum_{i=1}^d d |\nabla^2 h^{(i)}(\nu\theta + (1-\nu)\bar{\theta}) - \nabla^2 h^{(i)}(\bar{\theta})|^2 d\nu |\theta - \bar{\theta}|^2 \\
&\leq 2^{2(\rho-2)} d^2 L^2 (1 + |\theta| + |\bar{\theta}|)^{2(\rho-2)} |\theta - \bar{\theta}|^{2+2q}. \tag{96}
\end{aligned}$$

By using (96), we obtain that

$$\begin{aligned}
&\mathbb{E} \left[\left| \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t \left(H(\bar{\theta}_s^\lambda) - H(\bar{\theta}_{[s]}^\lambda) - \mathfrak{M}(\bar{\theta}_s^\lambda, \bar{\theta}_{[s]}^\lambda) \right) dB_s^\lambda \right|^2 \right] \\
&= 2\lambda\beta^{-1} \int_{[t]}^t \mathbb{E} \left[\left| H(\bar{\theta}_s^\lambda) - H(\bar{\theta}_{[s]}^\lambda) - \mathfrak{M}(\bar{\theta}_s^\lambda, \bar{\theta}_{[s]}^\lambda) \right|_{\mathbb{F}}^2 \right] ds \\
&\leq \lambda\beta^{-1} 2^{2\rho-3} d^2 L^2 \int_{[t]}^t \mathbb{E} \left[(1 + |\bar{\theta}_s^\lambda| + |\bar{\theta}_{[s]}^\lambda|)^{2(\rho-2)} |\bar{\theta}_s^\lambda - \bar{\theta}_{[s]}^\lambda|^{2+2q} \right] ds \\
&\leq \lambda\beta^{-1} 2^{2\rho-3} d^2 L^2 \int_{[t]}^t 3^{8(\rho+1)} \left(\mathbb{E} \left[1 + |\bar{\theta}_s^\lambda|^{16(\rho+1)} + |\bar{\theta}_{[s]}^\lambda|^{16(\rho+1)} \right] \right)^{1/2} \\
&\quad \times \left(\mathbb{E} \left[|\bar{\theta}_s^\lambda - \bar{\theta}_{[s]}^\lambda|^{4+4q} \right] \right)^{1/2} ds \\
&\leq \lambda\beta^{-1} 3^{16(\rho+1)} d^2 L^2 \left(2e^{-\lambda a_D \kappa [t]} \mathbb{E} \left[|\theta_0|^{16(\rho+1)} \right] + 2c_{8(\rho+1)} (1 + 1/(a_D \kappa)) + 1 \right)^{1/2} \\
&\quad \times \lambda^{1+q} \left(e^{-\lambda a_D \kappa [t]} \bar{\mathcal{C}}_{\mathbf{A}0, 2+2q} \mathbb{E} \left[|\theta_0|^{4[2+2q](\rho+1)} \right] + \tilde{\mathcal{C}}_{\mathbf{A}0, 2+2q} \right)^{1/2} \\
&\leq \lambda^{2+q} \left(e^{-a_D \kappa n/2} \bar{\mathcal{C}}_{\mathbf{A}2} \mathbb{E} \left[|\theta_0|^{16(\rho+1)} \right] + \tilde{\mathcal{C}}_{\mathbf{A}2} \right),
\end{aligned}$$

where the third inequality holds due to Lemma 6.3 and A.2, and $\bar{\mathcal{C}}_{\mathbf{A}2}$, $\tilde{\mathcal{C}}_{\mathbf{A}2}$ are given in (90).

(ii) To establish the second inequality (94), we recall Definition A.5 and use Remark 2.2 to obtain

$$\begin{aligned}
&\mathbb{E} \left[\left| \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t \mathfrak{M}(\bar{\theta}_s^\lambda, \bar{\theta}_{[s]}^\lambda) dB_s^\lambda \right|^2 \right] \\
&= 2\lambda\beta^{-1} \int_{[t]}^t \mathbb{E} \left[\sum_{i,j=1}^d \left| \langle \nabla H^{(i,j)}(\bar{\theta}_{[s]}^\lambda), \bar{\theta}_s^\lambda - \bar{\theta}_{[s]}^\lambda \rangle \right|^2 \right] ds \\
&\leq 2\lambda\beta^{-1} \int_{[t]}^t \mathbb{E} \left[d^2 \mathcal{K}_0^2 (1 + |\bar{\theta}_{[s]}^\lambda|)^{2(\rho+q-2)} |\bar{\theta}_s^\lambda - \bar{\theta}_{[s]}^\lambda|^2 \right] ds \\
&\leq \lambda\beta^{-1} d^2 \mathcal{K}_0^2 \int_{[t]}^t 2^{8(\rho+1)+1} \left(\mathbb{E} \left[1 + |\bar{\theta}_{[s]}^\lambda|^{16(\rho+1)} \right] \right)^{1/2} \left(\mathbb{E} \left[|\bar{\theta}_s^\lambda - \bar{\theta}_{[s]}^\lambda|^4 \right] \right)^{1/2} ds \\
&\leq \lambda\beta^{-1} d^2 \mathcal{K}_0^2 2^{8(\rho+1)+1} \left(e^{-\lambda a_D \kappa [t]} \mathbb{E} \left[|\theta_0|^{16(\rho+1)} \right] + c_{8(\rho+1)} (1 + 1/(a_D \kappa)) + 1 \right)^{1/2} \\
&\quad \times \lambda \left(e^{-\lambda a_D \kappa [t]} \bar{\mathcal{C}}_{\mathbf{A}0, 2} \mathbb{E} \left[|\theta_0|^{8(\rho+1)} \right] + \tilde{\mathcal{C}}_{\mathbf{A}0, 2} \right)^{1/2} \\
&\leq \lambda^2 \left(e^{-a_D \kappa n/2} \bar{\mathcal{C}}_{\mathbf{A}2} \mathbb{E} \left[|\theta_0|^{16(\rho+1)} \right] + \tilde{\mathcal{C}}_{\mathbf{A}2} \right),
\end{aligned}$$

where the third inequality holds due to Lemma 6.3 and A.2, and $\bar{\mathcal{C}}_{\mathbf{A}2}$, $\tilde{\mathcal{C}}_{\mathbf{A}2}$ are given in (90).

(iii) To obtain the third inequality (95), we use Definition A.5 and write the following

$$\begin{aligned}
& \mathbb{E} \left[2\lambda\beta^{-1} \left\langle \int_{[t]}^t \int_{[s]}^s \left(H(\bar{\zeta}_{[r]}^{\lambda,n}) - H(\bar{\theta}_{[r]}^\lambda) \right) dB_r^\lambda ds, \int_{[t]}^t \mathfrak{M}(\bar{\theta}_s^\lambda, \bar{\theta}_{[s]}^\lambda) dB_s^\lambda \right\rangle \right] \\
&= 2\lambda\beta^{-1} \mathbb{E} \left[\left\langle \int_{[t]}^t \int_{[s]}^s \left(H(\bar{\zeta}_{[r]}^{\lambda,n}) - H(\bar{\theta}_{[r]}^\lambda) \right) dB_r^\lambda ds, \right. \right. \\
&\quad \left. \int_{[t]}^t \left\langle \nabla H(\bar{\theta}_{[s]}^\lambda), \int_{[s]}^s -\lambda h_\lambda(\bar{\theta}_{[r]}^\lambda) dr \right\rangle dB_s^\lambda \right\rangle \\
&\quad + 2\lambda\beta^{-1} \mathbb{E} \left[\left\langle \int_{[t]}^t \int_{[s]}^s \left(H(\bar{\zeta}_{[r]}^{\lambda,n}) - H(\bar{\theta}_{[r]}^\lambda) \right) dB_r^\lambda ds, \right. \right. \\
&\quad \left. \int_{[t]}^t \left\langle \nabla H(\bar{\theta}_{[s]}^\lambda), \int_{[s]}^s \int_{[r]}^r \lambda^2 H_\lambda(\bar{\theta}_{[\nu]}^\lambda) h_\lambda(\bar{\theta}_{[\nu]}^\lambda) d\nu dr \right\rangle dB_s^\lambda \right\rangle \\
&\quad + 2\lambda\beta^{-1} \mathbb{E} \left[\left\langle \int_{[t]}^t \int_{[s]}^s \left(H(\bar{\zeta}_{[r]}^{\lambda,n}) - H(\bar{\theta}_{[r]}^\lambda) \right) dB_r^\lambda ds, \right. \right. \\
&\quad \left. \int_{[t]}^t \left\langle \nabla H(\bar{\theta}_{[s]}^\lambda), \int_{[s]}^s \int_{[r]}^r -\lambda^2 \beta^{-1} \Upsilon_\lambda(\bar{\theta}_{[\nu]}^\lambda) d\nu dr \right\rangle dB_s^\lambda \right\rangle \tag{97} \\
&\quad + 2\lambda\beta^{-1} \mathbb{E} \left[\left\langle \int_{[t]}^t \int_{[s]}^s \left(H(\bar{\zeta}_{[r]}^{\lambda,n}) - H(\bar{\theta}_{[r]}^\lambda) \right) dB_r^\lambda ds, \right. \right. \\
&\quad \left. \int_{[t]}^t \left\langle \nabla H(\bar{\theta}_{[s]}^\lambda), \int_{[s]}^s -\lambda \sqrt{2\lambda\beta^{-1}} \int_{[r]}^r H_\lambda(\bar{\theta}_{[\nu]}^\lambda) dB_\nu^\lambda dr \right\rangle dB_s^\lambda \right\rangle \\
&\quad + 2\lambda\beta^{-1} \mathbb{E} \left[\left\langle \int_{[t]}^t \int_{[s]}^s \left(H(\bar{\zeta}_{[r]}^{\lambda,n}) - H(\bar{\theta}_{[r]}^\lambda) \right) dB_r^\lambda ds, \right. \right. \\
&\quad \left. \int_{[t]}^t \left\langle \nabla H(\bar{\theta}_{[s]}^\lambda), \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s dB_r^\lambda \right\rangle dB_s^\lambda \right\rangle.
\end{aligned}$$

Recall that $(\mathcal{F}_t^\lambda)_{t \geq 0}$ is the completed natural filtration of $(B_t^\lambda)_{t \geq 0}$. Then, we note that, for any $i, j, k = 1, \dots, d$, it holds that

$$\begin{aligned}
& \mathbb{E} \left[\left(\int_{[t]}^t \int_{[s]}^s \left(H^{(i,j)}(\bar{\zeta}_{[r]}^{\lambda,n}) - H^{(i,j)}(\bar{\theta}_{[r]}^\lambda) \right) d(B_r^\lambda)^{(j)} ds \right) \right. \\
&\quad \left. \times \left(\int_{[t]}^t \left(\partial_{\theta^{(k)}} H^{(i,j)}(\bar{\theta}_{[s]}^\lambda) \int_{[s]}^s d(B_r^\lambda)^{(k)} \right) d(B_s^\lambda)^{(j)} \right) \right] \\
&= \mathbb{E} \left[\left(H^{(i,j)}(\bar{\zeta}_{[t]}^{\lambda,n}) - H^{(i,j)}(\bar{\theta}_{[t]}^\lambda) \right) \partial_{\theta^{(k)}} H^{(i,j)}(\bar{\theta}_{[t]}^\lambda) \right. \\
&\quad \left. \times \mathbb{E} \left[\int_{[t]}^t \int_{[s]}^s d(B_r^\lambda)^{(j)} ds \int_{[t]}^t \int_{[s]}^s d(B_r^\lambda)^{(k)} d(B_s^\lambda)^{(j)} \middle| \mathcal{F}_{[t]}^\lambda \right] \right] \\
&= \mathbb{E} \left[\left(H^{(i,j)}(\bar{\zeta}_{[t]}^{\lambda,n}) - H^{(i,j)}(\bar{\theta}_{[t]}^\lambda) \right) \partial_{\theta^{(k)}} H^{(i,j)}(\bar{\theta}_{[t]}^\lambda) \right. \\
&\quad \left. \times \mathbb{E} \left[\int_{[t]}^t (t-s) d(B_s^\lambda)^{(j)} \int_{[t]}^t \int_{[s]}^s d(B_r^\lambda)^{(k)} d(B_s^\lambda)^{(j)} \middle| \mathcal{F}_{[t]}^\lambda \right] \right] \\
&= \mathbb{E} \left[\left(H^{(i,j)}(\bar{\zeta}_{[t]}^{\lambda,n}) - H^{(i,j)}(\bar{\theta}_{[t]}^\lambda) \right) \partial_{\theta^{(k)}} H^{(i,j)}(\bar{\theta}_{[t]}^\lambda) \right]
\end{aligned}$$

$$\begin{aligned}
& \times \left(t \int_{[t]}^t \mathbb{E} \left[\int_{[t]}^s d(B_r^\lambda)^{(k)} \middle| \mathcal{F}_{[t]}^\lambda \right] ds - \int_{[t]}^t s \mathbb{E} \left[\int_{[t]}^s d(B_r^\lambda)^{(k)} \middle| \mathcal{F}_{[t]}^\lambda \right] ds \right) \\
& = 0.
\end{aligned}$$

This implies that the last term in (97) is zero. Indeed, we have that

$$\begin{aligned}
& 2\lambda\beta^{-1} \mathbb{E} \left[\left\langle \int_{[t]}^t \int_{[s]}^s \left(H(\bar{\zeta}_{[r]}^{\lambda,n}) - H(\bar{\theta}_{[r]}^\lambda) \right) dB_r^\lambda ds, \right. \right. \\
& \quad \left. \left. \int_{[t]}^t \left\langle \nabla H(\bar{\theta}_{[s]}^\lambda), \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s dB_r^\lambda \right\rangle dB_s^\lambda \right\rangle \right] \\
& = (2\lambda\beta^{-1})^{3/2} \mathbb{E} \left[\sum_{i=1}^d \left(\int_{[t]}^t \sum_{j=1}^d \int_{[s]}^s \left(H^{(i,j)}(\bar{\zeta}_{[r]}^{\lambda,n}) - H^{(i,j)}(\bar{\theta}_{[r]}^\lambda) \right) d(B_r^\lambda)^{(j)} ds \right) \right. \\
& \quad \left. \times \left(\sum_{j=1}^d \int_{[t]}^t \left(\sum_{k=1}^d \partial_{\theta^{(k)}} H^{(i,j)}(\bar{\theta}_{[s]}^\lambda) \int_{[s]}^s d(B_r^\lambda)^{(k)} \right) d(B_s^\lambda)^{(j)} \right) \right] = 0.
\end{aligned}$$

Then, by using Remark 2.2 and (97) with the result above, we obtain that

$$\begin{aligned}
& \mathbb{E} \left[2\lambda\beta^{-1} \left\langle \int_{[t]}^t \int_{[s]}^s \left(H(\bar{\zeta}_{[r]}^{\lambda,n}) - H(\bar{\theta}_{[r]}^\lambda) \right) dB_r^\lambda ds, \int_{[t]}^t \mathfrak{M}(\bar{\theta}_s^\lambda, \bar{\theta}_{[s]}^\lambda) dB_s^\lambda \right\rangle \right] \\
& \leq 4\lambda^2\beta^{-2} \mathbb{E} \left[\left| \int_{[t]}^t \int_{[s]}^s \left(H(\bar{\zeta}_{[r]}^{\lambda,n}) - H(\bar{\theta}_{[r]}^\lambda) \right) dB_r^\lambda ds \right|^2 \right] \\
& \quad + \lambda^2 \mathbb{E} \left[\left| \int_{[t]}^t \left\langle \nabla H(\bar{\theta}_{[s]}^\lambda), \int_{[s]}^s -h_\lambda(\bar{\theta}_{[r]}^\lambda) dr \right\rangle dB_s^\lambda \right|^2 \right] \\
& \quad + \lambda^4 \mathbb{E} \left[\left| \int_{[t]}^t \left\langle \nabla H(\bar{\theta}_{[s]}^\lambda), \int_{[s]}^s \int_{[r]}^r H_\lambda(\bar{\theta}_{[\nu]}^\lambda) h_\lambda(\bar{\theta}_{[\nu]}^\lambda) d\nu dr \right\rangle dB_s^\lambda \right|^2 \right] \\
& \quad + \lambda^4\beta^{-2} \mathbb{E} \left[\left| \int_{[t]}^t \left\langle \nabla H(\bar{\theta}_{[s]}^\lambda), \int_{[s]}^s \int_{[r]}^r -\Upsilon_\lambda(\bar{\theta}_{[\nu]}^\lambda) d\nu dr \right\rangle dB_s^\lambda \right|^2 \right] \\
& \quad + 2\lambda^3\beta^{-1} \mathbb{E} \left[\left| \int_{[t]}^t \left\langle \nabla H(\bar{\theta}_{[s]}^\lambda), \int_{[s]}^s - \int_{[r]}^r H_\lambda(\bar{\theta}_{[\nu]}^\lambda) dB_\nu^\lambda dr \right\rangle dB_s^\lambda \right|^2 \right] \\
& \leq 8\lambda^2\beta^{-2} \int_{[t]}^t \int_{[s]}^s \mathbb{E} \left[|H(\bar{\zeta}_{[r]}^{\lambda,n})|_{\mathbb{F}}^2 + |H(\bar{\theta}_{[r]}^\lambda)|_{\mathbb{F}}^2 \right] dr ds \\
& \quad + \lambda^2 \int_{[t]}^t \mathbb{E} \left[\sum_{i,j=1}^d \left| \left\langle \nabla H^{(i,j)}(\bar{\theta}_{[s]}^\lambda), h_\lambda(\bar{\theta}_{[s]}^\lambda) \right\rangle \right|^2 \right] ds \\
& \quad + \lambda^4 \int_{[t]}^t \mathbb{E} \left[\sum_{i,j=1}^d \left| \left\langle \nabla H^{(i,j)}(\bar{\theta}_{[s]}^\lambda), H_\lambda(\bar{\theta}_{[s]}^\lambda) h_\lambda(\bar{\theta}_{[s]}^\lambda) \right\rangle \right|^2 \right] ds \\
& \quad + \lambda^4\beta^{-2} \int_{[t]}^t \mathbb{E} \left[\sum_{i,j=1}^d \left| \left\langle \nabla H^{(i,j)}(\bar{\theta}_{[s]}^\lambda), \Upsilon_\lambda(\bar{\theta}_{[s]}^\lambda) \right\rangle \right|^2 \right] ds \\
& \quad + 2\lambda^3\beta^{-1} \int_{[t]}^t \mathbb{E} \left[\sum_{i,j=1}^d \left| \left\langle \nabla H^{(i,j)}(\bar{\theta}_{[s]}^\lambda), H_\lambda(\bar{\theta}_{[s]}^\lambda) \int_{[s]}^s \int_{[r]}^r dB_\nu^\lambda dr \right\rangle \right|^2 \right] ds
\end{aligned}$$

$$\begin{aligned}
&\leq 8\lambda^2\beta^{-2} \int_{[t]}^t \mathbb{E} \left[dK_H^2(1 + |\bar{\zeta}_{[s]}^{\lambda,n}|^{\rho+q-1})^2 + dK_H^2(1 + |\bar{\theta}_{[s]}^\lambda|^{\rho+q-1})^2 \right] ds \\
&\quad + \lambda^2 \int_{[t]}^t \mathbb{E} \left[d^2\mathsf{K}_0^2(1 + |\bar{\theta}_{[s]}^\lambda|)^{2(\rho+q-2)} K_h^2(1 + |\bar{\theta}_{[s]}^\lambda|^{\rho+q})^2 \right] ds \\
&\quad + \lambda^4 \int_{[t]}^t \mathbb{E} \left[d^2\mathsf{K}_0^2(1 + |\bar{\theta}_{[s]}^\lambda|)^{2(\rho+q-2)} K_H^2(1 + |\bar{\theta}_{[s]}^\lambda|^{\rho+q-1})^2 K_h^2(1 + |\bar{\theta}_{[s]}^\lambda|^{\rho+q})^2 \right] ds \\
&\quad + \lambda^4\beta^{-2} \int_{[t]}^t \mathbb{E} \left[d^2\mathsf{K}_0^2(1 + |\bar{\theta}_{[s]}^\lambda|)^{2(\rho+q-2)} \mathsf{K}_{3,d}^2(1 + |\bar{\theta}_{[s]}^\lambda|)^{2(\rho+q-2)} \right] ds \\
&\quad + 2\lambda^3\beta^{-1} \int_{[t]}^t \mathbb{E} \left[d^2\mathsf{K}_0^2(1 + |\bar{\theta}_{[s]}^\lambda|)^{2(\rho+q-2)} K_H^2(1 + |\bar{\theta}_{[s]}^\lambda|^{\rho+q-1})^2 \left| \int_{[s]}^s \int_{[r]}^r dB_\nu^\lambda dr \right|^2 \right] ds \\
&\leq 6\lambda^2\beta^{-2} dK_H^2 3^{16(\rho+1)} \int_{[t]}^t \mathbb{E} \left[1 + |\bar{\zeta}_{[s]}^{\lambda,n}|^{16(\rho+1)} + |\bar{\theta}_{[s]}^\lambda|^{16(\rho+1)} \right] ds \\
&\quad + \lambda^2(K_h^2 + K_H^2 K_h^2 + \beta^{-2} \mathsf{K}_{3,d}^2) d^2\mathsf{K}_0^2 2^{16(\rho+1)} \int_{[t]}^t \mathbb{E} \left[1 + |\bar{\theta}_{[s]}^\lambda|^{16(\rho+1)} \right] ds \\
&\quad + \lambda^2\beta^{-1} d^2\mathsf{K}_0^2 K_H^2 2^{8(\rho+1)+1} \int_{[t]}^t \left(\mathbb{E} \left[1 + |\bar{\theta}_{[s]}^\lambda|^{16(\rho+1)} \right] \right)^{1/2} (3(d+4)) ds \\
&\leq 6\lambda^2\beta^{-2} dK_H^2 3^{24(\rho+1)} \left(e^{-\lambda a_D \min\{\kappa, 1/2\} [t]} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + (c_{8(\rho+1)} (1 + 1/(a_D \kappa)) + 1) \right. \\
&\quad \left. + 3v_{16(\rho+1)}(M_V(16(\rho+1))) \right) \\
&\quad + \lambda^2(K_h^2 + K_H^2 K_h^2 + \beta^{-2} \mathsf{K}_{3,d}^2 + \beta^{-1} K_H^2) d^3\mathsf{K}_0^2 2^{16(\rho+1)} \\
&\quad \times \left(e^{-\lambda a_D \kappa [t]} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + c_{8(\rho+1)} (1 + 1/(a_D \kappa)) + 1 \right) \\
&\leq \lambda^2 \left(e^{-a_D \min\{\kappa, 1/2\} n/2} 10\bar{\mathbf{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + 10\tilde{\mathbf{C}}_{\mathbf{A}2} \right),
\end{aligned}$$

where the fifth inequality holds due to Lemma 6.3, 6.5 and A.2, and $\bar{\mathbf{C}}_{\mathbf{A}2}$, $\tilde{\mathbf{C}}_{\mathbf{A}2}$ are given in (90).

This completes the proof. \square

Proof of Lemma 6.6. By using the definitions of $\bar{\theta}_t^\lambda$ in (29) and $\bar{\zeta}_t^{\lambda,n}$ in Definition 6.2, and by applying Itô's formula, we obtain, for any $n \in \mathbb{N}_0$, $t \in (nT, (n+1)T]$,

$$\begin{aligned}
&W_2^2(\mathcal{L}(\bar{\theta}_t^\lambda), \mathcal{L}(\bar{\zeta}_t^{\lambda,n})) \\
&\leq \mathbb{E} \left[\left| \bar{\theta}_t^\lambda - \bar{\zeta}_t^{\lambda,n} \right|^2 \right] \\
&= -2\lambda \mathbb{E} \left[\int_{nT}^t \left\langle \bar{\theta}_s^\lambda - \bar{\zeta}_s^{\lambda,n}, h_\lambda(\bar{\theta}_{[s]}^\lambda) - h(\bar{\zeta}_s^{\lambda,n}) \right\rangle ds \right] \\
&\quad + 2\lambda \mathbb{E} \left[\int_{nT}^t \left\langle \bar{\theta}_s^\lambda - \bar{\zeta}_s^{\lambda,n}, \lambda \int_{[s]}^s \left(H_\lambda(\bar{\theta}_{[r]}^\lambda) h_\lambda(\bar{\theta}_{[r]}^\lambda) - \beta^{-1} \Upsilon_\lambda(\bar{\theta}_{[r]}^\lambda) \right) dr \right\rangle ds \right] \\
&\quad - 2\lambda \mathbb{E} \left[\int_{nT}^t \left\langle \bar{\theta}_s^\lambda - \bar{\zeta}_s^{\lambda,n}, \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s H_\lambda(\bar{\theta}_{[r]}^\lambda) dB_r^\lambda \right\rangle ds \right]
\end{aligned}$$

$$\begin{aligned}
&= -2\lambda\mathbb{E} \left[\int_{nT}^t \left\langle \bar{\theta}_s^\lambda - \bar{\zeta}_s^{\lambda,n}, h(\bar{\theta}_s^\lambda) - h(\bar{\zeta}_s^{\lambda,n}) \right\rangle ds \right] \\
&\quad - 2\lambda\mathbb{E} \left[\int_{nT}^t \left\langle \bar{\theta}_s^\lambda - \bar{\zeta}_s^{\lambda,n}, h(\bar{\theta}_{[s]}^\lambda) - h(\bar{\theta}_s^\lambda) \right\rangle ds \right] \\
&\quad - 2\lambda\mathbb{E} \left[\int_{nT}^t \left\langle \bar{\theta}_s^\lambda - \bar{\zeta}_s^{\lambda,n}, h_\lambda(\bar{\theta}_{[s]}^\lambda) - h(\bar{\theta}_{[s]}^\lambda) \right\rangle ds \right] \\
&\quad + 2\lambda\mathbb{E} \left[\int_{nT}^t \left\langle \bar{\theta}_s^\lambda - \bar{\zeta}_s^{\lambda,n}, \lambda \int_{[s]}^s \left(H_\lambda(\bar{\theta}_{[r]}^\lambda) h_\lambda(\bar{\theta}_{[r]}^\lambda) - \beta^{-1} \Upsilon_\lambda(\bar{\theta}_{[r]}^\lambda) \right) dr \right\rangle ds \right] \\
&\quad - 2\lambda\mathbb{E} \left[\int_{nT}^t \left\langle \bar{\theta}_s^\lambda - \bar{\zeta}_s^{\lambda,n}, \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s H_\lambda(\bar{\theta}_{[r]}^\lambda) dB_r^\lambda \right\rangle ds \right].
\end{aligned} \tag{98}$$

By applying Itô's formula to $h(\bar{\theta}_s^\lambda)$, we obtain (91). Substituting (91) into (98), applying Remark 2.5 and Young's inequality yield

$$\begin{aligned}
&\mathbb{E} \left[\left| \bar{\theta}_t^\lambda - \bar{\zeta}_t^{\lambda,n} \right|^2 \right] \\
&\leq 2\lambda\text{Los} \int_{nT}^t \mathbb{E} \left[\left| \bar{\theta}_s^\lambda - \bar{\zeta}_s^{\lambda,n} \right|^2 \right] ds \\
&\quad + 2\lambda\mathbb{E} \left[\int_{nT}^t \left\langle \bar{\theta}_s^\lambda - \bar{\zeta}_s^{\lambda,n}, -\lambda \int_{[s]}^s \left(H(\bar{\theta}_r^\lambda) - H_\lambda(\bar{\theta}_{[r]}^\lambda) \right) h_\lambda(\bar{\theta}_{[r]}^\lambda) dr \right\rangle ds \right] \\
&\quad + 2\lambda\mathbb{E} \left[\int_{nT}^t \left\langle \bar{\theta}_s^\lambda - \bar{\zeta}_s^{\lambda,n}, \lambda^2 \int_{[s]}^s H(\bar{\theta}_r^\lambda) \int_{[r]}^r H_\lambda(\bar{\theta}_{[\nu]}^\lambda) h_\lambda(\bar{\theta}_{[\nu]}^\lambda) d\nu dr \right\rangle ds \right] \\
&\quad + 2\lambda\mathbb{E} \left[\int_{nT}^t \left\langle \bar{\theta}_s^\lambda - \bar{\zeta}_s^{\lambda,n}, -\lambda^2\beta^{-1} \int_{[s]}^s H(\bar{\theta}_r^\lambda) \int_{[r]}^r \Upsilon_\lambda(\bar{\theta}_{[\nu]}^\lambda) d\nu dr \right\rangle ds \right] \\
&\quad + 2\lambda\mathbb{E} \left[\int_{nT}^t \left\langle \bar{\theta}_s^\lambda - \bar{\zeta}_s^{\lambda,n}, -\lambda\sqrt{2\lambda\beta^{-1}} \int_{[s]}^s H(\bar{\theta}_r^\lambda) \int_{[r]}^r H_\lambda(\bar{\theta}_{[\nu]}^\lambda) dB_\nu^\lambda dr \right\rangle ds \right] \\
&\quad + 2\lambda\mathbb{E} \left[\int_{nT}^t \left\langle \bar{\theta}_s^\lambda - \bar{\zeta}_s^{\lambda,n}, \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s \left(H(\bar{\theta}_r^\lambda) - H_\lambda(\bar{\theta}_{[r]}^\lambda) \right) dB_r^\lambda \right\rangle ds \right] \\
&\quad + 2\lambda\mathbb{E} \left[\int_{nT}^t \left\langle \bar{\theta}_s^\lambda - \bar{\zeta}_s^{\lambda,n}, \lambda\beta^{-1} \int_{[s]}^s \left(\Upsilon(\bar{\theta}_r^\lambda) - \Upsilon_\lambda(\bar{\theta}_{[r]}^\lambda) \right) dr \right\rangle ds \right] \\
&\quad - 2\lambda\mathbb{E} \left[\int_{nT}^t \left\langle \bar{\theta}_s^\lambda - \bar{\zeta}_s^{\lambda,n}, h_\lambda(\bar{\theta}_{[s]}^\lambda) - h(\bar{\theta}_{[s]}^\lambda) \right\rangle ds \right]
\end{aligned}$$

$$\begin{aligned}
&\leq \lambda(2\text{L}_{\text{OS}} + 6) \int_{nT}^t \mathbb{E} \left[|\bar{\theta}_s^\lambda - \bar{\zeta}_s^{\lambda,n}|^2 \right] ds \\
&\quad + \lambda \int_{nT}^t \mathbb{E} \left[\left| -\lambda \int_{[s]}^r \left(H(\bar{\theta}_r^\lambda) - H_\lambda(\bar{\theta}_{[r]}) \right) h_\lambda(\bar{\theta}_{[r]}) dr \right|^2 \right] ds \\
&\quad + \lambda \int_{nT}^t \mathbb{E} \left[\left| \lambda^2 \int_{[s]}^r H(\bar{\theta}_r^\lambda) \int_{[r]}^r H_\lambda(\bar{\theta}_{[\nu]}) h_\lambda(\bar{\theta}_{[\nu]}) d\nu dr \right|^2 \right] ds \\
&\quad + \lambda \int_{nT}^t \mathbb{E} \left[\left| -\lambda^2 \beta^{-1} \int_{[s]}^r H(\bar{\theta}_r^\lambda) \int_{[r]}^r \Upsilon_\lambda(\bar{\theta}_{[\nu]}) d\nu dr \right|^2 \right] ds \\
&\quad + \lambda \int_{nT}^t \mathbb{E} \left[\left| -\lambda \sqrt{2\lambda\beta^{-1}} \int_{[s]}^r H(\bar{\theta}_r^\lambda) \int_{[r]}^r H_\lambda(\bar{\theta}_{[\nu]}) dB_\nu^\lambda dr \right|^2 \right] ds \\
&\quad + \lambda \int_{nT}^t \mathbb{E} \left[\left| \lambda \beta^{-1} \int_{[s]}^r \left(\Upsilon(\bar{\theta}_r^\lambda) - \Upsilon_\lambda(\bar{\theta}_{[r]}) \right) dr \right|^2 \right] ds \\
&\quad + \lambda \int_{nT}^t \mathbb{E} \left[\left| h(\bar{\theta}_{[s]}) - h_\lambda(\bar{\theta}_{[s]}) \right|^2 \right] ds \\
&\quad + 2\lambda \mathbb{E} \left[\int_{nT}^t \left\langle \bar{\theta}_s^\lambda - \bar{\zeta}_s^{\lambda,n}, \sqrt{2\lambda\beta^{-1}} \int_{[s]}^r \left(H(\bar{\theta}_r^\lambda) - H_\lambda(\bar{\theta}_{[r]}) \right) dB_r^\lambda \right\rangle ds \right].
\end{aligned} \tag{99}$$

We note that, by using (89), for any $t \geq nT$,

$$\begin{aligned}
\mathbb{E} \left[|h(\bar{\theta}_{[t]}) - h_\lambda(\bar{\theta}_{[t]})|^2 \right] &\leq \mathbb{E} \left[\lambda^3 |\bar{\theta}_{[t]}^\lambda|^{6(\rho+q-1)} K_h^2 (1 + |\bar{\theta}_{[t]}^\lambda|^{\rho+q})^2 \right] \\
&\leq \lambda^3 K_h^2 2^{16(\rho+1)} \mathbb{E} \left[1 + |\bar{\theta}_{[t]}^\lambda|^{16(\rho+1)} \right] \\
&\leq \lambda^3 K_h^2 2^{16(\rho+1)} \left(e^{-\lambda \mathfrak{a}_D \kappa [t]} \mathbb{E} \left[|\theta_0|^{16(\rho+1)} \right] + \mathfrak{c}_{8(\rho+1)} (1 + 1/(\mathfrak{a}_D \kappa)) + 1 \right) \\
&\leq \lambda^3 \left(e^{-\mathfrak{a}_D \kappa n/2} \bar{\mathfrak{C}}_{\mathbf{A}2} \mathbb{E} \left[|\theta_0|^{16(\rho+1)} \right] + \tilde{\mathfrak{C}}_{\mathbf{A}2} \right),
\end{aligned} \tag{100}$$

where the third inequality holds due to Lemma 6.3, and $\bar{\mathfrak{C}}_{\mathbf{A}2}$, $\tilde{\mathfrak{C}}_{\mathbf{A}2}$ are given in (90). By using Lemma A.3 and (100), (99) becomes

$$\begin{aligned}
\mathbb{E} \left[|\bar{\theta}_t^\lambda - \bar{\zeta}_t^{\lambda,n}|^2 \right] &\leq \lambda(2\text{L}_{\text{OS}} + 6) \int_{nT}^t \mathbb{E} \left[|\bar{\theta}_s^\lambda - \bar{\zeta}_s^{\lambda,n}|^2 \right] ds \\
&\quad + 6\lambda^{2+q} \left(e^{-\mathfrak{a}_D \kappa n/2} \bar{\mathfrak{C}}_{\mathbf{A}2} \mathbb{E} \left[|\theta_0|^{16(\rho+1)} \right] + \tilde{\mathfrak{C}}_{\mathbf{A}2} \right) \\
&\quad + \mathfrak{J}_1 + \mathfrak{J}_2,
\end{aligned} \tag{101}$$

where

$$\begin{aligned}
\mathfrak{J}_1^\lambda(t) &:= 2\lambda \mathbb{E} \left[\int_{nT}^t \left\langle \bar{\theta}_s^\lambda - \bar{\zeta}_s^{\lambda,n}, \sqrt{2\lambda\beta^{-1}} \int_{[s]}^r \left(H(\bar{\theta}_r^\lambda) - H_\lambda(\bar{\theta}_{[r]}) - \mathfrak{M}(\bar{\theta}_r^\lambda, \bar{\theta}_{[r]}) \right) dB_r^\lambda \right\rangle ds \right], \\
\mathfrak{J}_2^\lambda(t) &:= 2\lambda \mathbb{E} \left[\int_{nT}^t \left\langle \bar{\theta}_s^\lambda - \bar{\zeta}_s^{\lambda,n}, \sqrt{2\lambda\beta^{-1}} \int_{[s]}^r \mathfrak{M}(\bar{\theta}_r^\lambda, \bar{\theta}_{[r]}) dB_r^\lambda \right\rangle ds \right]
\end{aligned}$$

with \mathfrak{M} defined in Definition A.5. By using Young's inequality and Lemma A.6, we have that

$$\mathfrak{J}_1^\lambda(t) \leq \lambda \int_{nT}^t \mathbb{E} \left[|\bar{\theta}_s^\lambda - \bar{\zeta}_s^{\lambda,n}|^2 \right] ds + \lambda^{2+q} \left(e^{-\mathfrak{a}_D \kappa n/2} \bar{\mathfrak{C}}_{\mathbf{A}2} \mathbb{E} \left[|\theta_0|^{16(\rho+1)} \right] + \tilde{\mathfrak{C}}_{\mathbf{A}2} \right). \tag{102}$$

To establish an upper bound for $\mathfrak{J}_2^\lambda(t)$, we recall the definitions of $(\bar{\theta}_t^\lambda)_{t \geq 0}$ and $(\bar{\zeta}_t^{\lambda,n})_{t \geq 0}$ given in (29) and Definition 6.2, respectively, and consider the following splitting:

$$\begin{aligned}
\mathfrak{J}_2^\lambda(t) &= 2\lambda \int_{nT}^t \mathbb{E} \left[\left\langle \bar{\theta}_s^\lambda - \bar{\theta}_{[s]}^\lambda - (\bar{\zeta}_s^{\lambda,n} - \bar{\zeta}_{[s]}^{\lambda,n}), \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s \mathfrak{M}(\bar{\theta}_r^\lambda, \bar{\theta}_{[r]}^\lambda) dB_r^\lambda \right\rangle \right] ds \\
&= 2\lambda \int_{nT}^t \mathbb{E} \left[\left\langle \int_{[s]}^s \left(\lambda(h(\bar{\zeta}_r^{\lambda,n}) - h_\lambda(\bar{\theta}_{[r]}^\lambda)) + \lambda^2 \int_{[r]}^r (H_\lambda(\bar{\theta}_{[v]}^\lambda)h_\lambda(\bar{\theta}_{[v]}^\lambda) - \beta^{-1}\Upsilon_\lambda(\bar{\theta}_{[v]}^\lambda)) \, dv \right. \right. \\
&\quad \left. \left. - \lambda\sqrt{2\lambda\beta^{-1}} \int_{[r]}^r H_\lambda(\bar{\theta}_{[v]}^\lambda) dB_v^\lambda \right) dr, \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s \mathfrak{M}(\bar{\theta}_r^\lambda, \bar{\theta}_{[r]}^\lambda) dB_r^\lambda \right\rangle \right] ds \\
&= 2\lambda \int_{nT}^t \mathbb{E} \left[\left\langle \int_{[s]}^s \left(\lambda(h(\bar{\theta}_r^\lambda) - h(\bar{\theta}_{[r]}^\lambda)) + \lambda^2 \int_{[r]}^r (H_\lambda(\bar{\theta}_{[v]}^\lambda)h_\lambda(\bar{\theta}_{[v]}^\lambda) - \beta^{-1}\Upsilon_\lambda(\bar{\theta}_{[v]}^\lambda)) \, dv \right. \right. \\
&\quad \left. \left. - \lambda\sqrt{2\lambda\beta^{-1}} \int_{[r]}^r H_\lambda(\bar{\theta}_{[v]}^\lambda) dB_v^\lambda \right) dr, \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s \mathfrak{M}(\bar{\theta}_r^\lambda, \bar{\theta}_{[r]}^\lambda) dB_r^\lambda \right\rangle \right] ds \\
&\quad + 2\lambda \int_{nT}^t \mathbb{E} \left[\left\langle \int_{[s]}^s \lambda(h(\bar{\theta}_{[r]}^\lambda) - h_\lambda(\bar{\theta}_{[r]}^\lambda)) dr, \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s \mathfrak{M}(\bar{\theta}_r^\lambda, \bar{\theta}_{[r]}^\lambda) dB_r^\lambda \right\rangle \right] ds + \mathfrak{J}_{2,1}^\lambda(t),
\end{aligned}$$

where the first equality holds due to the following:

$$\begin{aligned}
&\mathbb{E} \left[\left\langle \bar{\theta}_{[s]}^\lambda + \bar{\zeta}_{[s]}^{\lambda,n}, \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s \mathfrak{M}(\bar{\theta}_r^\lambda, \bar{\theta}_{[r]}^\lambda) dB_r^\lambda \right\rangle \right] \\
&= \mathbb{E} \left[\left\langle \bar{\theta}_{[s]}^\lambda + \bar{\zeta}_{[s]}^{\lambda,n}, \mathbb{E} \left[\sqrt{2\lambda\beta^{-1}} \int_{[s]}^s \mathfrak{M}(\bar{\theta}_r^\lambda, \bar{\theta}_{[r]}^\lambda) dB_r^\lambda \middle| \mathcal{F}_{[s]}^\lambda \right] \right\rangle \right] = 0,
\end{aligned} \tag{103}$$

and where

$$\mathfrak{J}_{2,1}^\lambda(t) = 2\lambda \int_{nT}^t \mathbb{E} \left[\left\langle \int_{[s]}^s \lambda(h(\bar{\zeta}_r^{\lambda,n}) - h(\bar{\theta}_r^\lambda)) dr, \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s \mathfrak{M}(\bar{\theta}_r^\lambda, \bar{\theta}_{[r]}^\lambda) dB_r^\lambda \right\rangle \right] ds.$$

By applying Cauchy-Schwarz inequality, Corollary A.4, Lemma A.6 and (100), we further obtain that

$$\begin{aligned}
\mathfrak{J}_2^\lambda(t) &\leq 2\lambda \int_{nT}^t \left(\mathbb{E} \left[\left| \int_{[s]}^s \lambda \left(h(\bar{\theta}_r^\lambda) - h(\bar{\theta}_{[r]}^\lambda) + \lambda \int_{[r]}^r (H_\lambda(\bar{\theta}_{[v]}^\lambda)h_\lambda(\bar{\theta}_{[v]}^\lambda) - \beta^{-1}\Upsilon_\lambda(\bar{\theta}_{[v]}^\lambda)) \, dv \right. \right. \right. \\
&\quad \left. \left. \left. - \sqrt{2\lambda\beta^{-1}} \int_{[r]}^r H_\lambda(\bar{\theta}_{[v]}^\lambda) dB_v^\lambda \right) dr \right|^2 \right] \right)^{1/2} \\
&\quad \times \left(\mathbb{E} \left[\left| \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s \mathfrak{M}(\bar{\theta}_r^\lambda, \bar{\theta}_{[r]}^\lambda) dB_r^\lambda \right|^2 \right] \right)^{1/2} ds \\
&\quad + 2\lambda \int_{nT}^t \left(\mathbb{E} \left[\left| \int_{[s]}^s \lambda(h(\bar{\theta}_{[r]}^\lambda) - h_\lambda(\bar{\theta}_{[r]}^\lambda)) dr \right|^2 \right] \right)^{1/2} \\
&\quad \times \left(\mathbb{E} \left[\left| \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s \mathfrak{M}(\bar{\theta}_r^\lambda, \bar{\theta}_{[r]}^\lambda) dB_r^\lambda \right|^2 \right] \right)^{1/2} ds + \mathfrak{J}_{2,1}^\lambda(t) \\
&\leq 2\lambda \int_{nT}^t \left(\lambda^4 \left(e^{-\mathfrak{a}_D \kappa n/2} 36 \bar{\mathfrak{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + 36 \tilde{\mathfrak{C}}_{\mathbf{A}2} \right) \right)^{1/2} \\
&\quad \times \left(\lambda^2 \left(e^{-\mathfrak{a}_D \kappa n/2} \bar{\mathfrak{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + \tilde{\mathfrak{C}}_{\mathbf{A}2} \right) \right)^{1/2} ds \\
&\quad + 2\lambda \int_{nT}^t \left(\lambda^5 \left(e^{-\mathfrak{a}_D \kappa n/2} \bar{\mathfrak{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + \tilde{\mathfrak{C}}_{\mathbf{A}2} \right) \right)^{1/2}
\end{aligned}$$

$$\begin{aligned}
& \times \left(\lambda^2 \left(e^{-\text{aD}\kappa n/2} \bar{\mathbf{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + \tilde{\mathbf{C}}_{\mathbf{A}2} \right) \right)^{1/2} ds + \mathfrak{J}_{2,1}^\lambda(t) \\
& \leq 14\lambda^3 \left(e^{-\text{aD}\kappa n/2} \bar{\mathbf{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + \tilde{\mathbf{C}}_{\mathbf{A}2} \right) + \mathfrak{J}_{2,1}^\lambda(t). \tag{104}
\end{aligned}$$

At this stage, the task reduces to upper bound $\mathfrak{J}_{2,1}^\lambda(t)$. To achieve this, we apply Cauchy-Schwarz inequality, Corollary A.4 and Lemma A.6 to obtain

$$\begin{aligned}
\mathfrak{J}_{2,1}^\lambda(t) &= 2\lambda \int_{nT}^t \mathbb{E} \left[\left\langle \int_{[s]}^s \lambda \left(h(\bar{\zeta}_r^{\lambda,n}) - h(\bar{\zeta}_{[r]}^{\lambda,n}) - \sqrt{2\lambda\beta^{-1}} \int_{[r]}^r H(\bar{\zeta}_{[v]}^{\lambda,n}) dB_\nu^\lambda - \left(h(\bar{\theta}_r^\lambda) \right. \right. \right. \\
&\quad \left. \left. \left. - h(\bar{\theta}_{[r]}^\lambda) - \sqrt{2\lambda\beta^{-1}} \int_{[r]}^r H(\bar{\theta}_{[v]}^\lambda) dB_\nu^\lambda \right) dr, \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s \mathfrak{M}(\bar{\theta}_r^\lambda, \bar{\theta}_{[r]}^\lambda) dB_r^\lambda \right\rangle \right] ds \\
&\quad + 2\lambda \int_{nT}^t \mathbb{E} \left[\left\langle \int_{[s]}^s \lambda \sqrt{2\lambda\beta^{-1}} \int_{[r]}^r \left(H(\bar{\zeta}_{[v]}^{\lambda,n}) - H(\bar{\theta}_{[v]}^\lambda) \right) dB_\nu^\lambda dr, \right. \right. \\
&\quad \left. \left. \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s \mathfrak{M}(\bar{\theta}_r^\lambda, \bar{\theta}_{[r]}^\lambda) dB_r^\lambda \right\rangle \right] ds \\
&\leq 2\lambda \int_{nT}^t \left(\mathbb{E} \left[\left| \int_{[s]}^s \lambda \left(h(\bar{\zeta}_r^{\lambda,n}) - h(\bar{\zeta}_{[r]}^{\lambda,n}) - \sqrt{2\lambda\beta^{-1}} \int_{[r]}^r H(\bar{\zeta}_{[v]}^{\lambda,n}) dB_\nu^\lambda - \left(h(\bar{\theta}_r^\lambda) \right. \right. \right. \right. \right. \\
&\quad \left. \left. \left. - h(\bar{\theta}_{[r]}^\lambda) - \sqrt{2\lambda\beta^{-1}} \int_{[r]}^r H(\bar{\theta}_{[v]}^\lambda) dB_\nu^\lambda \right) dr \right|^2 \right] \right)^{1/2} \\
&\quad \times \left(\mathbb{E} \left[\left| \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s \mathfrak{M}(\bar{\theta}_r^\lambda, \bar{\theta}_{[r]}^\lambda) dB_r^\lambda \right|^2 \right] \right)^{1/2} ds \\
&\quad + 2\lambda^3 \left(e^{-\text{aD} \min\{\kappa, 1/2\} n/2} 10 \bar{\mathbf{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + 10 \tilde{\mathbf{C}}_{\mathbf{A}2} \right) \\
&\leq 2\lambda \int_{nT}^t \left(\lambda^4 \left(e^{-\text{aD} \min\{\kappa, 1/2\} n/2} 72 \bar{\mathbf{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + 72 \tilde{\mathbf{C}}_{\mathbf{A}2} \right) \right)^{1/2} \\
&\quad \times \left(\lambda^2 \left(e^{-\text{aD}\kappa n/2} \bar{\mathbf{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + \tilde{\mathbf{C}}_{\mathbf{A}2} \right) \right)^{1/2} ds \\
&\quad + 20\lambda^3 \left(e^{-\text{aD} \min\{\kappa, 1/2\} n/2} \bar{\mathbf{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + \tilde{\mathbf{C}}_{\mathbf{A}2} \right) \\
&\leq 44\lambda^3 \left(e^{-\text{aD} \min\{\kappa, 1/2\} n/2} \bar{\mathbf{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + \tilde{\mathbf{C}}_{\mathbf{A}2} \right). \tag{105}
\end{aligned}$$

Substituting (105) into (104) yields

$$\mathfrak{J}_2^\lambda(t) \leq 58\lambda^3 \left(e^{-\text{aD} \min\{\kappa, 1/2\} n/2} \bar{\mathbf{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + \tilde{\mathbf{C}}_{\mathbf{A}2} \right). \tag{106}$$

By applying (102) and (106) to (101), we obtain that

$$\begin{aligned}
\mathbb{E} \left[\left| \bar{\theta}_t^\lambda - \bar{\zeta}_t^{\lambda,n} \right|^2 \right] &\leq \lambda(2L_{\text{OS}} + 7) \int_{nT}^t \mathbb{E} \left[\left| \bar{\theta}_s^\lambda - \bar{\zeta}_s^{\lambda,n} \right|^2 \right] ds \\
&\quad + 65\lambda^{2+q} \left(e^{-\text{aD} \min\{\kappa, 1/2\} n/2} \bar{\mathbf{C}}_{\mathbf{A}2} \mathbb{E}[|\theta_0|^{16(\rho+1)}] + \tilde{\mathbf{C}}_{\mathbf{A}2} \right),
\end{aligned}$$

which, by applying Grönwall's lemma, yields

$$W_2^2(\mathcal{L}(\bar{\theta}_t^\lambda), \mathcal{L}(\bar{\zeta}_t^{\lambda,n})) \leq \mathbb{E} \left[\left| \bar{\theta}_t^\lambda - \bar{\zeta}_t^{\lambda,n} \right|^2 \right] \leq \lambda^{2+q} \left(e^{-\text{aD} \min\{\kappa, 1/2\} n/2} \mathbf{C}_0 \mathbb{E}[|\theta_0|^{16(\rho+1)}] + \mathbf{C}_1 \right),$$

where κ is given in (75) and

$$\mathbf{C}_0 := 65e^{2L_{\text{OS}}+7} \bar{\mathbf{C}}_{\mathbf{A}2}, \quad \mathbf{C}_1 := 65e^{2L_{\text{OS}}+7} \tilde{\mathbf{C}}_{\mathbf{A}2} \tag{107}$$

with $\bar{\mathbf{C}}_{\mathbf{A}2}, \tilde{\mathbf{C}}_{\mathbf{A}2}$ given in (90). \square

APPENDIX B. ANALYTIC EXPRESSION OF CONSTANTS FOR AHOLA

CONSTANT	FULL EXPRESSION
Remark 2.2	K_0 $2^{1-q} \max\{L, \nabla^2 h^{(1)}(0) , \dots, \nabla^2 h^{(d)}(0) \}$ K_1 $\max\{K_0, \nabla h^{(1)}(0) , \dots, \nabla h^{(d)}(0) \}$ K_2 $\max\{K_1, h^{(1)}(0) , \dots, h^{(d)}(0) \}$ $K_{3,d}$ $\max\{d^{3/2}L, \Upsilon(0)\}$
Remark 2.4	a_D $a/2$ b_D $(a/2 + b)R_D^{\bar{r}+2} + h(0) ^2/(2a)$ R_D $\max\{(4b/a)^{1/(r-\bar{r})}, 2^{1/r}\}$ \bar{b}_D $a_D + b_D$
Remark 2.5	L_{os} $\sqrt{d}K_1(1 + 2R_{os})^{\rho+q-1}$ R_{os} $(b/a)^{1/(r-\bar{r})}$
Theorem 2.7	C_0 $\min\{\hat{c}, a_D\kappa, a_D/2\}/4$ C_1 $e^{\min\{\hat{c}, a_D\kappa, a_D/2\}/4} \left[C_0^{1/2} + C_2 + \hat{c} \left(3 + \int_{\mathbb{R}^d} V_2(\theta) \pi_\beta(d\theta) \right) \right]$ C_2 $C_1^{1/2} + C_3$
Corollary 2.8	C_3 $\min\{\hat{c}, a_D\kappa, a_D/2\}/8$ C_4 $e^{\min\{\hat{c}, a_D\kappa, a_D/2\}/8} \left[C_0^{1/2} + C_4 + \sqrt{2\hat{c}} \left(3 + \int_{\mathbb{R}^d} V_2(\theta) \pi_\beta(d\theta) \right)^{1/2} \right]$ C_5 $C_1^{1/2} + C_5$ C_4 $\sqrt{\hat{c}} \left(1 + \frac{8}{\min\{\hat{c}, a_D\kappa, a_D/2\}} \right) e^{\min\{\hat{c}, a_D\kappa, a_D/2\}/8} \left(\sqrt{2}C_0^{1/2} + 2^{4\rho+5} \right)$ C_5 $4(\sqrt{\hat{c}}/\hat{c})e^{\hat{c}/4}(\sqrt{2}C_1^{1/2} + 3\sqrt{2} + 2^{4\rho+5}(c_{8(\rho+1)}(1 + 1/(a_D\kappa)) + 1)^{1/2} + \sqrt{6}v_{16(\rho+1)}^{1/2}(M_V(16(\rho+1))))$
Lemma 6.3	κ $M_1^{\rho+q-1} / \left(1 + M_1^{3(\rho+q-1)} \right)^{1/3}$ (with $M_1 > 0$) c_0 $a_D\kappa M_1^2 + c_1 + 2^{2(\rho+q)}\beta^{-1}d(1 + K_1)^2$ c_1 $2b_D + 4K_h K_H + \beta^{-1}2^{\rho+q-1}K_{3,d} + 2K_h^2 K_H^2 + 3K_H^2 K_h^2 + 6K_h^2 K_H + 2^{2(\rho+q)-6}\beta^{-2}K_{3,d}^2 + \beta^{-1}K_h K_{3,d}2^{\rho+q} + \beta^{-1}K_H K_h K_{3,d}2^{\rho+q-1}$ c_p $(1 + 2/(a_D\kappa))^{p-1} (a_D\kappa M_1^2 + c_1)^p + c_\Xi(p) + 2^{2(p+\rho+q)-3}p(2p-1)\beta^{-1}d(1 + K_1)^2((1 + 2/(a_D\kappa))^{p-2} (a_D\kappa M_1^2 + c_1)^{p-1} + M_2(p)^{2p-2})$ $c_\Xi(p)$ $2^{2p-4}(\beta^{-1}d)^p(2p(2p-1))^{p+1}3^{p-1} \left(1 + 2^{p(\rho+q-1)}d^{p/2}K_1^p \right)^2$ $M_2(p)$ $(2^{2(p+\rho+q)-1}p(2p-1)\beta^{-1}d(1 + K_1)^2/(a_D\kappa))^{1/2}$
Lemma 6.4	$c_{V,1}(p)$ $a_D p/4$ $c_{V,2}(p)$ $(3/4)a_D p v_p(M_V(p))$ $M_V(p)$ $(1/3 + 4\bar{b}_D/(3a_D) + 4d/(3a_D\beta) + 4(p-2)/(3a_D\beta))^{1/2}$
Lemma 6.6	C_0 $65e^{2L_{os}+7}\bar{C}_{A2}$ C_1 $65e^{2L_{os}+7}\tilde{C}_{A2}$ \bar{C}_{A2} $3^{24(\rho+1)}d^3(1 + K_0)^2(1 + K_h)^2(1 + K_H)^4(1 + \beta^{-1})^2(1 + L)^2(1 + K_{3,d})^2(2 + \max\{\bar{C}_{A0,2}, \bar{C}_{A0,2q}, \bar{C}_{A1,2}, \bar{C}_{A0,2+2q}\})$ \tilde{C}_{A2} $3^{24(\rho+1)}d^3(1 + K_0)^2(1 + K_h)^2(1 + K_H)^4(1 + \beta^{-1})^2(1 + L)^2(1 + K_{3,d})^2(2c_{8(\rho+1)}(1 + 1/(a_D\kappa)) + 2 \max\{\tilde{C}_{A0,2}, \tilde{C}_{A0,2q}, \tilde{C}_{A1,2}, \tilde{C}_{A0,2+2q}\} + 1 + v_{16(\rho+1)}(M_V(16(\rho+1))))$ $\bar{C}_{A0,p}$ $5^{2p}2^{4[p](\rho+1)}(K_h^{2p} + \beta^{-2p}K_{3,d}^{2p} + K_H^{2p}K_h^{2p} + (\beta^{-1}(4p+2)(d+4p))^p(1 + K_H^{2p}))$ $\tilde{C}_{A0,p}$ $\bar{C}_{A0,p} (c_{2[p](\rho+1)}(1 + 1/(a_D\kappa)) + 1)$ $\bar{C}_{A1,p}$ $2^{2p+[p](\rho+1)-1}K_h^{2p}$ $\tilde{C}_{A1,p}$ $\bar{C}_{A1,p} (c_{[p](\rho+1)}(1 + 1/(a_D\kappa)) + 1) + 2^{2p}K_h^{2p}3v_{2[p](\rho+1)}(M_V(2[p](\rho+1)))) + 2^{2p}(2\beta^{-1}(p+1)(d+2p))^p$
Proposition 6.7	\dot{c} $\min\{\bar{\phi}, c_{V,1}(2), 4c_{V,2}(2)\epsilon c_{V,1}(2)\}/2$ $\bar{\phi}$ $\left(\sqrt{8\pi/(\beta L_{os})}\dot{c}_0 \exp \left(\left(\dot{c}_0 \sqrt{\beta L_{os}/8} + \sqrt{8/(\beta L_{os})} \right)^2 \right) \right)^{-1}$ ϵ $1 \wedge (4c_{V,2}(2)\sqrt{2\beta\pi/L_{os}} \int_0^{\dot{c}_1} \exp \left(\left(s\sqrt{\beta L_{os}/8} + \sqrt{8/(\beta L_{os})} \right)^2 \right) ds)^{-1}$ \dot{c}_0 $2(4c_{V,2}(2)(1 + c_{V,1}(2))/c_{V,1}(2) - 1)^{1/2}$ \dot{c}_1 $2(2c_{V,2}(2)/c_{V,1}(2) - 1)^{1/2}$ \hat{c} $2(1 + \dot{c}_0) \exp(\beta L_{os}\dot{c}_0^2/8 + 2\dot{c}_0)/\epsilon$
Lemma 6.8	C_2 $\hat{c} \left(1 + \frac{4}{\min\{\hat{c}, a_D\kappa, a_D/2\}} \right) e^{\min\{\hat{c}, a_D\kappa, a_D/2\}/4} (C_0 + 2^{8\rho+11})$ C_3 $2(\hat{c}/\dot{c})e^{\hat{c}/2}(C_1 + 15 + 2^{8\rho+11}(c_{8(\rho+1)}(1 + 1/(a_D\kappa)) + 1) + 18v_{16(\rho+1)}(M_V(16(\rho+1))))$

APPENDIX C. ANALYTIC EXPRESSION OF CONSTANTS FOR AHOLLA

CONSTANT	FULL EXPRESSION	
Remark 3.2	\bar{K}_0 \bar{K}_1	$\max\{\bar{L}_1, \nabla^2 h^{(1)}(0) , \dots, \nabla^2 h^{(d)}(0) \}$ $\max\{\bar{L}_3, h(0) \}$
Theorem 3.3	$C_{\text{Lin},0}$ $C_{\text{Lin},1}$ $C_{\text{Lin},2}$	$\min\{\dot{c}_{\text{Lin}}, \bar{a}/2\}/4$ $e^{\min\{\dot{c}_{\text{Lin}}, \bar{a}/2\}/4} \left[C_{\text{Lin},0}^{1/2} + C_{\text{Lin},2} + \dot{c}_{\text{Lin}} \left(3 + \int_{\mathbb{R}^d} V_2(\theta) \pi_\beta(d\theta) \right) \right]$ $C_{\text{Lin},1}^{1/2} + C_{\text{Lin},3}$
Corollary 3.4	$C_{\text{Lin},3}$ $C_{\text{Lin},4}$ $C_{\text{Lin},5}$ $C_{\text{Lin},4}$ $C_{\text{Lin},5}$	$\min\{\dot{c}_{\text{Lin}}, \bar{a}/2\}/8$ $e^{\min\{\dot{c}_{\text{Lin}}, \bar{a}/2\}/8} \left[C_{\text{Lin},0}^{1/2} + C_{\text{Lin},4} + \sqrt{2\dot{c}_{\text{Lin}}} \left(3 + \int_{\mathbb{R}^d} V_2(\theta) \pi_\beta(d\theta) \right)^{1/2} \right]$ $C_{\text{Lin},1}^{1/2} + C_{\text{Lin},5}$ $\sqrt{\dot{c}_{\text{Lin}}} \left(1 + \frac{8}{\min\{\dot{c}_{\text{Lin}}, \bar{a}/2\}} \right) e^{\min\{\dot{c}_{\text{Lin}}, \bar{a}/2\}/8} \left(C_{\text{Lin},0}^{1/2} + 3 \right)$ $4(\sqrt{\dot{c}_{\text{Lin}}}/\dot{c}_{\text{Lin}}) e^{\dot{c}_{\text{Lin}}/4} \left(C_{\text{Lin},1}^{1/2} + 1 + 3(\bar{c}_2(1 + 1/\bar{a}) + 1)^{1/2} + \sqrt{3}v_4^{1/2}(\bar{M}_V(4)) \right)$
Lemma D.3	\bar{c}_0 \bar{c}_1 \bar{M}_1 \bar{c}_p $\bar{c}_{\Xi}(p)$ $\bar{M}_2(p)$	$\beta^{-1}d(\bar{L}_2 + \bar{L}_2\bar{K}_1 + \bar{L}_2\bar{L}_3\bar{K}_1/2)\bar{M}_1 + \bar{c}_1 + 2\beta^{-1}d(1 + \bar{L}_3)^2$ $2\bar{b} + 2\bar{L}_3\bar{K}_1 + 2\bar{K}_1^2 + \bar{L}_3\bar{K}_1^2/2 + \beta^{-2}d^2\bar{L}_2^2/4 + 2\bar{L}_3\bar{K}_1^2 + \beta^{-1}d(\bar{L}_2\bar{K}_1 + \bar{L}_2\bar{L}_3\bar{K}_1/2)$ $2(\bar{a}\beta)^{-1}d(\bar{L}_2 + \bar{L}_2\bar{K}_1 + \bar{L}_2\bar{L}_3\bar{K}_1/2)$ $\bar{c}_{\Xi}(p) + (1 + 2/\bar{a})^{p-1}(\bar{c}_1 + \beta^{-1}d(\bar{L}_2 + \bar{L}_2\bar{K}_1 + \bar{L}_2\bar{L}_3\bar{K}_1/2)\bar{M}_1)^p + 2^{2p-2}p(2p - 1)\beta^{-1}d(1 + \bar{L}_3)^2((1 + 2/\bar{a})^{p-2}(\bar{c}_1 + \beta^{-1}d(\bar{L}_2 + \bar{L}_2\bar{K}_1 + \bar{L}_2\bar{L}_3\bar{K}_1/2)\bar{M}_1)^{p-1} + \bar{M}_2(p)^{2p-2})$ $2^{2p-3}p(2p - 1)(2\beta^{-1}dp(2p - 1)(1 + \bar{L}_3)^2)^p$ $(2^{2p}p(2p - 1)\beta^{-1}d(1 + \bar{L}_3)^2/\bar{a})^{1/2}$
Lemma D.4	$\bar{c}_{V,1}(p)$ $\bar{c}_{V,2}(p)$ $\bar{M}_V(p)$	$\bar{a}p/4$ $(3/4)\bar{a}p v_p(\bar{M}_V(p))$ $(1/3 + 4b/(3\bar{a}) + 4d/(3\bar{a}\beta) + 4(p - 2)/(3\bar{a}\beta))^{1/2}$
Lemma D.6	$C_{\text{Lin},0}$ $C_{\text{Lin},1}$ \bar{C}_{S_2} \tilde{C}_{S_2} $\bar{C}_{S_0,p}$ $\tilde{C}_{S_0,p}$ $\bar{C}_{S_1,p}$ $\tilde{C}_{S_1,p}$	$52e^{2\bar{L}_3+6}\bar{C}_{S_2}$ $52e^{2\bar{L}_3+6}\bar{C}_{S_2}$ $2^7d^4(1 + \beta^{-1})^2(1 + \bar{L}_3)^4(1 + \bar{L}_2)^2(1 + \bar{L}_1)^2(1 + \bar{K}_1)^2(1 + \bar{K}_0)^2(1 + \max\{\bar{C}_{S_0,2}, \bar{C}_{S_0,2q}, \bar{C}_{S_0,1+q}, \bar{C}_{S_1,2}\})$ $2^7d^4(1 + \beta^{-1})^2(1 + \bar{L}_3)^4(1 + \bar{L}_2)^2(1 + \bar{L}_1)^2(1 + \bar{K}_1)^2(1 + \bar{K}_0)^2(\bar{c}_2(1 + 1/\bar{a}) + \max\{\tilde{C}_{S_0,2}, \tilde{C}_{S_0,2q}, \tilde{C}_{S_0,1+q}, \tilde{C}_{S_1,2}\} + 1)$ $10^{2p}(\bar{K}_1^{2p} + (\bar{L}_3\bar{K}_1)^{2p})$ $\bar{C}_{S_0,p}(\bar{c}_{\lceil p \rceil}(1 + 1/\bar{a}) + 1) + 5^{2p}((\beta^{-1}d\bar{L}_2)^{2p} + (2\beta^{-1}(p + 1)(d + 2p))^p(1 + \bar{L}_3^{2p}))$ $2^{4\lceil p \rceil}(1 + \bar{K}_1)^{2p}$ $\bar{C}_{S_1,p}(\bar{c}_{\lceil p \rceil}(1 + 1/\bar{a}) + 1) + 2^{3\lceil p \rceil}(1 + \bar{K}_1)^{2p}3v_{2\lceil p \rceil}(\bar{M}_V(2\lceil p \rceil)) + 2^{2p}(2\beta^{-1}(p + 1)(d + 2p))^p$
Proposition D.7	\dot{c}_{Lin} $\bar{\phi}_{\text{Lin}}$ $\bar{\epsilon}$ $\dot{c}_{\text{Lin},0}$ $\dot{c}_{\text{Lin},1}$ \dot{c}_{Lin}	$\min\{\bar{\phi}_{\text{Lin}}, \bar{c}_{V,1}(2), 4\bar{c}_{V,2}(2)\bar{\epsilon}\bar{c}_{V,1}(2)\}/2$ $\left(\sqrt{8\pi/(\beta\bar{L}_3)}\dot{c}_{\text{Lin},0} \exp \left(\left(\dot{c}_{\text{Lin},0}\sqrt{\beta\bar{L}_3/8} + \sqrt{8/(\beta\bar{L}_3)} \right)^2 \right) \right)^{-1}$ $1 \wedge \left(4\bar{c}_{V,2}(2)\sqrt{2\beta\pi/\bar{L}_3} \int_0^{\dot{c}_{\text{Lin},1}} \exp \left(\left(s\sqrt{\beta\bar{L}_3/8} + \sqrt{8/(\beta\bar{L}_3)} \right)^2 \right) ds \right)^{-1}$ $2(4\bar{c}_{V,2}(2)(1 + \bar{c}_{V,1}(2))/\bar{c}_{V,1}(2) - 1)^{1/2}$ $2(2\bar{c}_{V,2}(2)/\bar{c}_{V,1}(2) - 1)^{1/2}$ $2(1 + \dot{c}_{\text{Lin},0}) \exp(\beta\bar{L}_3\dot{c}_{\text{Lin},0}^2/8 + 2\dot{c}_{\text{Lin},0})/\bar{\epsilon}$
Lemma D.8	$C_{\text{Lin},2}$ $C_{\text{Lin},3}$	$\dot{c}_{\text{Lin}} \left(1 + \frac{4}{\min\{\dot{c}_{\text{Lin}}, \bar{a}/2\}} \right) e^{\min\{\dot{c}_{\text{Lin}}, \bar{a}/2\}/4} (C_{\text{Lin},0} + 9)$ $2(\dot{c}_{\text{Lin}}/\dot{c}_{\text{Lin}}) e^{\dot{c}_{\text{Lin}}/2} (C_{\text{Lin},1} + 9 + 9\bar{c}_2(1 + 1/\bar{a}) + 9v_4(\bar{M}_V(4)))$

APPENDIX D. PROOF OF MAIN RESULTS FOR aHOLLA

We provide the proofs for Theorem 3.3 and Corollary 3.4, which follow the same lines as those for aHOLA. We first introduce auxiliary processes which we use throughout the convergence analysis in Appendix D.1. Then, we provide moment estimates for the newly introduced processes in Appendix D.2, which are followed by the detailed proofs for the main results in Appendix D.3. We postpone the proofs of the results in Appendices D.2 and D.3 to Appendix D.4.

D.1. Auxiliary processes. Fix $\beta > 0$. Consider the Langevin SDE $(Z_t)_{t \geq 0}$ given by

$$Z_0 := \theta_0, \quad dZ_t = -h(Z_t) dt + \sqrt{2\beta^{-1}} dB_t, \quad (108)$$

where $(B_t)_{t \geq 0}$ is a d -dimensional Brownian motion on (Ω, \mathcal{F}, P) with its completed natural filtration denoted by $(\mathcal{F}_t)_{t \geq 0}$. Moreover, we assume that $(\mathcal{F}_t)_{t \geq 0}$ is independent of $\sigma(\theta_0)$. Under Assumption 5, it is a well-known result that the Langevin SDE (108) admits a unique solution, which is adapted to $\mathcal{F}_t \vee \sigma(\theta_0)$, $t \geq 0$. Its $2p$ -th moment estimate with $p \in \mathbb{N}$ is provided in Lemma D.9, which can be used to further deduce the $2p$ -th moment estimate of $\pi_\beta(\theta) \propto e^{-\beta U(\theta)}$.

For each $\lambda > 0$, recall $B_t^\lambda := B_{\lambda t}/\sqrt{\lambda}$, $t \geq 0$. Denote by $(\mathcal{F}_t^\lambda)_{t \geq 0}$ with $\mathcal{F}_t^\lambda := \mathcal{F}_{\lambda t}$, $t \geq 0$, its completed natural filtration, which is independent of $\sigma(\theta_0)$. Moreover, we denote by $Z_t^\lambda := Z_{\lambda t}$, $t \geq 0$, the time-changed version of Langevin SDE (108), which is given by

$$Z_0^\lambda := \theta_0, \quad dZ_t^\lambda = -\lambda h(Z_t^\lambda) dt + \sqrt{2\lambda\beta^{-1}} dB_t^\lambda. \quad (109)$$

Furthermore, we denote by $(\tilde{\Theta}_t^\lambda)_{t \geq 0}$ the continuous-time interpolation of aHOLLA (15)-(17) given by

$$\tilde{\Theta}_0^\lambda := \theta_0, \quad d\tilde{\Theta}_t^\lambda = \lambda \phi_{\text{Lin}}^\lambda(\tilde{\Theta}_{[t]}^\lambda) dt + \sqrt{2\lambda\beta^{-1}} \psi_{\text{Lin}}^\lambda(\tilde{\Theta}_{[t]}^\lambda) dB_t^\lambda, \quad (110)$$

where $\phi_{\text{Lin}}^\lambda$ and $\psi_{\text{Lin}}^\lambda$ are defined in (16) and (17), respectively.

Remark D.1. Similarly, denote by $(\bar{\Theta}_t^\lambda)_{t \geq 0}$ the continuous-time interpolation of the order 1.5 scheme (18) given by

$$\begin{aligned} d\bar{\Theta}_t^\lambda &= -\lambda h(\bar{\Theta}_{[t]}^\lambda) dt + \lambda^2 \int_{[t]}^t \left(H(\bar{\Theta}_{[s]}^\lambda) h(\bar{\Theta}_{[s]}^\lambda) - \beta^{-1} \Upsilon(\bar{\Theta}_{[s]}^\lambda) \right) ds dt \\ &\quad - \lambda \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t H(\bar{\Theta}_{[s]}^\lambda) dB_s^\lambda dt + \sqrt{2\lambda\beta^{-1}} dB_t^\lambda \end{aligned} \quad (111)$$

with $\bar{\Theta}_0^\lambda := \theta_0$. We note that $\mathcal{L}(\tilde{\Theta}_n^\lambda) = \mathcal{L}(\Theta_n^{\text{aHOLLA}}) = \mathcal{L}(\Theta_n^\lambda) = \mathcal{L}(\bar{\Theta}_n^\lambda)$, for each $n \in \mathbb{N}_0$.

Finally, for any $s \geq 0$, consider a continuous-time process $(\zeta_t^{s,v,\lambda})_{t \geq s}$ defined by

$$\zeta_s^{s,v,\lambda} := v \in \mathbb{R}^d, \quad d\zeta_t^{s,v,\lambda} = -\lambda h(\zeta_t^{s,v,\lambda}) dt + \sqrt{2\lambda\beta^{-1}} dB_t^\lambda. \quad (112)$$

Definition D.2. Fix $\lambda > 0$. Define $T \equiv T(\lambda) := \lfloor 1/\lambda \rfloor$. Then, for any $n \in \mathbb{N}_0$ and $t \geq nT$, define

$$\tilde{\zeta}_t^{\lambda,n} := \zeta_t^{nT, \bar{\Theta}_{nT}^\lambda, \lambda}.$$

D.2. Moment estimates. Recall the following Lyapunov functions: for each $p \in [2, \infty) \cap \mathbb{N}$, define $V_p(\theta) := (1 + |\theta|^2)^{p/2}$, for all $\theta \in \mathbb{R}^d$, and moreover, define $v_p(w) := (1 + w^2)^{p/2}$, for all $w \geq 0$. We observe that V_p is twice continuously differentiable and satisfies:

$$\sup_{\theta \in \mathbb{R}^d} |\nabla V_p(\theta)|/V_p(\theta) < \infty, \quad \lim_{|\theta| \rightarrow \infty} \nabla V_p(\theta)/V_p(\theta) = 0. \quad (113)$$

Furthermore, we denote by $\mathcal{P}_{V_p}(\mathbb{R}^d)$ the set of probability measures $\mu \in \mathcal{P}(\mathbb{R}^d)$ which satisfies $\int_{\mathbb{R}^d} V_p(\theta) \mu(d\theta) < \infty$.

Next, we establish moment estimates for $(\tilde{\Theta}_t^\lambda)_{t \geq 0}$ given in (110). The results with explicit constants are provided below. We note that for any $p \in [2, \infty) \cap \mathbb{N}$ and $t \geq 0$, we have that $\mathbb{E}[|\tilde{\Theta}_t^\lambda|^{2p}] = \mathbb{E}[|\bar{\Theta}_t^\lambda|^{2p}]$.

Lemma D.3. Let Assumptions 4, 5, and 6 hold. Then, we obtain the following estimates:

(i) For any $0 < \lambda \leq \bar{\lambda}_{\max}$, $n \in \mathbb{N}_0$, and $t \in (n, n+1]$,

$$\mathbb{E} \left[|\tilde{\Theta}_t^\lambda|^2 \right] \leq (1 - \lambda(t-n)\bar{a}) (1 - \lambda\bar{a})^n \mathbb{E} [|\theta_0|^2] + \bar{c}_0 (1 + 1/\bar{a}),$$

where the constant \bar{c}_0 is given explicitly in (139). In particular, the above inequality implies

$$\sup_{t \geq 0} \mathbb{E} \left[|\tilde{\Theta}_t^\lambda|^2 \right] \leq \mathbb{E} [|\theta_0|^2] + \bar{c}_0 (1 + 1/\bar{a}) < \infty.$$

(ii) For any $p \in [2, \infty) \cap \mathbb{N}$, $0 < \lambda \leq \bar{\lambda}_{\max}$, $n \in \mathbb{N}_0$, and $t \in (n, n+1]$,

$$\mathbb{E} \left[|\tilde{\Theta}_t^\lambda|^{2p} \right] \leq (1 - \lambda(t-n)\bar{a}) (1 - \lambda\bar{a})^n \mathbb{E} [|\theta_0|^{2p}] + \bar{c}_p (1 + 1/\bar{a}),$$

where the constant \bar{c}_p is given explicitly in (147). In particular, the above estimate implies

$$\sup_{t \geq 0} \mathbb{E} \left[|\tilde{\Theta}_t^\lambda|^{2p} \right] \leq \mathbb{E} [|\theta_0|^{2p}] + \bar{c}_p (1 + 1/\bar{a}) < \infty.$$

Proof. See Appendix D.4. □

We provide below a drift condition for V_p (defined in the beginning of Appendix D.2).

Lemma D.4. *Let Assumptions 4, 5, and 6 hold. Then, for any $p \in [2, \infty) \cap \mathbb{N}$, $\theta \in \mathbb{R}^d$, we obtain*

$$\Delta V_p(\theta)/\beta - \langle \nabla V_p(\theta), h(\theta) \rangle \leq -\bar{c}_{V,1}(p)V_p(\theta) + \bar{c}_{V,2}(p),$$

where $\bar{c}_{V,1}(p) := \bar{a}p/4$, $\bar{c}_{V,2}(p) := (3/4)\bar{a}p v_p(\bar{M}_V(p))$ with $\bar{M}_V(p) := (1/3 + 4\bar{b}/(3\bar{a}) + 4d/(3\bar{a}\beta) + 4(p-2)/(3\bar{a}\beta))^{1/2}$.

Proof. See [6, Lemma 3.5]. □

By applying Lemma D.3 and D.4, we obtain moment estimates for $(\tilde{\zeta}_t^{\lambda,n})_{t \geq nT}$ defined in Definition D.2.

Lemma D.5. *Let Assumptions 4, 5, and 6 hold. Then, for any $p \in \mathbb{N}$, $0 < \lambda \leq \bar{\lambda}_{\max}$, $n \in \mathbb{N}_0$, and $t \geq nT$, we obtain*

$$\mathbb{E}[V_{2p}(\tilde{\zeta}_t^{\lambda,n})] \leq 2^{p-1}e^{-\lambda\bar{a}t/2}\mathbb{E}[|\theta_0|^{2p}] + 2^{p-1}(\bar{c}_p(1 + 1/\bar{a}) + 1) + 3v_{2p}(\bar{M}_V(2p)),$$

where \bar{c}_p is given in (147) (see also Lemma D.3) and $\bar{M}_V(2p)$ is given in Lemma D.4.

Proof. See [33, Corollary 4.6]. □

D.3. Proof of main results. In this section, we present key results used to obtain Theorem 3.3. To this end, we split $W_1(\mathcal{L}(\bar{\Theta}_t^\lambda), \pi_\beta)$, for any $t \in (nT, (n+1)T]$ and $n \in \mathbb{N}_0$, by using the law of $\tilde{\zeta}_t^{\lambda,n}$ defined in Definition D.2 and the law of Z_t^λ given in (109) as follows:

$$W_1(\mathcal{L}(\bar{\Theta}_t^\lambda), \pi_\beta) \leq W_1(\mathcal{L}(\bar{\Theta}_t^\lambda), \mathcal{L}(\tilde{\zeta}_t^{\lambda,n})) + W_1(\mathcal{L}(\tilde{\zeta}_t^{\lambda,n}), \mathcal{L}(Z_t^\lambda)) + W_1(\mathcal{L}(Z_t^\lambda), \pi_\beta). \quad (114)$$

In the following lemma, we provide a non-asymptotic estimate for $W_2(\mathcal{L}(\bar{\Theta}_t^\lambda), \mathcal{L}(\tilde{\zeta}_t^{\lambda,n}))$, which can be used to upper bound the first term on the RHS of (114).

Lemma D.6. *Let Assumptions 4, 5, and 6 hold. Then, for any $0 < \lambda \leq \bar{\lambda}_{\max}$, $n \in \mathbb{N}_0$, and $t \in (nT, (n+1)T]$, we obtain*

$$W_2(\mathcal{L}(\bar{\Theta}_t^\lambda), \mathcal{L}(\tilde{\zeta}_t^{\lambda,n})) \leq \lambda^{1+q/2} \left(e^{-\bar{a}n/4} C_{\text{Lin},0} \mathbb{E}[|\theta_0|^4] + C_{\text{Lin},1} \right)^{1/2},$$

where $C_{\text{Lin},0}, C_{\text{Lin},1}$ are given explicitly in (169).

Proof. See Appendix D.4. □

For the last two terms on the RHS of (114), we observe that they can be viewed as Wasserstein-1 distances between distributions of Langevin processes starting from different initial points. Therefore, to obtain their upper bounds, we introduce a semi-metric which allows us to establish a contraction result for the Langevin SDE (108) under our assumptions.

We consider the following semi-metric: for any $p \in [2, \infty) \cap \mathbb{N}$, $\mu, \nu \in \mathcal{P}_{V_p}(\mathbb{R}^d)$, let

$$w_{1,p}(\mu, \nu) := \inf_{\zeta \in \mathcal{C}(\mu, \nu)} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} [1 \wedge |\theta - \theta'|] (1 + V_p(\theta) + V_p(\theta')) \zeta(d\theta, d\theta'). \quad (115)$$

Then, we provide a result which states the contraction property of the Langevin SDE (108) in $w_{1,2}$.

Proposition D.7. *Let Assumptions 4, 5, and 6 hold. Moreover, let $\theta'_0 \in L^2$, and let $(Z'_t)_{t \geq 0}$ be the solution of SDE (108) whose starting point $Z'_0 := \theta'_0$ is assumed to be independent of $\mathcal{F}_\infty := \sigma(\bigcup_{t \geq 0} \mathcal{F}_t)$. Then, we obtain*

$$w_{1,2}(\mathcal{L}(Z_t), \mathcal{L}(Z'_t)) \leq \hat{c}_{\text{Lin}} e^{-\hat{c}_{\text{Lin}} t} w_{1,2}(\mathcal{L}(\theta_0), \mathcal{L}(\theta'_0)), \quad (116)$$

where the explicit expressions for \hat{c}_{Lin} , \bar{c}_{Lin} are given below.

The contraction constant \hat{c}_{Lin} is given by:

$$\hat{c}_{\text{Lin}} := \min\{\bar{\phi}_{\text{Lin}}, \bar{c}_{V,1}(2), 4\bar{c}_{V,2}(2)\bar{c}_{V,1}(2)\}/2,$$

where $\bar{c}_{V,1}(2) := \bar{a}/2$, $\bar{c}_{V,2}(2) := 3\bar{a}v_2(\bar{M}_V(2))/2$ with $\bar{M}_V(2) := (1/3 + 4\bar{b}/(3\bar{a}) + 4d/(3\bar{a}\beta))^{1/2}$, the constant $\bar{\phi}_{\text{Lin}}$ is given by

$$\bar{\phi}_{\text{Lin}} := \left(\sqrt{8\pi/(\beta\bar{L}_3)} \hat{c}_{\text{Lin},0} \exp \left(\left(\hat{c}_{\text{Lin},0} \sqrt{\beta\bar{L}_3/8} + \sqrt{8/(\beta\bar{L}_3)} \right)^2 \right) \right)^{-1},$$

and $\bar{\epsilon} > 0$ is chosen such that

$$\bar{\epsilon} \leq 1 \wedge \left(4\bar{c}_{V,2}(2) \sqrt{2\beta\pi/\bar{L}_3} \int_0^{\hat{c}_{\text{Lin},1}} \exp \left(\left(s \sqrt{\beta\bar{L}_3/8} + \sqrt{8/(\beta\bar{L}_3)} \right)^2 \right) ds \right)^{-1}$$

with $\hat{c}_{\text{Lin},0} := 2(4\bar{c}_{V,2}(2)(1 + \bar{c}_{V,1}(2))/\bar{c}_{V,1}(2) - 1)^{1/2}$ and $\hat{c}_{\text{Lin},1} := 2(2\bar{c}_{V,2}(2)/\bar{c}_{V,1}(2) - 1)^{1/2}$.

Moreover, the constant \hat{c}_{Lin} is given by:

$$\hat{c}_{\text{Lin}} := 2(1 + \hat{c}_{\text{Lin},0}) \exp(\beta\bar{L}_3 \hat{c}_{\text{Lin},0}^2/8 + 2\hat{c}_{\text{Lin},0})/\bar{\epsilon}.$$

Proof. We note that [16, Assumption 2.1] holds with $\kappa = \bar{L}_3$ due to Assumption 5, [16, Assumption 2.2] holds with $V = V_2$ due to Remark D.4, and [16, Assumptions 2.4 and 2.5] hold due to (113). Therefore, we can obtain (116) following the same arguments as in the proof of [6, Proposition 3.14] based on [16, Theorem 2.2, Corollary 2.3]. In addition, \hat{c}_{Lin} , \bar{c}_{Lin} can be obtained following the arguments in the proof of [27, Proposition 4.6]. \square

By using the above result and $W_1 \leq w_{1,2}$ (see [27, Lemma A.3]), we can establish a non-asymptotic error bound for the second term on the RHS of (114). The explicit statement is given below.

Lemma D.8. *Let Assumptions 4, 5, and 6 hold. Then, for any $0 < \lambda \leq \bar{\lambda}_{\text{max}}$, $n \in \mathbb{N}_0$, and $t \in (nT, (n+1)T]$, we obtain*

$$W_1(\mathcal{L}(\tilde{\zeta}_t^{\lambda,n}), \mathcal{L}(Z_t^\lambda)) \leq \lambda^{1+q/2} \left(e^{-\min\{\hat{c}_{\text{Lin}}, \bar{a}/2\}n/4} C_{\text{Lin},2} \mathbb{E}[|\theta_0|^4] + C_{\text{Lin},3} \right),$$

where

$$C_{\text{Lin},2} := \hat{c}_{\text{Lin}} \left(1 + \frac{4}{\min\{\hat{c}_{\text{Lin}}, \bar{a}/2\}} \right) e^{\min\{\hat{c}_{\text{Lin}}, \bar{a}/2\}/4} (C_{\text{Lin},0} + 9), \quad (117)$$

$$C_{\text{Lin},3} := 2(\hat{c}_{\text{Lin}}/\hat{c}_{\text{Lin}}) e^{\hat{c}_{\text{Lin}}/2} (C_{\text{Lin},1} + 9 + 9\bar{c}_2(1 + 1/\bar{a}) + 9v_4(\bar{M}_V(4)))$$

with \hat{c}_{Lin} , \bar{c}_{Lin} given in Proposition D.7, $C_{\text{Lin},0}$, $C_{\text{Lin},1}$ given in (169) (see also Lemma D.6), \bar{c}_2 given in Lemma D.3, and $\bar{M}_V(4)$ given in Lemma D.5.

Proof. See [27, Lemma 4.7]. \square

To obtain an upper bound for the last term on the RHS of (114), we observe that π_β is the invariant measure of the Langevin SDE (109). Thus, by applying Proposition D.7, we have that

$$W_1(\mathcal{L}(Z_t^\lambda), \pi_\beta) \leq \hat{c}_{\text{Lin}} e^{-\hat{c}_{\text{Lin}} \lambda t} w_{1,2}(\mathcal{L}(\theta_0), \pi_\beta) \leq \hat{c}_{\text{Lin}} e^{-\hat{c}_{\text{Lin}} \lambda t} \left[1 + \mathbb{E}[V_2(\theta_0)] + \int_{\mathbb{R}^d} V_2(\theta) \pi_\beta(d\theta) \right]. \quad (118)$$

By using Lemma D.6, D.8 and (118), we can obtain an upper bound for each $W_2(\mathcal{L}(\Theta_n^{\text{aHOLLA}}), \pi_\beta)$, $n \in \mathbb{N}_0$, as stated in Theorem 3.3.

Proof of Theorem 3.3. Substituting the results in Lemma D.6, D.8 and (118) into (114), for any $0 < \lambda \leq \bar{\lambda}_{\text{max}}$, $n \in \mathbb{N}_0$, and $t \in (nT, (n+1)T]$, we have that

$$W_1(\mathcal{L}(\bar{\Theta}_t^\lambda), \pi_\beta) \leq \lambda^{1+q/2} \left(e^{-\bar{a}n/4} C_{\text{Lin},0} \mathbb{E}[|\theta_0|^4] + C_{\text{Lin},1} \right)^{1/2}$$

$$\begin{aligned}
& + \lambda^{1+q/2} \left(e^{-\min\{\dot{c}_{\text{Lin}}, \bar{a}/2\}n/4} C_{\text{Lin},2} \mathbb{E}[|\theta_0|^4] + C_{\text{Lin},3} \right) \\
& + \hat{c}_{\text{Lin}} e^{-\dot{c}_{\text{Lin}} \lambda t} \left[1 + \mathbb{E}[V_2(\theta_0)] + \int_{\mathbb{R}^d} V_2(\theta) \pi_\beta(d\theta) \right] \\
& \leq C_{\text{Lin},1} e^{-C_{\text{Lin},0}(n+1)} (\mathbb{E}[|\theta_0|^4] + 1) + C_{\text{Lin},2} \lambda^{1+q/2},
\end{aligned}$$

where

$$\begin{aligned}
C_{\text{Lin},0} & := \min\{\dot{c}_{\text{Lin}}, \bar{a}/2\}/4, \\
C_{\text{Lin},1} & := e^{\min\{\dot{c}_{\text{Lin}}, \bar{a}/2\}/4} \left[C_{\text{Lin},0}^{1/2} + C_{\text{Lin},2} + \hat{c}_{\text{Lin}} \left(3 + \int_{\mathbb{R}^d} V_2(\theta) \pi_\beta(d\theta) \right) \right], \\
C_{\text{Lin},2} & := C_{\text{Lin},1}^{1/2} + C_{\text{Lin},3}
\end{aligned} \tag{119}$$

with $\dot{c}_{\text{Lin}}, \hat{c}_{\text{Lin}}$ given in Proposition D.7, $C_{\text{Lin},0}, C_{\text{Lin},1}$ given in (169) (see also Lemma D.6), $C_{\text{Lin},2}, C_{\text{Lin},3}$ given in (117) (see also Lemma D.8). The above result implies that, for each $n \in \mathbb{N}_0$,

$$W_1(\mathcal{L}(\bar{\Theta}_{nT}^\lambda), \pi_\beta) \leq C_{\text{Lin},1} e^{-C_{\text{Lin},0}n} (\mathbb{E}[|\theta_0|^4] + 1) + C_{\text{Lin},2} \lambda^{1+q/2},$$

which further yields, by setting nT to n on the LHS and n to n/T on the RHS, that

$$\begin{aligned}
W_1(\mathcal{L}(\bar{\Theta}_n^\lambda), \pi_\beta) & = W_1(\mathcal{L}(\Theta_n^\lambda), \pi_\beta) = W_1(\mathcal{L}(\Theta_n^{\text{aHOLLA}}), \pi_\beta) \\
& \leq C_{\text{Lin},1} e^{-C_{\text{Lin},0}\lambda n} (\mathbb{E}[|\theta_0|^4] + 1) + C_{\text{Lin},2} \lambda^{1+q/2},
\end{aligned}$$

where the inequality holds due to $n\lambda \leq n/T$. This completes the proof. \square

We can obtain the upper bound for $W_2(\mathcal{L}(\Theta_n^{\text{aHOLLA}}), \pi_\beta)$, $n \in \mathbb{N}_0$, as stated in Corollary 3.4, by applying similar arguments as in the proof of Theorem 3.3.

Proof of Corollary 3.4. To establish a non-asymptotic error bound for $W_2(\mathcal{L}(\Theta_n^{\text{aHOLLA}}), \pi_\beta)$, we consider the following splitting: for any $0 < \lambda \leq \bar{\lambda}_{\text{max}}$, $n \in \mathbb{N}_0$, and $t \in (nT, (n+1)T]$,

$$W_2(\mathcal{L}(\bar{\Theta}_t^\lambda), \pi_\beta) \leq W_2(\mathcal{L}(\bar{\Theta}_t^\lambda), \mathcal{L}(\tilde{\zeta}_t^{\lambda,n})) + W_2(\mathcal{L}(\tilde{\zeta}_t^{\lambda,n}), \mathcal{L}(Z_t^\lambda)) + W_2(\mathcal{L}(Z_t^\lambda), \pi_\beta). \tag{120}$$

An upper bound for the first term on the RHS of (120) is provided in Lemma D.6. To establish an estimate for the second term on the RHS of (120), we use $W_2 \leq \sqrt{2w_{1,2}}$ (see [27, Lemma A.3] for the proof) and follow the same arguments as that in the proof of [27, Lemma 4.7]. Consequently, for any $0 < \lambda \leq \bar{\lambda}_{\text{max}}$, $n \in \mathbb{N}_0$, and $t \in (nT, (n+1)T]$, we obtain that,

$$W_2(\mathcal{L}(\tilde{\zeta}_t^{\lambda,n}), \mathcal{L}(Z_t^\lambda)) \leq \lambda^{1/2+q/4} \left(e^{-\min\{\dot{c}_{\text{Lin}}, \bar{a}/2\}n/8} C_{\text{Lin},4} \mathbb{E}^{1/2}[|\theta_0|^4] + C_{\text{Lin},5} \right), \tag{121}$$

where

$$\begin{aligned}
C_{\text{Lin},4} & := \sqrt{\hat{c}_{\text{Lin}}} \left(1 + \frac{8}{\min\{\dot{c}_{\text{Lin}}, \bar{a}/2\}} \right) e^{\min\{\dot{c}_{\text{Lin}}, \bar{a}/2\}/8} \left(C_{\text{Lin},0}^{1/2} + 3 \right), \\
C_{\text{Lin},5} & := 4(\sqrt{\hat{c}_{\text{Lin}}}/\dot{c}_{\text{Lin}}) e^{\dot{c}_{\text{Lin}}/4} \left(C_{\text{Lin},1}^{1/2} + 1 + 3(\bar{c}_2(1 + 1/\bar{a}) + 1)^{1/2} + \sqrt{3v_4^{1/2}}(\bar{M}_V(4)) \right)
\end{aligned} \tag{122}$$

with $\dot{c}_{\text{Lin}}, \hat{c}_{\text{Lin}}$ given in Proposition D.7, $C_{\text{Lin},0}, C_{\text{Lin},1}$ given in (169) (see also Lemma D.6), \bar{c}_2 given in Lemma D.3, and $\bar{M}_V(4)$ given in Lemma D.5. An upper bound for the last term on the RHS of (120) can be obtained by using $W_2 \leq \sqrt{2w_{1,2}}$ and Proposition D.7:

$$\begin{aligned}
W_2(\mathcal{L}(Z_t^\lambda), \pi_\beta) & \leq \sqrt{2\hat{c}_{\text{Lin}}} e^{-\dot{c}_{\text{Lin}} \lambda t/2} w_{1,2}^{1/2}(\mathcal{L}(\theta_0), \pi_\beta) \\
& \leq \sqrt{2\hat{c}_{\text{Lin}}} e^{-\dot{c}_{\text{Lin}} \lambda t/2} \left[1 + \mathbb{E}[V_2(\theta_0)] + \int_{\mathbb{R}^d} V_2(\theta) \pi_\beta(d\theta) \right]^{1/2}.
\end{aligned} \tag{123}$$

Applying the results in Lemma D.6, (121), (123) to (120) yields, for any $0 < \lambda \leq \bar{\lambda}_{\text{max}}$, $n \in \mathbb{N}_0$, and $t \in (nT, (n+1)T]$, that

$$\begin{aligned}
W_2(\mathcal{L}(\bar{\Theta}_t^\lambda), \pi_\beta) & \leq \lambda^{1+q/2} \left(e^{-\bar{a}n/4} C_{\text{Lin},0} \mathbb{E}[|\theta_0|^4] + C_{\text{Lin},1} \right)^{1/2} \\
& + \lambda^{1/2+q/4} \left(e^{-\min\{\dot{c}_{\text{Lin}}, \bar{a}/2\}n/8} C_{\text{Lin},4} \mathbb{E}^{1/2}[|\theta_0|^4] + C_{\text{Lin},5} \right)
\end{aligned}$$

$$\begin{aligned}
& + \sqrt{2\hat{c}_{\text{Lin}}} e^{-\hat{c}_{\text{Lin}}\lambda t/2} \left[1 + \mathbb{E}[V_2(\theta_0)] + \int_{\mathbb{R}^d} V_2(\theta)\pi_\beta(d\theta) \right]^{1/2} \\
& \leq C_{\text{Lin},4} e^{-C_{\text{Lin},3}(n+1)} (\mathbb{E}[|\theta_0|^4] + 1)^{1/2} + C_{\text{Lin},5} \lambda^{1/2+q/4},
\end{aligned}$$

where

$$\begin{aligned}
C_{\text{Lin},3} & := \min\{\hat{c}_{\text{Lin}}, \bar{a}/2\}/8, \\
C_{\text{Lin},4} & := e^{\min\{\hat{c}_{\text{Lin}}, \bar{a}/2\}/8} \left[C_{\text{Lin},0}^{1/2} + C_{\text{Lin},4} + \sqrt{2\hat{c}_{\text{Lin}}} \left(3 + \int_{\mathbb{R}^d} V_2(\theta)\pi_\beta(d\theta) \right)^{1/2} \right], \\
C_{\text{Lin},5} & := C_{\text{Lin},1}^{1/2} + C_{\text{Lin},5}
\end{aligned} \tag{124}$$

with $\hat{c}_{\text{Lin}}, \bar{c}_{\text{Lin}}$ given in Proposition D.7, $C_{\text{Lin},0}, C_{\text{Lin},1}$ given in (169) (see also Lemma D.6), $C_{\text{Lin},4}, C_{\text{Lin},5}$ given in (122). This further implies that, for each $n \in \mathbb{N}_0$,

$$\begin{aligned}
W_2(\mathcal{L}(\bar{\Theta}_n^\lambda), \pi_\beta) & = W_2(\mathcal{L}(\Theta_n^\lambda), \pi_\beta) = W_2(\mathcal{L}(\Theta_n^{\text{aHOLLA}}), \pi_\beta) \\
& \leq C_{\text{Lin},4} e^{-C_{\text{Lin},3}\lambda n} (\mathbb{E}[|\theta_0|^4] + 1)^{1/2} + C_{\text{Lin},5} \lambda^{1/2+q/4},
\end{aligned}$$

which completes the proof. \square

D.4. Proof of auxiliary results in Appendices D.2 and D.3.

Lemma D.9. *Let Assumption 6 hold. Then, for any $p \in \mathbb{N}, t \geq 0$, we obtain that*

$$\mathbb{E}[|Z_t^\lambda|^{2p}] \leq e^{-\lambda p \bar{a} t} \mathbb{E}[|\theta_0|^{2p}] + 2(\bar{b} + \beta^{-1}(d + 2(p-1))) \bar{M}_0^{2p-2} / \bar{a} < \infty,$$

where $\bar{M}_0 := (2(\bar{b} + \beta^{-1}(d + 2(p-1))) / \bar{a})^{1/2}$.

Proof. See [27, Lemma A.1]. \square

Proof of Lemma D.3-(i). For any $0 < \lambda \leq \bar{\lambda}_{\text{max}} \leq 1$ with $\bar{\lambda}_{\text{max}}$ given in (19), $t \in (n, n+1], n \in \mathbb{N}_0$, we define

$$\bar{\Delta}_{n,t}^\lambda := \tilde{\Theta}_n^\lambda + \lambda \phi_{\text{Lin}}^\lambda(\tilde{\Theta}_n^\lambda)(t-n), \quad \bar{\Xi}_{n,t}^\lambda := \sqrt{2\lambda\beta^{-1}} \psi_{\text{Lin}}^\lambda(\tilde{\Theta}_n^\lambda)(B_t^\lambda - B_n^\lambda), \tag{125}$$

where for all $\theta \in \mathbb{R}^d$,

$$\phi_{\text{Lin}}^\lambda(\theta) := -h(\theta) + (\lambda/2) (H(\theta)h(\theta) - \beta^{-1}\Upsilon(\theta)), \tag{126}$$

and

$$\psi_{\text{Lin}}^\lambda(\theta) := \sqrt{1_d - \lambda H(\theta) + (\lambda^2/3)(H(\theta))^2}. \tag{127}$$

Then, by using (110), (125) – (127), and by noticing $\mathbb{E}[\langle \bar{\Delta}_{n,t}^\lambda, \bar{\Xi}_{n,t}^\lambda \mid \tilde{\Theta}_n^\lambda \rangle] = 0$, we have that

$$\mathbb{E}[|\tilde{\Theta}_t^\lambda|^2 \mid \tilde{\Theta}_n^\lambda] = |\bar{\Delta}_{n,t}^\lambda|^2 + \mathbb{E}[|\bar{\Xi}_{n,t}^\lambda|^2 \mid \tilde{\Theta}_n^\lambda]. \tag{128}$$

The second term on the RHS of (128) can be further upper bounded as follows:

$$\begin{aligned}
\mathbb{E}[|\bar{\Xi}_{n,t}^\lambda|^2 \mid \tilde{\Theta}_n^\lambda] & = 2\lambda\beta^{-1} \mathbb{E} \left[\left\langle \psi_{\text{Lin}}^\lambda(\tilde{\Theta}_n^\lambda)(B_t^\lambda - B_n^\lambda), \psi_{\text{Lin}}^\lambda(\tilde{\Theta}_n^\lambda)(B_t^\lambda - B_n^\lambda) \right\rangle \mid \tilde{\Theta}_n^\lambda \right] \\
& = 2\lambda\beta^{-1} \mathbb{E} \left[\left\langle B_t^\lambda - B_n^\lambda, \left(1_d - \lambda H(\tilde{\Theta}_n^\lambda) + (\lambda^2/3)(H(\tilde{\Theta}_n^\lambda))^2 \right) (B_t^\lambda - B_n^\lambda) \right\rangle \mid \tilde{\Theta}_n^\lambda \right] \\
& \leq 2\lambda\beta^{-1} \left(d(t-n) + \lambda \bar{L}_3 d(t-n) + (\lambda^2/3) \bar{L}_3^2 d(t-n) \right) \\
& \leq 2\lambda\beta^{-1} (t-n) d(1 + \bar{L}_3)^2,
\end{aligned} \tag{129}$$

where the first inequality holds due to Remark 3.2. Next, to upper bound the first term on the RHS of (128), we use (125) to obtain

$$|\bar{\Delta}_{n,t}^\lambda|^2 = |\tilde{\Theta}_n^\lambda|^2 + 2\lambda(t-n) \langle \tilde{\Theta}_n^\lambda, \phi_{\text{Lin}}^\lambda(\tilde{\Theta}_n^\lambda) \rangle + \lambda^2(t-n)^2 |\phi_{\text{Lin}}^\lambda(\tilde{\Theta}_n^\lambda)|^2. \tag{130}$$

By using (126), the second term on the RHS of (130) can be estimated as follows:

$$\begin{aligned}
\langle \tilde{\Theta}_n^\lambda, \phi_{\text{Lin}}^\lambda(\tilde{\Theta}_n^\lambda) \rangle & = -\langle \tilde{\Theta}_n^\lambda, h(\tilde{\Theta}_n^\lambda) \rangle + (\lambda/2) \langle \tilde{\Theta}_n^\lambda, H(\tilde{\Theta}_n^\lambda)h(\tilde{\Theta}_n^\lambda) \rangle - (\lambda/2)\beta^{-1} \langle \tilde{\Theta}_n^\lambda, \Upsilon(\tilde{\Theta}_n^\lambda) \rangle \\
& \leq -\bar{a} |\tilde{\Theta}_n^\lambda|^2 + \bar{b} + \lambda \bar{L}_3 \bar{K}_1 (1 + |\tilde{\Theta}_n^\lambda|^2) + (\lambda/2)\beta^{-1} d \bar{L}_2 |\tilde{\Theta}_n^\lambda|,
\end{aligned} \tag{131}$$

where the last inequality holds due to Assumption 6, Remark 3.2 and the following calculations: for all $\theta \in \mathbb{R}^d$,

$$|\theta| |H(\theta)h(\theta)| \leq \bar{L}_3 |\theta| |h(\theta)| = \bar{L}_3 \bar{K}_1 (|\theta| + |\theta|^2) \leq 2\bar{L}_3 \bar{K}_1 (1 + |\theta|^2).$$

To provide an upper bound for the third term on the RHS of (130), we write by straightforward calculations that

$$\begin{aligned} |\phi_{\text{Lin}}^\lambda(\tilde{\Theta}_n^\lambda)|^2 &= | -h(\tilde{\Theta}_n^\lambda) + (\lambda/2)H(\tilde{\Theta}_n^\lambda)h(\tilde{\Theta}_n^\lambda) - (\lambda/2)\beta^{-1}\Upsilon(\tilde{\Theta}_n^\lambda) |^2 \\ &= |h(\tilde{\Theta}_n^\lambda)|^2 + (\lambda^2/4)|H(\tilde{\Theta}_n^\lambda)h(\tilde{\Theta}_n^\lambda)|^2 + (\lambda^2/4)\beta^{-2}|\Upsilon(\tilde{\Theta}_n^\lambda)|^2 \\ &\quad - \lambda \left\langle h(\tilde{\Theta}_n^\lambda), H(\tilde{\Theta}_n^\lambda)h(\tilde{\Theta}_n^\lambda) \right\rangle + \lambda\beta^{-1} \left\langle h(\tilde{\Theta}_n^\lambda), \Upsilon(\tilde{\Theta}_n^\lambda) \right\rangle \\ &\quad - (\lambda^2/2)\beta^{-1} \left\langle H(\tilde{\Theta}_n^\lambda)h(\tilde{\Theta}_n^\lambda), \Upsilon(\tilde{\Theta}_n^\lambda) \right\rangle. \end{aligned} \quad (132)$$

We then provide upper bounds for each of the terms on the RHS of (132). By using Remark 3.2, we obtain that

$$\begin{aligned} |h(\tilde{\Theta}_n^\lambda)|^2 &\leq 2\bar{K}_1^2(1 + |\tilde{\Theta}_n^\lambda|^2), \\ (\lambda^2/4)|H(\tilde{\Theta}_n^\lambda)h(\tilde{\Theta}_n^\lambda)|^2 &\leq (\lambda^2/2)\bar{L}_3^2\bar{K}_1^2(1 + |\tilde{\Theta}_n^\lambda|^2), \\ (\lambda^2/4)\beta^{-2}|\Upsilon(\tilde{\Theta}_n^\lambda)|^2 &\leq (\lambda^2/4)\beta^{-2}d^2\bar{L}_2^2, \\ -\lambda \left\langle h(\tilde{\Theta}_n^\lambda), H(\tilde{\Theta}_n^\lambda)h(\tilde{\Theta}_n^\lambda) \right\rangle &\leq 2\lambda\bar{L}_3\bar{K}_1^2(1 + |\tilde{\Theta}_n^\lambda|^2), \\ \lambda\beta^{-1} \left\langle h(\tilde{\Theta}_n^\lambda), \Upsilon(\tilde{\Theta}_n^\lambda) \right\rangle &\leq \lambda\beta^{-1}d\bar{L}_2\bar{K}_1(1 + |\tilde{\Theta}_n^\lambda|), \\ -(\lambda^2/2)\beta^{-1} \left\langle H(\tilde{\Theta}_n^\lambda)h(\tilde{\Theta}_n^\lambda), \Upsilon(\tilde{\Theta}_n^\lambda) \right\rangle &\leq (\lambda^2/2)\beta^{-1}d\bar{L}_2\bar{L}_3\bar{K}_1(1 + |\tilde{\Theta}_n^\lambda|). \end{aligned} \quad (133)$$

Substituting (133) into (132) yields

$$\begin{aligned} |\phi_{\text{Lin}}^\lambda(\tilde{\Theta}_n^\lambda)|^2 &\leq 2\bar{K}_1^2(1 + |\tilde{\Theta}_n^\lambda|^2) + (\lambda^2/2)\bar{L}_3^2\bar{K}_1^2(1 + |\tilde{\Theta}_n^\lambda|^2) + (\lambda^2/4)\beta^{-2}d^2\bar{L}_2^2 \\ &\quad + 2\lambda\bar{L}_3\bar{K}_1^2(1 + |\tilde{\Theta}_n^\lambda|^2) + \lambda\beta^{-1}d\bar{L}_2\bar{K}_1(1 + |\tilde{\Theta}_n^\lambda|) + (\lambda^2/2)\beta^{-1}d\bar{L}_2\bar{L}_3\bar{K}_1(1 + |\tilde{\Theta}_n^\lambda|). \end{aligned} \quad (134)$$

Combining the results in (131) and (134), we obtain the following upper bound for (130):

$$\begin{aligned} |\bar{\Delta}_{n,t}^\lambda|^2 &\leq |\tilde{\Theta}_n^\lambda|^2 - 2\lambda(t-n)\bar{a}|\tilde{\Theta}_n^\lambda|^2 + \lambda(t-n) \left(2\lambda\bar{L}_3\bar{K}_1 + 2\lambda\bar{K}_1^2 + \lambda^3\bar{L}_3^2\bar{K}_1^2/2 + 2\lambda^2\bar{L}_3\bar{K}_1^2 \right) |\tilde{\Theta}_n^\lambda|^2 \\ &\quad + \lambda(t-n) \left(\beta^{-1}d\bar{L}_2 + \beta^{-1}d\bar{L}_2\bar{K}_1 + \beta^{-1}d\bar{L}_2\bar{L}_3\bar{K}_1/2 \right) |\tilde{\Theta}_n^\lambda| + \lambda(t-n) \left(2\bar{b} + 2\bar{L}_3\bar{K}_1 \right. \\ &\quad \left. + 2\bar{K}_1^2 + \bar{L}_3^2\bar{K}_1^2/2 + \beta^{-2}d^2\bar{L}_2^2/4 + 2\bar{L}_3\bar{K}_1^2 + \beta^{-1}d\bar{L}_2\bar{K}_1 + \beta^{-1}d\bar{L}_2\bar{L}_3\bar{K}_1/2 \right) \\ &= (1 - \lambda(t-n)\bar{a})|\tilde{\Theta}_n^\lambda|^2 - \lambda(t-n)\bar{\mathcal{J}}_1^\lambda(\tilde{\Theta}_n^\lambda) - \lambda(t-n)\bar{\mathcal{J}}_2^\lambda(\tilde{\Theta}_n^\lambda) + \lambda(t-n)\bar{c}_1, \end{aligned} \quad (135)$$

where, for all $\theta \in \mathbb{R}^d$,

$$\begin{aligned} \bar{\mathcal{J}}_1^\lambda(\theta) &:= (\bar{a}/2)|\theta|^2 - \left(2\lambda\bar{L}_3\bar{K}_1 + 2\lambda\bar{K}_1^2 + \lambda^3\bar{L}_3^2\bar{K}_1^2/2 + 2\lambda^2\bar{L}_3\bar{K}_1^2 \right) |\theta|^2 \\ \bar{\mathcal{J}}_2^\lambda(\theta) &:= (\bar{a}/2)|\theta|^2 - \left(\beta^{-1}d\bar{L}_2 + \beta^{-1}d\bar{L}_2\bar{K}_1 + \beta^{-1}d\bar{L}_2\bar{L}_3\bar{K}_1/2 \right) |\theta|, \end{aligned}$$

and where $\bar{c}_1 := 2\bar{b} + 2\bar{L}_3\bar{K}_1 + 2\bar{K}_1^2 + \bar{L}_3^2\bar{K}_1^2/2 + \beta^{-2}d^2\bar{L}_2^2/4 + 2\bar{L}_3\bar{K}_1^2 + \beta^{-1}d\bar{L}_2\bar{K}_1 + \beta^{-1}d\bar{L}_2\bar{L}_3\bar{K}_1/2$. We note that, for all $\theta \in \mathbb{R}^d$, $0 < \lambda \leq \bar{\lambda}_{\max} \leq \min\{\bar{a}/(16\bar{L}_3\bar{K}_1), \bar{a}/(16\bar{K}_1^2), \bar{a}^{1/3}/(4\bar{L}_3\bar{K}_1^2)^{1/3}, \bar{a}^{1/2}/(16\bar{L}_3\bar{K}_1^2)^{1/3}\}$,

$$\bar{\mathcal{J}}_1^\lambda(\theta) = \left(\bar{a}/8 - 2\lambda\bar{L}_3\bar{K}_1 \right) + \left(\bar{a}/8 - 2\lambda\bar{K}_1^2 \right) + \left(\bar{a}/8 - \lambda^3\bar{L}_3^2\bar{K}_1^2/2 \right) + \left(\bar{a}/8 - 2\lambda^2\bar{L}_3\bar{K}_1^2 \right) |\theta|^2 \geq 0. \quad (136)$$

Substituting (136) into (135) yields

$$|\bar{\Delta}_{n,t}^\lambda|^2 \leq (1 - \lambda(t-n)\bar{a})|\tilde{\Theta}_n^\lambda|^2 - \lambda(t-n)\bar{\mathcal{J}}_2^\lambda(\tilde{\Theta}_n^\lambda) + \lambda(t-n)\bar{c}_1.$$

Denote by $\bar{M}_1 := 2(\bar{a}\beta)^{-1}d(\bar{L}_2 + \bar{L}_2\bar{K}_1 + \bar{L}_2\bar{L}_3\bar{K}_1/2)$ and $\bar{S}_{n,\bar{M}_1} := \{\omega \in \Omega : |\tilde{\Theta}_n^\lambda(\omega)| > \bar{M}_1\}$. By observing the fact that

$$\bar{J}_2^\lambda(\theta) > 0 \quad \iff \quad |\theta| > \bar{M},$$

we obtain the following:

$$|\bar{\Delta}_{n,t}^\lambda|^2 \mathbb{1}_{\bar{S}_{n,\bar{M}_1}} \leq (1 - \lambda(t-n)\bar{a})|\tilde{\Theta}_n^\lambda|^2 \mathbb{1}_{\bar{S}_{n,\bar{M}_1}} + \lambda(t-n)\bar{c}_1 \mathbb{1}_{\bar{S}_{n,\bar{M}_1}}.$$

Similarly, we have that

$$\begin{aligned} |\bar{\Delta}_{n,t}^\lambda|^2 \mathbb{1}_{\bar{S}_{n,\bar{M}_1}^c} &\leq (1 - \lambda(t-n)\bar{a})|\tilde{\Theta}_n^\lambda|^2 \mathbb{1}_{\bar{S}_{n,\bar{M}_1}^c} + \lambda(t-n)\bar{c}_1 \mathbb{1}_{\bar{S}_{n,\bar{M}_1}^c} \\ &\quad + \lambda(t-n) (\beta^{-1}d\bar{L}_2 + \beta^{-1}d\bar{L}_2\bar{K}_1 + \beta^{-1}d\bar{L}_2\bar{L}_3\bar{K}_1/2) \bar{M}_1 \mathbb{1}_{\bar{S}_{n,\bar{M}_1}^c}. \end{aligned}$$

Combining the two cases yields

$$|\bar{\Delta}_{n,t}^\lambda|^2 \leq (1 - \lambda(t-n)\bar{a})|\tilde{\Theta}_n^\lambda|^2 + \lambda(t-n)\bar{c}_1 + \lambda(t-n)\beta^{-1}d(\bar{L}_2 + \bar{L}_2\bar{K}_1 + \bar{L}_2\bar{L}_3\bar{K}_1/2) \bar{M}_1. \quad (137)$$

Finally, by substituting (129) and (137) into (128), we obtain

$$\mathbb{E} \left[|\tilde{\Theta}_t^\lambda|^2 \mid \tilde{\Theta}_n^\lambda \right] \leq (1 - \lambda(t-n)\bar{a})|\tilde{\Theta}_n^\lambda|^2 + \lambda(t-n)\bar{c}_0, \quad (138)$$

where

$$\begin{aligned} \bar{c}_0 &:= \beta^{-1}d(\bar{L}_2 + \bar{L}_2\bar{K}_1 + \bar{L}_2\bar{L}_3\bar{K}_1/2) \bar{M}_1 + \bar{c}_1 + 2\beta^{-1}d(1 + \bar{L}_3)^2, \\ \bar{c}_1 &:= 2\bar{b} + 2\bar{L}_3\bar{K}_1 + 2\bar{K}_1^2 + \bar{L}_3^2\bar{K}_1^2/2 + \beta^{-2}d^2\bar{L}_2^2/4 + 2\bar{L}_3\bar{K}_1^2 + \beta^{-1}d(\bar{L}_2\bar{K}_1 + \bar{L}_2\bar{L}_3\bar{K}_1/2), \\ \bar{M}_1 &:= 2(\bar{a}\beta)^{-1}d(\bar{L}_2 + \bar{L}_2\bar{K}_1 + \bar{L}_2\bar{L}_3\bar{K}_1/2). \end{aligned} \quad (139)$$

We observe that, for $0 < \lambda \leq \bar{\lambda}_{\max} \leq 1/\bar{a}$,

$$1 > 1 - \lambda(t-n)\bar{a} > 1 - \lambda\bar{a} \geq 0,$$

then, by induction, (138) implies, for $t \in (n, n+1]$, $n \in \mathbb{N}_0$, $0 < \lambda \leq \bar{\lambda}_{\max} \leq 1$, that,

$$\begin{aligned} \mathbb{E} \left[|\tilde{\Theta}_t^\lambda|^2 \right] &\leq (1 - \lambda(t-n)\bar{a})\mathbb{E} \left[|\tilde{\Theta}_n^\lambda|^2 \right] + \lambda(t-n)\bar{c}_0 \\ &\leq (1 - \lambda(t-n)\bar{a})(1 - \lambda\bar{a})\mathbb{E} \left[|\tilde{\Theta}_{n-1}^\lambda|^2 \right] + \bar{c}_0 + \lambda\bar{c}_0 \\ &\leq (1 - \lambda(t-n)\bar{a})(1 - \lambda\bar{a})^2\mathbb{E} \left[|\tilde{\Theta}_{n-2}^\lambda|^2 \right] + \bar{c}_0 + \lambda\bar{c}_0(1 + (1 - \lambda\bar{a})) \\ &\leq \dots \\ &\leq (1 - \lambda(t-n)\bar{a})(1 - \lambda\bar{a})^n\mathbb{E} \left[|\theta_0|^2 \right] + \bar{c}_0(1 + 1/\bar{a}), \end{aligned}$$

which completes the proof. \square

Proof of Lemma D.3-(ii). For any $p \in [2, \infty) \cap \mathbb{N}$, $0 < \lambda \leq \bar{\lambda}_{\max} \leq 1$ with $\bar{\lambda}_{\max}$ given in (19), $t \in (n, n+1]$, $n \in \mathbb{N}_0$, by using the same arguments as in the proof of [27, Lemma 4.2-(ii)] up to the inequality before [27, Eq. (134)] and by using (110) with (125), we obtain that

$$\begin{aligned} \mathbb{E} \left[|\tilde{\Theta}_t^\lambda|^{2p} \mid \tilde{\Theta}_n^\lambda \right] &\leq |\bar{\Delta}_{n,t}^\lambda|^{2p} + 2^{2p-3}p(2p-1)|\bar{\Delta}_{n,t}^\lambda|^{2p-2}\mathbb{E} \left[|\bar{\Xi}_{n,t}^\lambda|^2 \mid \tilde{\Theta}_n^\lambda \right] \\ &\quad + 2^{2p-3}p(2p-1)\mathbb{E} \left[|\bar{\Xi}_{n,t}^\lambda|^{2p} \mid \tilde{\Theta}_n^\lambda \right]. \end{aligned} \quad (140)$$

Then, by using (125) and (127), we can obtain an upper estimate for the last term in (140) as follows:

$$\begin{aligned} \mathbb{E} \left[|\bar{\Xi}_{n,t}^\lambda|^{2p} \mid \tilde{\Theta}_n^\lambda \right] &= (2\lambda\beta^{-1})^p \mathbb{E} \left[\left\langle (B_t^\lambda - B_n^\lambda), \left(\psi_{\text{Lin}}^\lambda(\tilde{\Theta}_n^\lambda) \right)^2 (B_t^\lambda - B_n^\lambda) \right\rangle^p \mid \tilde{\Theta}_n^\lambda \right] \\ &\leq (2\lambda\beta^{-1})^p \mathbb{E} \left[\left(|B_t^\lambda - B_n^\lambda|^2 + \lambda\bar{L}_3|B_t^\lambda - B_n^\lambda|^2 + (\lambda^2/3)\bar{L}_3^2|B_t^\lambda - B_n^\lambda|^2 \right)^p \mid \tilde{\Theta}_n^\lambda \right] \\ &\leq (2\lambda\beta^{-1})^p (1 + \bar{L}_3)^{2p} \mathbb{E} \left[|B_t^\lambda - B_n^\lambda|^{2p} \right] \\ &\leq (2\lambda\beta^{-1})^p dp(2p-1)(t-n)^p (1 + \bar{L}_3)^{2p} \\ &\leq \lambda(t-n)(2\beta^{-1})^p dp(2p-1)(1 + \bar{L}_3)^{2p}. \end{aligned} \quad (141)$$

Substituting (129) and (141) into (140) yields

$$\mathbb{E} \left[|\tilde{\Theta}_t^\lambda|^{2p} \left| \tilde{\Theta}_n^\lambda \right| \right] \leq |\bar{\Delta}_{n,t}^\lambda|^{2p} + \lambda(t-n)2^{2p-2}p(2p-1)\beta^{-1}d(1+\bar{L}_3)^2|\bar{\Delta}_{n,t}^\lambda|^{2p-2} + \lambda(t-n)\bar{c}_{\Xi}(p), \quad (142)$$

where $\bar{c}_{\Xi}(p) := 2^{2p-3}p(2p-1)(2\beta^{-1}dp(2p-1)(1+\bar{L}_3)^2)^p$. Next, we apply (137) to obtain

$$\begin{aligned} |\bar{\Delta}_{n,t}^\lambda|^{2p} &\leq \left((1-\lambda(t-n)\bar{a})|\tilde{\Theta}_n^\lambda|^2 + \lambda(t-n)\bar{c}_1 + \lambda(t-n)\beta^{-1}d(\bar{L}_2 + \bar{L}_2\bar{K}_1 + \bar{L}_2\bar{L}_3\bar{K}_1/2)\bar{M}_1 \right)^p \\ &\leq (1+\lambda(t-n)\bar{a}/2)^{p-1}(1-\lambda(t-n)\bar{a})^p|\tilde{\Theta}_n^\lambda|^{2p} \\ &\quad + (1+2/(\lambda(t-n)\bar{a}))^{p-1}(\lambda(t-n))^p(\bar{c}_1 + \beta^{-1}d(\bar{L}_2 + \bar{L}_2\bar{K}_1 + \bar{L}_2\bar{L}_3\bar{K}_1/2)\bar{M}_1)^p \\ &\leq (1-\lambda(t-n)\bar{a}/2)^{p-1}(1-\lambda(t-n)\bar{a})|\tilde{\Theta}_n^\lambda|^{2p} \\ &\quad + \lambda(t-n)(1+2/\bar{a})^{p-1}(\bar{c}_1 + \beta^{-1}d(\bar{L}_2 + \bar{L}_2\bar{K}_1 + \bar{L}_2\bar{L}_3\bar{K}_1/2)\bar{M}_1)^p \\ &= \bar{c}_{n,t}^\lambda(p)|\tilde{\Theta}_n^\lambda|^{2p} + \tilde{c}_{n,t}^\lambda(p), \end{aligned} \quad (143)$$

where the second inequality holds due to $(u+v)^p \leq (1+\varepsilon)^{p-1}u^p + (1+\varepsilon^{-1})^{p-1}v^p$, $u, v \geq 0$, $\varepsilon > 0$ with $\varepsilon = \lambda(t-n)\bar{a}/2$, and where

$$\begin{aligned} \bar{c}_{n,t}^\lambda(p) &:= (1-\lambda(t-n)\bar{a}/2)^{p-1}(1-\lambda(t-n)\bar{a}), \\ \tilde{c}_{n,t}^\lambda(p) &:= \lambda(t-n)(1+2/\bar{a})^{p-1}(\bar{c}_1 + \beta^{-1}d(\bar{L}_2 + \bar{L}_2\bar{K}_1 + \bar{L}_2\bar{L}_3\bar{K}_1/2)\bar{M}_1)^p. \end{aligned}$$

In addition, we observe that by (143),

$$|\bar{\Delta}_{n,t}^\lambda|^{2p-2} \leq \bar{c}_{n,t}^\lambda(p-1)|\tilde{\Theta}_n^\lambda|^{2p-2} + \tilde{c}_{n,t}^\lambda(p-1), \quad (144)$$

and, in particular, when $p = 2$, (144) yields $|\bar{\Delta}_{n,t}^\lambda|^2 \leq \bar{c}_{n,t}^\lambda(1)|\tilde{\Theta}_n^\lambda|^2 + \tilde{c}_{n,t}^\lambda(1)$ which is exactly the upper bound (137). By substituting (143) and (144) into (142), we have that

$$\begin{aligned} \mathbb{E} \left[|\tilde{\Theta}_t^\lambda|^{2p} \left| \tilde{\Theta}_n^\lambda \right| \right] &\leq \bar{c}_{n,t}^\lambda(p)|\tilde{\Theta}_n^\lambda|^{2p} + \tilde{c}_{n,t}^\lambda(p) + \lambda(t-n)\bar{c}_{\Xi}(p) \\ &\quad + \lambda(t-n)2^{2p-2}p(2p-1)\beta^{-1}d(1+\bar{L}_3)^2 \left(\bar{c}_{n,t}^\lambda(p-1)|\tilde{\Theta}_n^\lambda|^{2p-2} + \tilde{c}_{n,t}^\lambda(p-1) \right). \end{aligned} \quad (145)$$

Denote by $\bar{M}_2(p) := (2^{2p}p(2p-1)\beta^{-1}d(1+\bar{L}_3)^2/\bar{a})^{1/2}$. For all $|\theta| > \bar{M}_2(p)$, we have that

$$(\lambda(t-n)\bar{a}/4)|\theta|^{2p} > (\lambda(t-n)2^{2p}p(2p-1)\beta^{-1}d(1+\bar{L}_3)^2/4)|\theta|^{2p-2}.$$

Denote by $\bar{S}_{n,\bar{M}_2(p)} := \{\omega \in \Omega : |\tilde{\Theta}_n^\lambda(\omega)| > \bar{M}_2(p)\}$. By using the above inequality, (145) can be further bounded as follows:

$$\begin{aligned} &\mathbb{E} \left[|\tilde{\Theta}_t^\lambda|^{2p} \mathbb{1}_{\bar{S}_{n,\bar{M}_2(p)}} \left| \tilde{\Theta}_n^\lambda \right| \right] \\ &\leq (1-\lambda(t-n)\bar{a}/4)\bar{c}_{n,t}^\lambda(p-1)|\tilde{\Theta}_n^\lambda|^{2p} \mathbb{1}_{\bar{S}_{n,\bar{M}_2(p)}} + \left(\tilde{c}_{n,t}^\lambda(p) + \lambda(t-n)\bar{c}_{\Xi}(p) \right) \mathbb{1}_{\bar{S}_{n,\bar{M}_2(p)}} \\ &\quad + \lambda(t-n)2^{2p-2}p(2p-1)\beta^{-1}d(1+\bar{L}_3)^2\tilde{c}_{n,t}^\lambda(p-1)\mathbb{1}_{\bar{S}_{n,\bar{M}_2(p)}} \\ &\quad - (\lambda(t-n)\bar{a}/4)\bar{c}_{n,t}^\lambda(p-1)|\tilde{\Theta}_n^\lambda|^{2p} \mathbb{1}_{\bar{S}_{n,\bar{M}_2(p)}} \\ &\quad + (\lambda(t-n)2^{2p}p(2p-1)\beta^{-1}d(1+\bar{L}_3)^2/4)\bar{c}_{n,t}^\lambda(p-1)|\tilde{\Theta}_n^\lambda|^{2p-2} \mathbb{1}_{\bar{S}_{n,\bar{M}_2(p)}} \\ &\leq (1-\lambda(t-n)\bar{a})|\tilde{\Theta}_n^\lambda|^{2p} \mathbb{1}_{\bar{S}_{n,\bar{M}_2(p)}} + \lambda(t-n) \left[\bar{c}_{\Xi}(p) + (1+2/\bar{a})^{p-1} \right. \\ &\quad \times (\bar{c}_1 + \beta^{-1}d(\bar{L}_2 + \bar{L}_2\bar{K}_1 + \bar{L}_2\bar{L}_3\bar{K}_1/2)\bar{M}_1)^p + 2^{2p-2}p(2p-1)\beta^{-1}d(1+\bar{L}_3)^2 \\ &\quad \left. \times \left((1+2/\bar{a})^{p-2}(\bar{c}_1 + \beta^{-1}d(\bar{L}_2 + \bar{L}_2\bar{K}_1 + \bar{L}_2\bar{L}_3\bar{K}_1/2)\bar{M}_1)^{p-1} + \bar{M}_2(p)^{2p-2} \right) \right] \mathbb{1}_{\bar{S}_{n,\bar{M}_2(p)}} \\ &\leq (1-\lambda(t-n)\bar{a})|\tilde{\Theta}_n^\lambda|^{2p} \mathbb{1}_{\bar{S}_{n,\bar{M}_2(p)}} + \lambda(t-n)\bar{c}_p \mathbb{1}_{\bar{S}_{n,\bar{M}_2(p)}}, \end{aligned} \quad (146)$$

where

$$\begin{aligned} \bar{c}_p &:= \bar{c}_{\Xi}(p) + (1 + 2/\bar{a})^{p-1} (\bar{c}_1 + \beta^{-1}d (\bar{L}_2 + \bar{L}_2\bar{K}_1 + \bar{L}_2\bar{L}_3\bar{K}_1/2) \bar{M}_1)^p \\ &\quad + 2^{2p-2}p(2p-1)\beta^{-1}d(1 + \bar{L}_3)^2 \\ &\quad \times \left((1 + 2/\bar{a})^{p-2} (\bar{c}_1 + \beta^{-1}d (\bar{L}_2 + \bar{L}_2\bar{K}_1 + \bar{L}_2\bar{L}_3\bar{K}_1/2) \bar{M}_1)^{p-1} + \bar{M}_2(p)^{2p-2} \right), \end{aligned} \quad (147)$$

$$\bar{c}_{\Xi}(p) := 2^{2p-3}p(2p-1)(2\beta^{-1}dp(2p-1)(1 + \bar{L}_3)^2)^p,$$

$$\bar{M}_2(p) := (2^{2p}p(2p-1)\beta^{-1}d(1 + \bar{L}_3)^2/\bar{a})^{1/2}$$

with \bar{c}_1 given in (139). Similarly, by using (145), we have that

$$\mathbb{E} \left[|\tilde{\Theta}_t^\lambda|^{2p} \mathbb{1}_{\bar{S}_{n, \bar{M}_2(p)}^c} \mid \tilde{\Theta}_n^\lambda \right] \leq (1 - \lambda(t-n)\bar{a}) |\tilde{\Theta}_n^\lambda|^{2p} \mathbb{1}_{\bar{S}_{n, \bar{M}_2(p)}^c} + \lambda(t-n)\bar{c}_p \mathbb{1}_{\bar{S}_{n, \bar{M}_2(p)}^c}. \quad (148)$$

Combing (146) and (148) yields the desired result. \square

Lemma D.10. *Let Assumptions 4, 5, and 6 hold. Then, for any $0 < \lambda \leq \bar{\lambda}_{\max}$, $p > 0$, $t \geq 0$, we obtain*

$$\mathbb{E} \left[|\bar{\Theta}_t^\lambda - \bar{\Theta}_{[t]}^\lambda|^{2p} \right] \leq \lambda^p \left(e^{-\lambda\bar{a}[t]} \bar{C}_{S_{0,p}} \mathbb{E}[|\theta_0|^{2[p]}] + \tilde{C}_{S_{0,p}} \right), \quad (149)$$

$$\mathbb{E} \left[|\tilde{\zeta}_t^{\lambda,n} - \tilde{\zeta}_{[t]}^{\lambda,n}|^{2p} \right] \leq \lambda^p \left(e^{-\lambda\bar{a}[t]/2} \bar{C}_{S_{1,p}} \mathbb{E}[|\theta_0|^{2[p]}] + \tilde{C}_{S_{1,p}} \right), \quad (150)$$

where $\bar{C}_{S_{0,p}}$ and $\tilde{C}_{S_{0,p}}$ are given in (151), and $\bar{C}_{S_{1,p}}$ and $\tilde{C}_{S_{1,p}}$ are given in (152).

Proof. To show that (149) holds, we use the definition of $(\bar{\Theta}_t^\lambda)_{t \geq 0}$ given in (111) and obtain that, for any $t \geq 0$,

$$\begin{aligned} &\mathbb{E} \left[|\bar{\Theta}_t^\lambda - \bar{\Theta}_{[t]}^\lambda|^{2p} \right] \\ &= \mathbb{E} \left[\left| -\lambda \int_{[t]}^t h(\bar{\Theta}_{[s]}^\lambda) ds + \lambda^2 \int_{[t]}^t \int_{[s]}^s \left(H(\bar{\Theta}_{[r]}^\lambda) h(\bar{\Theta}_{[r]}^\lambda) - \beta^{-1} \Upsilon(\bar{\Theta}_{[r]}^\lambda) \right) dr ds \right. \right. \\ &\quad \left. \left. - \lambda \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t \int_{[s]}^s H(\bar{\Theta}_{[r]}^\lambda) dB_r^\lambda ds + \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t dB_s^\lambda \right|^{2p} \right] \\ &\leq 5^{2p} \left(\lambda^{2p} \mathbb{E} \left[|h(\bar{\Theta}_{[t]}^\lambda)|^{2p} \right] + \lambda^{4p} \mathbb{E} \left[|H(\bar{\Theta}_{[r]}^\lambda) h(\bar{\Theta}_{[t]}^\lambda)|^{2p} \right] \right. \\ &\quad \left. + \lambda^{4p} \beta^{-2p} \mathbb{E} \left[|\Upsilon(\bar{\Theta}_{[t]}^\lambda)|^{2p} \right] + \lambda^p (2\beta^{-1})^p \mathbb{E} \left[\left| \int_{[t]}^t dB_s^\lambda \right|^{2p} \right] \right. \\ &\quad \left. + \lambda^{3p} (2\beta^{-1})^p \mathbb{E} \left[\left| H(\bar{\Theta}_{[t]}^\lambda) \int_{[t]}^t \int_{[s]}^s dB_r^\lambda ds \right|^{2p} \right] \right) \\ &\leq 5^{2p} \lambda^p \left(\bar{K}_1^{2p} \mathbb{E} \left[(1 + |\bar{\Theta}_{[t]}^\lambda|)^{2p} \right] + (\bar{L}_3\bar{K}_1)^{2p} \mathbb{E} \left[(1 + |\bar{\Theta}_{[t]}^\lambda|)^{2p} \right] + (\beta^{-1}d\bar{L}_2)^{2p} \right. \\ &\quad \left. + (2\beta^{-1}(p+1)(d+2p))^p + (2\beta^{-1}\bar{L}_3^2(p+1)(d+2p))^p \right) \\ &\leq \lambda^p 10^{2p} (\bar{K}_1^{2p} + (\bar{L}_3\bar{K}_1)^{2p}) \left(\mathbb{E} \left[|\bar{\Theta}_{[t]}^\lambda|^{2[p]} \right] + 1 \right) \\ &\quad + \lambda^p 5^{2p} \left((\beta^{-1}d\bar{L}_2)^{2p} + (2\beta^{-1}(p+1)(d+2p))^p + (2\beta^{-1}\bar{L}_3^2(p+1)(d+2p))^p \right) \\ &\leq \lambda^p \left(e^{-\lambda\bar{a}[t]} \bar{C}_{S_{0,p}} \mathbb{E}[|\theta_0|^{2[p]}] + \tilde{C}_{S_{0,p}} \right), \end{aligned}$$

where the first inequality holds due to $(\sum_{l=1}^v u_l)^w \leq v^w \sum_{l=1}^v u_l^w$, $v \in \mathbb{N}$, $u_l \geq 0$, $w > 0$, the second inequality holds due to Remark 3.2, Cauchy-Schwarz inequality, and the following inequality:

$$\mathbb{E} \left[\left| \int_{[t]}^t dB_s^\lambda \right|^{2p} \right] \leq \max\{d^p, (p(d+2p-2))^p\} \leq ((p+1)(d+2p))^p,$$

the fourth inequality holds due to Lemma D.3, and where

$$\begin{aligned}\bar{C}_{\mathbf{S}0,p} &:= 10^{2p}(\bar{K}_1^{2p} + (\bar{L}_3\bar{K}_1)^{2p}), \\ \tilde{C}_{\mathbf{S}0,p} &:= \bar{C}_{\mathbf{S}0,p}(\bar{c}_{[p]}(1 + 1/\bar{a}) + 1) + 5^{2p} \left((\beta^{-1}d\bar{L}_2)^{2p} + (2\beta^{-1}(p+1)(d+2p))^p(1 + \bar{L}_3^{2p}) \right).\end{aligned}\tag{151}$$

The inequality (150) can be obtained by using similar arguments. More precisely, by using Definition D.2 with (112), we obtain that, for any $t \geq 0$,

$$\begin{aligned}\mathbb{E} \left[|\tilde{\zeta}_t^{\lambda,n} - \tilde{\zeta}_{[t]}^{\lambda,n}|^{2p} \right] &= \mathbb{E} \left[\left| -\lambda \int_{[t]}^t h(\tilde{\zeta}_s^{\lambda,n}) ds + \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t dB_s^\lambda \right|^{2p} \right] \\ &\leq 2^{2p} \left(\lambda^{2p} \mathbb{E} \left[\int_{[t]}^t |h(\tilde{\zeta}_s^{\lambda,n})|^{2p} ds \right] + \lambda^p (2\beta^{-1})^p \mathbb{E} \left[\left| \int_{[t]}^t dB_s^\lambda \right|^{2p} \right] \right) \\ &\leq 2^{2p} \lambda^p \left(\bar{K}_1^{2p} \int_{[t]}^t \mathbb{E} \left[(1 + |\tilde{\zeta}_s^{\lambda,n}|)^{2[p]} \right] ds + (2\beta^{-1}(p+1)(d+2p))^p \right) \\ &\leq \lambda^p \left(2^{3[p]} \bar{K}_1^{2p} \int_{[t]}^t \mathbb{E} \left[V_{2[p]}(\tilde{\zeta}_s^{\lambda,n}) \right] ds + 2^{2p} (2\beta^{-1}(p+1)(d+2p))^p \right) \\ &\leq \lambda^p \left(e^{-\lambda\bar{a}[t]/2} \bar{C}_{\mathbf{S}1,p} \mathbb{E}[|\theta_0|^{2[p]}] + \tilde{C}_{\mathbf{S}1,p} \right),\end{aligned}$$

where the second inequality holds due to Remark 3.2 and the last inequality holds due to Lemma D.5 and where

$$\begin{aligned}\bar{C}_{\mathbf{S}1,p} &:= 2^{4[p]}(1 + \bar{K}_1)^{2p}, \\ \tilde{C}_{\mathbf{S}1,p} &:= \bar{C}_{\mathbf{S}1,p}(\bar{c}_{[p]}(1 + 1/\bar{a}) + 1) + 2^{3[p]}(1 + \bar{K}_1)^{2p} 3v_{2[p]}(\bar{M}_V(2[p])) \\ &\quad + 2^{2p}(2\beta^{-1}(p+1)(d+2p))^p.\end{aligned}\tag{152}$$

This completes the proof. \square

Lemma D.11. *Let Assumptions 4, 5, and 6 hold. Then, for any $0 < \lambda \leq \bar{\lambda}_{\max}$, $n \in \mathbb{N}_0$, $t \geq nT$, we obtain the following inequalities:*

$$\begin{aligned}\mathbb{E} \left[\left| -\lambda \int_{[t]}^t \left(H(\bar{\Theta}_s^\lambda) - H(\bar{\Theta}_{[s]}^\lambda) \right) h(\bar{\Theta}_{[s]}^\lambda) ds \right|^2 \right] &\leq \lambda^3 \left(e^{-\bar{a}n/2} \bar{C}_{\mathbf{S}2} \mathbb{E}[|\theta_0|^4] + \tilde{C}_{\mathbf{S}2} \right), \\ \mathbb{E} \left[\left| \lambda^2 \int_{[t]}^t H(\bar{\Theta}_s^\lambda) \int_{[s]}^s H(\bar{\Theta}_{[r]}^\lambda) h(\bar{\Theta}_{[r]}^\lambda) dr ds \right|^2 \right] &\leq \lambda^3 \left(e^{-\bar{a}n/2} \bar{C}_{\mathbf{S}2} \mathbb{E}[|\theta_0|^4] + \tilde{C}_{\mathbf{S}2} \right), \\ \mathbb{E} \left[\left| -\lambda^2 \beta^{-1} \int_{[t]}^t H(\bar{\Theta}_s^\lambda) \int_{[s]}^s \Upsilon(\bar{\Theta}_{[r]}^\lambda) dr ds \right|^2 \right] &\leq \lambda^3 \left(e^{-\bar{a}n/2} \bar{C}_{\mathbf{S}2} \mathbb{E}[|\theta_0|^4] + \tilde{C}_{\mathbf{S}2} \right), \\ \mathbb{E} \left[\left| -\lambda \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t H(\bar{\Theta}_s^\lambda) \int_{[s]}^s H(\bar{\Theta}_{[r]}^\lambda) dB_r^\lambda ds \right|^2 \right] &\leq \lambda^3 \left(e^{-\bar{a}n/2} \bar{C}_{\mathbf{S}2} \mathbb{E}[|\theta_0|^4] + \tilde{C}_{\mathbf{S}2} \right), \\ \mathbb{E} \left[\left| \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t \left(H(\bar{\Theta}_s^\lambda) - H(\bar{\Theta}_{[s]}^\lambda) \right) dB_s^\lambda \right|^2 \right] &\leq \lambda^2 \left(e^{-\bar{a}n/2} \bar{C}_{\mathbf{S}2} \mathbb{E}[|\theta_0|^4] + \tilde{C}_{\mathbf{S}2} \right), \\ \mathbb{E} \left[\left| \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t \left(H(\tilde{\zeta}_s^{\lambda,n}) - H(\tilde{\zeta}_{[s]}^{\lambda,n}) \right) dB_s^\lambda \right|^2 \right] &\leq \lambda^2 \left(e^{-\bar{a}n/4} \bar{C}_{\mathbf{S}2} \mathbb{E}[|\theta_0|^4] + \tilde{C}_{\mathbf{S}2} \right), \\ \mathbb{E} \left[\left| \lambda \beta^{-1} \int_{[t]}^t \left(\Upsilon(\bar{\Theta}_s^\lambda) - \Upsilon(\bar{\Theta}_{[s]}^\lambda) \right) ds \right|^2 \right] &\leq \lambda^{2+q} \left(e^{-\bar{a}n/2} \bar{C}_{\mathbf{S}2} \mathbb{E}[|\theta_0|^4] + \tilde{C}_{\mathbf{S}2} \right),\end{aligned}$$

$$\begin{aligned} & \mathbb{E} \left[\left| -\lambda \int_{[t]}^t H(\bar{\Theta}_s^\lambda) h(\bar{\Theta}_{[s]}^\lambda) ds \right|^2 + \left| \lambda \beta^{-1} \int_{[t]}^t \Upsilon(\bar{\Theta}_s^\lambda) ds \right|^2 \right] \leq \lambda^2 \left(e^{-\bar{a}n/2} \bar{\mathcal{C}}_{\mathbf{S}_2} \mathbb{E}[|\theta_0|^4] + \tilde{\mathcal{C}}_{\mathbf{S}_2} \right), \\ & \mathbb{E} \left[\left| -\lambda \int_{[t]}^t H(\tilde{\zeta}_s^{\lambda,n}) h(\tilde{\zeta}_s^{\lambda,n}) ds \right|^2 + \left| \lambda \beta^{-1} \int_{[t]}^t \Upsilon(\tilde{\zeta}_s^{\lambda,n}) ds \right|^2 \right] \leq \lambda^2 \left(e^{-\bar{a}n/4} \bar{\mathcal{C}}_{\mathbf{S}_2} \mathbb{E}[|\theta_0|^4] + \tilde{\mathcal{C}}_{\mathbf{S}_2} \right), \end{aligned}$$

where $\bar{\mathcal{C}}_{\mathbf{S}_2}$ and $\tilde{\mathcal{C}}_{\mathbf{S}_2}$ are given in (153).

Proof. The inequalities can be obtained by using the following arguments:

(i) To show that the first inequality holds, by using Assumption 5, Remark 3.2, we have that

$$\begin{aligned} & \mathbb{E} \left[\left| -\lambda \int_{[t]}^t \left(H(\bar{\Theta}_s^\lambda) - H(\bar{\Theta}_{[s]}^\lambda) \right) h(\bar{\Theta}_{[s]}^\lambda) ds \right|^2 \right] \\ & \leq \lambda^2 \int_{[t]}^t \mathbb{E} \left[|H(\bar{\Theta}_s^\lambda) - H(\bar{\Theta}_{[s]}^\lambda)|^2 |h(\bar{\Theta}_{[s]}^\lambda)|^2 \right] ds \\ & \leq \lambda^2 \int_{[t]}^t \mathbb{E} \left[\bar{L}_2^2 |\bar{\Theta}_s^\lambda - \bar{\Theta}_{[s]}^\lambda|^2 \bar{\mathcal{K}}_1^2 (1 + |\bar{\Theta}_{[s]}^\lambda|)^2 \right] ds \\ & \leq 2^{3/2} \lambda^2 (\bar{L}_2 \bar{\mathcal{K}}_1)^2 \int_{[t]}^t \left(\mathbb{E} [1 + |\bar{\Theta}_{[s]}^\lambda|^4] \right)^{1/2} \left(\mathbb{E} [|\bar{\Theta}_s^\lambda - \bar{\Theta}_{[s]}^\lambda|^4] \right)^{1/2} ds \\ & \leq 2^{3/2} \lambda^2 (\bar{L}_2 \bar{\mathcal{K}}_1)^2 \left(e^{-\lambda \bar{a} [t]} \mathbb{E} [|\theta_0|^4] + \bar{c}_2 (1 + 1/\bar{a}) + 1 \right)^{1/2} \\ & \quad \times \lambda \left(e^{-\lambda \bar{a} [t]} \bar{\mathcal{C}}_{\mathbf{S}_{0,2}} \mathbb{E}[|\theta_0|^4] + \tilde{\mathcal{C}}_{\mathbf{S}_{0,2}} \right)^{1/2} \\ & \leq \lambda^3 3 (\bar{L}_2 \bar{\mathcal{K}}_1)^2 \left(e^{-\lambda \bar{a} [t]} (1 + \bar{\mathcal{C}}_{\mathbf{S}_{0,2}}) \mathbb{E} [|\theta_0|^4] + \bar{c}_2 (1 + 1/\bar{a}) + \tilde{\mathcal{C}}_{\mathbf{S}_{0,2}} + 1 \right) \\ & \leq \lambda^3 \left(e^{-\bar{a}n/2} \bar{\mathcal{C}}_{\mathbf{S}_2} \mathbb{E} [|\theta_0|^4] + \tilde{\mathcal{C}}_{\mathbf{S}_2} \right), \end{aligned}$$

where the fourth inequality holds by applying Lemma D.3 and D.10, the last inequality holds due to $\lambda [t] \geq \lambda n T \geq n/2$, and where

$$\begin{aligned} \bar{\mathcal{C}}_{\mathbf{S}_2} & := 2^7 d^4 (1 + \beta^{-1})^2 (1 + \bar{L}_3)^4 (1 + \bar{L}_2)^2 (1 + \bar{L}_1)^2 (1 + \bar{\mathcal{K}}_1)^2 (1 + \bar{\mathcal{K}}_0)^2 \\ & \quad \times (1 + \max\{\bar{\mathcal{C}}_{\mathbf{S}_{0,2}}, \bar{\mathcal{C}}_{\mathbf{S}_{0,2q}}, \bar{\mathcal{C}}_{\mathbf{S}_{0,1+q}}, \bar{\mathcal{C}}_{\mathbf{S}_{1,2}}\}), \\ \tilde{\mathcal{C}}_{\mathbf{S}_2} & := 2^7 d^4 (1 + \beta^{-1})^2 (1 + \bar{L}_3)^4 (1 + \bar{L}_2)^2 (1 + \bar{L}_1)^2 (1 + \bar{\mathcal{K}}_1)^2 (1 + \bar{\mathcal{K}}_0)^2 \\ & \quad \times (\bar{c}_2 (1 + 1/\bar{a}) + \max\{\tilde{\mathcal{C}}_{\mathbf{S}_{0,2}}, \tilde{\mathcal{C}}_{\mathbf{S}_{0,2q}}, \tilde{\mathcal{C}}_{\mathbf{S}_{0,1+q}}, \tilde{\mathcal{C}}_{\mathbf{S}_{1,2}}\} + 1) \end{aligned} \tag{153}$$

with $\bar{\mathcal{C}}_{\mathbf{S}_{0,p}}, \tilde{\mathcal{C}}_{\mathbf{S}_{0,p}}, \bar{\mathcal{C}}_{\mathbf{S}_{1,p}}, \tilde{\mathcal{C}}_{\mathbf{S}_{1,p}}, p > 0$, given in (151) and (152), and $\bar{c}_p, \bar{\mathcal{M}}_V(p), p \in [2, \infty) \cap \mathbb{N}$ given in (147) (see also Lemma D.3) and Lemma D.4.

(ii) To establish the second inequality, we apply Remark 3.2 and Lemma D.3 to obtain

$$\begin{aligned} & \mathbb{E} \left[\left| \lambda^2 \int_{[t]}^t H(\bar{\Theta}_s^\lambda) \int_{[s]}^s H(\bar{\Theta}_{[r]}^\lambda) h(\bar{\Theta}_{[r]}^\lambda) dr ds \right|^2 \right] \\ & \leq \lambda^4 \int_{[t]}^t \mathbb{E} \left[|H(\bar{\Theta}_s^\lambda) H(\bar{\Theta}_{[s]}^\lambda) h(\bar{\Theta}_{[s]}^\lambda)|^2 \right] ds \\ & \leq \lambda^4 \int_{[t]}^t \mathbb{E} \left[\bar{L}_3^4 \bar{\mathcal{K}}_1^2 (1 + |\bar{\Theta}_{[s]}^\lambda|)^2 \right] ds \\ & \leq 2 \lambda^4 \bar{L}_3^4 \bar{\mathcal{K}}_1^2 \int_{[t]}^t \mathbb{E} [1 + |\bar{\Theta}_{[s]}^\lambda|^2] ds \\ & \leq 4 \lambda^4 \bar{L}_3^4 \bar{\mathcal{K}}_1^2 \int_{[t]}^t \mathbb{E} [1 + |\bar{\Theta}_{[s]}^\lambda|^4] ds \end{aligned}$$

$$\begin{aligned}
&\leq \lambda^3 4\bar{L}_3^4 \bar{K}_1^2 \left(e^{-\lambda\bar{a}[t]} \mathbb{E}[|\theta_0|^4] + \bar{c}_2(1 + 1/\bar{a}) + 1 \right) \\
&\leq \lambda^3 \left(e^{-\bar{a}n/2} \bar{C}_{S_2} \mathbb{E}[|\theta_0|^4] + \tilde{C}_{S_2} \right),
\end{aligned}$$

where \bar{C}_{S_2} and \tilde{C}_{S_2} are given in (153).

(iii) To obtain the third inequality, we apply Remark 3.2 and write

$$\begin{aligned}
&\mathbb{E} \left[\left| -\lambda^2 \beta^{-1} \int_{[t]}^t H(\bar{\Theta}_s^\lambda) \int_{[s]}^s \Upsilon(\bar{\Theta}_{[r]}^\lambda) dr ds \right|^2 \right] \\
&\leq \lambda^4 \beta^{-2} \int_{[t]}^t \mathbb{E} \left[|H(\bar{\Theta}_s^\lambda) \Upsilon(\bar{\Theta}_{[s]}^\lambda)|^2 \right] ds \\
&\leq \lambda^4 \beta^{-2} \int_{[t]}^t \mathbb{E} \left[\bar{L}_3^2 |\Upsilon(\bar{\Theta}_{[s]}^\lambda)|^2 \right] ds \\
&\leq \lambda^4 (\beta^{-1} \bar{L}_3 d \bar{L}_2)^2 \\
&\leq \lambda^3 \left(e^{-\bar{a}n/2} \bar{C}_{S_2} \mathbb{E}[|\theta_0|^4] + \tilde{C}_{S_2} \right),
\end{aligned}$$

where \bar{C}_{S_2} and \tilde{C}_{S_2} are given in (153).

(iv) To obtain the fourth inequality, we use Cauchy-Schwarz inequality and Lemma D.3:

$$\begin{aligned}
&\mathbb{E} \left[\left| -\lambda \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t H(\bar{\Theta}_s^\lambda) \int_{[s]}^s H(\bar{\Theta}_{[r]}^\lambda) dB_r^\lambda ds \right|^2 \right] \\
&\leq 2\lambda^3 \beta^{-1} \int_{[t]}^t \mathbb{E} \left[\bar{L}_3^2 \left| H(\bar{\Theta}_{[s]}^\lambda) \int_{[s]}^s dB_r^\lambda \right|^2 \right] ds \\
&\leq 2\lambda^3 \beta^{-1} \int_{[t]}^t \mathbb{E} \left[\bar{L}_3^4 \left| \int_{[s]}^s dB_r^\lambda \right|^2 \right] ds \\
&\leq 2\lambda^3 \beta^{-1} \bar{L}_3^4 d \\
&\leq \lambda^3 \left(e^{-\bar{a}n/2} \bar{C}_{S_2} \mathbb{E}[|\theta_0|^4] + \tilde{C}_{S_2} \right),
\end{aligned}$$

where \bar{C}_{S_2} and \tilde{C}_{S_2} are given in (153).

(v) To obtain the fifth inequality, we apply Assumption 5 and Cauchy-Schwarz inequality:

$$\begin{aligned}
&\mathbb{E} \left[\left| \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t \left(H(\bar{\Theta}_s^\lambda) - H(\bar{\Theta}_{[s]}^\lambda) \right) dB_s^\lambda \right|^2 \right] \\
&= 2\lambda\beta^{-1} \int_{[t]}^t \mathbb{E} \left[|H(\bar{\Theta}_s^\lambda) - H(\bar{\Theta}_{[s]}^\lambda)|_F^2 \right] ds \\
&\leq 2\lambda\beta^{-1} \int_{[t]}^t \mathbb{E} \left[d\bar{L}_2^2 |\bar{\Theta}_s^\lambda - \bar{\Theta}_{[s]}^\lambda|^2 \right] ds \\
&\leq 2\lambda\beta^{-1} d\bar{L}_2^2 \int_{[t]}^t \left(\mathbb{E} \left[|\bar{\Theta}_s^\lambda - \bar{\Theta}_{[s]}^\lambda|^4 \right] \right)^{1/2} ds \\
&\leq 2\lambda^2 \beta^{-1} d\bar{L}_2^2 \left(e^{-\lambda\bar{a}[t]} \bar{C}_{S_{0,2}} \mathbb{E}[|\theta_0|^4] + \tilde{C}_{S_{0,2}} \right)^{1/2} \\
&\leq \lambda^2 \left(e^{-\bar{a}n/2} \bar{C}_{S_2} \mathbb{E}[|\theta_0|^4] + \tilde{C}_{S_2} \right),
\end{aligned}$$

where the third inequality holds due to Lemma D.10, and where \bar{C}_{S_2} and \tilde{C}_{S_2} are given in (153).

(vi) To obtain the sixth inequality, we use the same arguments as in (v):

$$\begin{aligned}
& \mathbb{E} \left[\left| \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t \left(H(\tilde{\zeta}_s^{\lambda,n}) - H(\tilde{\zeta}_{[s]}^{\lambda,n}) \right) dB_s^\lambda \right|^2 \right] \\
&= 2\lambda\beta^{-1} \int_{[t]}^t \mathbb{E} \left[|H(\tilde{\zeta}_s^{\lambda,n}) - H(\tilde{\zeta}_{[s]}^{\lambda,n})|_F^2 \right] ds \\
&\leq 2\lambda\beta^{-1} \int_{[t]}^t \mathbb{E} \left[d\bar{L}_2^2 |\tilde{\zeta}_s^{\lambda,n} - \tilde{\zeta}_{[s]}^{\lambda,n}|^2 \right] ds \\
&\leq 2\lambda\beta^{-1} d\bar{L}_2^2 \int_{[t]}^t \left(\mathbb{E} \left[|\tilde{\zeta}_s^{\lambda,n} - \tilde{\zeta}_{[s]}^{\lambda,n}|^4 \right] \right)^{1/2} ds \\
&\leq 2\lambda^2\beta^{-1} d\bar{L}_2^2 \left(e^{-\lambda\bar{a}[t]/2} \bar{C}_{S1,2} \mathbb{E}[|\theta_0|^4] + \tilde{C}_{S1,2} \right)^{1/2} \\
&\leq \lambda^2 \left(e^{-\bar{a}n/4} \bar{C}_{S2} \mathbb{E}[|\theta_0|^4] + \tilde{C}_{S2} \right),
\end{aligned}$$

where the third inequality holds due to Lemma D.10, and where \bar{C}_{S2} and \tilde{C}_{S2} are given in (153).

(vii) To establish the seventh inequality, we apply Remark 3.2 and Cauchy-Schwarz inequality to obtain

$$\begin{aligned}
& \mathbb{E} \left[\left| \lambda\beta^{-1} \int_{[t]}^t \left(\Upsilon(\bar{\Theta}_s^\lambda) - \Upsilon(\bar{\Theta}_{[s]}^\lambda) \right) ds \right|^2 \right] \\
&\leq \lambda^2\beta^{-2} \int_{[t]}^t \mathbb{E} \left[|\Upsilon(\bar{\Theta}_s^\lambda) - \Upsilon(\bar{\Theta}_{[s]}^\lambda)|^2 \right] ds \\
&\leq \lambda^2\beta^{-2} \int_{[t]}^t \mathbb{E} \left[d^3\bar{L}_1^2 |\bar{\Theta}_s^\lambda - \bar{\Theta}_{[s]}^\lambda|^{2q} \right] ds \\
&\leq \lambda^2\beta^{-2} d^3\bar{L}_1^2 \int_{[t]}^t \left(\mathbb{E} \left[|\bar{\Theta}_s^\lambda - \bar{\Theta}_{[s]}^\lambda|^{4q} \right] \right)^{1/2} ds \\
&\leq \lambda^{2+q}\beta^{-2} d^3\bar{L}_1^2 \left(e^{-\lambda\bar{a}[t]} \bar{C}_{S0,2q} \mathbb{E}[|\theta_0|^4] + \tilde{C}_{S0,2q} \right)^{1/2} \\
&\leq \lambda^{2+q} \left(e^{-\bar{a}n/2} \bar{C}_{S2} \mathbb{E}[|\theta_0|^4] + \tilde{C}_{S2} \right),
\end{aligned}$$

where the fourth inequality holds due to Lemma D.10, and \bar{C}_{S2} , \tilde{C}_{S2} are given in (153).

(viii) To obtain the eighth inequality, we apply Remark 3.2 and write

$$\begin{aligned}
& \mathbb{E} \left[\left| -\lambda \int_{[t]}^t H(\bar{\Theta}_s^\lambda) h(\bar{\Theta}_{[s]}^\lambda) ds \right|^2 + \left| \lambda\beta^{-1} \int_{[t]}^t \Upsilon(\bar{\Theta}_s^\lambda) ds \right|^2 \right] \\
&\leq \lambda^2 \int_{[t]}^t \mathbb{E} \left[\bar{L}_3^2 |h(\bar{\Theta}_{[s]}^\lambda)|^2 \right] ds + \lambda^2\beta^{-2} d^2\bar{L}_2^2 \\
&\leq \lambda^2\bar{L}_3^2 \int_{[t]}^t \mathbb{E} \left[\bar{K}_1^2 (1 + |\bar{\Theta}_{[s]}^\lambda|)^2 \right] ds + \lambda^2\beta^{-2} d^2\bar{L}_2^2 \\
&\leq \lambda^2 \left(4\bar{L}_3^2\bar{K}_1^2 \int_{[t]}^t \mathbb{E} \left[(1 + |\bar{\Theta}_{[s]}^\lambda|^4) \right] ds + \beta^{-2} d^2\bar{L}_2^2 \right) \\
&\leq \lambda^2 \left(4\bar{L}_3^2\bar{K}_1^2 \left(e^{-\lambda\bar{a}[t]} \mathbb{E}[|\theta_0|^4] + \bar{c}_2 (1 + 1/\bar{a}) + 1 \right) + \beta^{-2} d^2\bar{L}_2^2 \right) \\
&\leq \lambda^2 4(1 + \bar{L}_3)^2 (1 + \bar{K}_1)^2 (1 + \beta^{-1})^2 d^2 (1 + \bar{L}_2)^2 \left(e^{-\lambda\bar{a}[t]} \mathbb{E}[|\theta_0|^4] + \bar{c}_2 (1 + 1/\bar{a}) + 1 \right) \\
&\leq \lambda^2 \left(e^{-\bar{a}n/2} \bar{C}_{S2} \mathbb{E}[|\theta_0|^4] + \tilde{C}_{S2} \right),
\end{aligned}$$

where the fourth inequality holds due to Lemma D.3, and where \bar{C}_{S2} , \tilde{C}_{S2} are given in (153).

(ix) To establish the last inequality, we follow the arguments in (viii):

$$\begin{aligned}
& \mathbb{E} \left[\left| -\lambda \int_{[t]}^t H(\tilde{\zeta}_s^{\lambda,n}) h(\tilde{\zeta}_s^{\lambda,n}) ds \right|^2 + \left| \lambda \beta^{-1} \int_{[t]}^t \Upsilon(\tilde{\zeta}_s^{\lambda,n}) ds \right|^2 \right] \\
& \leq \lambda^2 \left(2\bar{L}_3^2 \bar{K}_1^2 \int_{[t]}^t \mathbb{E} \left[(1 + |\tilde{\zeta}_s^{\lambda,n}|^2) \right] ds + \beta^{-2} d^2 \bar{L}_2^2 \right) \\
& \leq \lambda^2 \left(2\bar{L}_3^2 \bar{K}_1^2 \int_{[t]}^t \mathbb{E} \left[V_4(\tilde{\zeta}_s^{\lambda,n}) \right] ds + \beta^{-2} d^2 \bar{L}_2^2 \right) \\
& \leq \lambda^2 \left(2\bar{L}_3^2 \bar{K}_1^2 \left(2e^{-\lambda \bar{a} [t]/2} \mathbb{E}[|\theta_0|^4] + 2(\bar{c}_2(1 + 1/\bar{a}) + 1) + 3v_4(\bar{M}_V(4)) \right) + \beta^{-2} d^2 \bar{L}_2^2 \right) \\
& \leq \lambda^2 4(1 + \bar{L}_3)^2 (1 + \bar{K}_1)^2 (1 + \beta^{-1})^2 d^2 (1 + \bar{L}_2)^2 \\
& \quad \times \left(e^{-\lambda \bar{a} [t]/2} \mathbb{E}[|\theta_0|^4] + \bar{c}_2(1 + 1/\bar{a}) + 3v_4(\bar{M}_V(4))/2 + 2 \right) \\
& \leq \lambda^2 \left(e^{-\bar{a}n/4} \bar{C}_{S2} \mathbb{E}[|\theta_0|^4] + \tilde{C}_{S2} \right),
\end{aligned}$$

where the second inequality holds due to Lemma D.5, and where \bar{C}_{S2} and \tilde{C}_{S2} are given in (153).

This completes the proof. \square

Corollary D.12. *Let Assumptions 4, 5, and 6 hold. Then, for any $0 < \lambda \leq \bar{\lambda}_{\max}$, $n \in \mathbb{N}_0$, $t \geq nT$, we obtain the following inequalities:*

$$\begin{aligned}
& \mathbb{E} \left[\left| h(\bar{\Theta}_t^\lambda) - h(\bar{\Theta}_{[t]}^\lambda) + \lambda \int_{[t]}^t \left(H(\bar{\Theta}_{[s]}^\lambda) h(\bar{\Theta}_{[s]}^\lambda) - \beta^{-1} \Upsilon(\bar{\Theta}_{[s]}^\lambda) \right) ds \right. \right. \\
& \quad \left. \left. - \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t H(\bar{\Theta}_{[s]}^\lambda) dB_s^\lambda \right|^2 \right] \leq \lambda^2 \left(e^{-\bar{a}n/2} 36\bar{C}_{S2} \mathbb{E}[|\theta_0|^4] + 36\tilde{C}_{S2} \right), \\
& \mathbb{E} \left[\left| h(\tilde{\zeta}_t^{\lambda,n}) - h(\tilde{\zeta}_{[t]}^{\lambda,n}) - \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t H(\tilde{\zeta}_{[s]}^{\lambda,n}) dB_s^\lambda - \left(h(\bar{\Theta}_t^\lambda) - h(\bar{\Theta}_{[t]}^\lambda) \right. \right. \right. \\
& \quad \left. \left. - \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t H(\bar{\Theta}_{[s]}^\lambda) dB_s^\lambda \right|^2 \right] \leq \lambda^2 \left(e^{-\bar{a}n/4} 72\bar{C}_{S2} \mathbb{E}[|\theta_0|^4] + 72\tilde{C}_{S2} \right),
\end{aligned}$$

where \bar{C}_{S2} and \tilde{C}_{S2} are given in (153).

Proof. For any $t \geq nT$, by applying Itô's formula to $h(\bar{\Theta}_t^\lambda)$, we obtain, almost surely

$$\begin{aligned}
h(\bar{\Theta}_t^\lambda) - h(\bar{\Theta}_{[t]}^\lambda) &= -\lambda \int_{[t]}^t H(\bar{\Theta}_s^\lambda) h(\bar{\Theta}_{[s]}^\lambda) ds + \lambda^2 \int_{[t]}^t H(\bar{\Theta}_s^\lambda) \int_{[s]}^s H(\bar{\Theta}_{[r]}^\lambda) h(\bar{\Theta}_{[r]}^\lambda) dr ds \\
& \quad - \lambda^2 \beta^{-1} \int_{[t]}^t H(\bar{\Theta}_s^\lambda) \int_{[s]}^s \Upsilon(\bar{\Theta}_{[r]}^\lambda) dr ds \\
& \quad - \lambda \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t H(\bar{\Theta}_s^\lambda) \int_{[s]}^s H(\bar{\Theta}_{[r]}^\lambda) dB_r^\lambda ds \\
& \quad + \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t H(\bar{\Theta}_s^\lambda) dB_s^\lambda + \lambda \beta^{-1} \int_{[t]}^t \Upsilon(\bar{\Theta}_s^\lambda) ds.
\end{aligned} \tag{154}$$

Similarly, applying Itô's formula to $h(\tilde{\zeta}_t^{\lambda,n})$ yields, almost surely

$$\begin{aligned}
h(\tilde{\zeta}_t^{\lambda,n}) - h(\tilde{\zeta}_{[t]}^{\lambda,n}) &= -\lambda \int_{[t]}^t H(\tilde{\zeta}_s^{\lambda,n}) h(\tilde{\zeta}_s^{\lambda,n}) ds + \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t H(\tilde{\zeta}_s^{\lambda,n}) dB_s^\lambda \\
& \quad + \lambda \beta^{-1} \int_{[t]}^t \Upsilon(\tilde{\zeta}_s^{\lambda,n}) ds.
\end{aligned} \tag{155}$$

(i) To obtain the first inequality, by using (154) with $(\sum_{l=1}^v u_l)^2 \leq v \sum_{l=1}^v u_l^2$, $v \in \mathbb{N}$, $u_l \geq 0$, we obtain that

$$\begin{aligned}
& \mathbb{E} \left[\left| h(\bar{\Theta}_t^\lambda) - h(\bar{\Theta}_{[t]}^\lambda) + \lambda \int_{[t]}^t \left(H(\bar{\Theta}_{[s]}^\lambda) h(\bar{\Theta}_{[s]}^\lambda) - \beta^{-1} \Upsilon(\bar{\Theta}_{[s]}^\lambda) \right) ds \right. \right. \\
& \quad \left. \left. - \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t H(\bar{\Theta}_{[s]}^\lambda) dB_s^\lambda \right|^2 \right] \\
& \leq 6\mathbb{E} \left[\left| -\lambda \int_{[t]}^t \left(H(\bar{\Theta}_s^\lambda) - H(\bar{\Theta}_{[s]}^\lambda) \right) h(\bar{\Theta}_{[s]}^\lambda) ds \right|^2 \right] \\
& \quad + 6\mathbb{E} \left[\left| \lambda^2 \int_{[t]}^t H(\bar{\Theta}_s^\lambda) \int_{[s]}^s H(\bar{\Theta}_{[r]}^\lambda) h(\bar{\Theta}_{[r]}^\lambda) dr ds \right|^2 \right] \\
& \quad + 6\mathbb{E} \left[\left| -\lambda^2 \beta^{-1} \int_{[t]}^t H(\bar{\Theta}_s^\lambda) \int_{[s]}^s \Upsilon(\bar{\Theta}_{[r]}^\lambda) dr ds \right|^2 \right] \\
& \quad + 6\mathbb{E} \left[\left| -\lambda \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t H(\bar{\Theta}_s^\lambda) \int_{[s]}^s H(\bar{\Theta}_{[r]}^\lambda) dB_r^\lambda ds \right|^2 \right] \\
& \quad + 6\mathbb{E} \left[\left| \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t \left(H(\bar{\Theta}_s^\lambda) - H(\bar{\Theta}_{[s]}^\lambda) \right) dB_s^\lambda \right|^2 \right] \\
& \quad + 6\mathbb{E} \left[\left| \lambda \beta^{-1} \int_{[t]}^t \left(\Upsilon(\bar{\Theta}_s^\lambda) - \Upsilon(\bar{\Theta}_{[s]}^\lambda) \right) ds \right|^2 \right] \\
& \leq \lambda^2 \left(e^{-\bar{a}n/2} 36\bar{\mathcal{C}}_{\mathbf{S}2} \mathbb{E}[|\theta_0|^4] + 36\tilde{\mathcal{C}}_{\mathbf{S}2} \right),
\end{aligned}$$

where the last inequality holds due to Lemma D.11.

(ii) To establish the second inequality, we use (154) and (155) to obtain

$$\begin{aligned}
& \mathbb{E} \left[\left| h(\tilde{\zeta}_t^{\lambda,n}) - h(\tilde{\zeta}_{[t]}^{\lambda,n}) - \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t H(\tilde{\zeta}_{[s]}^{\lambda,n}) dB_s^\lambda \right. \right. \\
& \quad \left. \left. - \left(h(\bar{\Theta}_t^\lambda) - h(\bar{\Theta}_{[t]}^\lambda) - \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t H(\bar{\Theta}_{[s]}^\lambda) dB_s^\lambda \right) \right|^2 \right] \\
& \leq 2\mathbb{E} \left[\left| h(\tilde{\zeta}_t^{\lambda,n}) - h(\tilde{\zeta}_{[t]}^{\lambda,n}) - \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t H(\tilde{\zeta}_{[s]}^{\lambda,n}) dB_s^\lambda \right|^2 \right] \\
& \quad + 2\mathbb{E} \left[\left| h(\bar{\Theta}_t^\lambda) - h(\bar{\Theta}_{[t]}^\lambda) - \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t H(\bar{\Theta}_{[s]}^\lambda) dB_s^\lambda \right|^2 \right] \\
& \leq 6\mathbb{E} \left[\left| -\lambda \int_{[t]}^t H(\tilde{\zeta}_s^{\lambda,n}) h(\tilde{\zeta}_s^{\lambda,n}) ds \right|^2 \right] + 6\mathbb{E} \left[\left| \lambda \beta^{-1} \int_{[t]}^t \Upsilon(\tilde{\zeta}_s^{\lambda,n}) ds \right|^2 \right] \\
& \quad + 6\mathbb{E} \left[\left| \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t \left(H(\tilde{\zeta}_s^{\lambda,n}) - H(\tilde{\zeta}_{[s]}^{\lambda,n}) \right) dB_s^\lambda \right|^2 \right] \\
& \quad + 12\mathbb{E} \left[\left| -\lambda \int_{[t]}^t H(\bar{\Theta}_s^\lambda) h(\bar{\Theta}_{[s]}^\lambda) ds \right|^2 \right] + 12\mathbb{E} \left[\left| \lambda \beta^{-1} \int_{[t]}^t \Upsilon(\bar{\Theta}_s^\lambda) ds \right|^2 \right]
\end{aligned}$$

$$\begin{aligned}
& + 12\mathbb{E} \left[\left| \lambda^2 \int_{[t]}^t H(\bar{\Theta}_s^\lambda) \int_{[s]}^s H(\bar{\Theta}_{[r]}^\lambda) h(\bar{\Theta}_{[r]}^\lambda) dr ds \right|^2 \right] \\
& + 12\mathbb{E} \left[\left| -\lambda^2 \beta^{-1} \int_{[t]}^t H(\bar{\Theta}_s^\lambda) \int_{[s]}^s \Upsilon(\bar{\Theta}_{[r]}^\lambda) dr ds \right|^2 \right] \\
& + 12\mathbb{E} \left[\left| -\lambda \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t H(\bar{\Theta}_s^\lambda) \int_{[s]}^s H(\bar{\Theta}_{[r]}^\lambda) dB_r^\lambda ds \right|^2 \right] \\
& + 12\mathbb{E} \left[\left| \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t \left(H(\bar{\Theta}_s^\lambda) - H(\bar{\Theta}_{[s]}^\lambda) \right) dB_s^\lambda \right|^2 \right] \\
& \leq 12\lambda^2 \left(e^{-\bar{a}n/4} \bar{\mathcal{C}}_{\mathbf{S}_2} \mathbb{E}[|\theta_0|^4] + \tilde{\mathcal{C}}_{\mathbf{S}_2} \right) + 60\lambda^2 \left(e^{-\bar{a}n/4} \bar{\mathcal{C}}_{\mathbf{S}_2} \mathbb{E}[|\theta_0|^4] + \tilde{\mathcal{C}}_{\mathbf{S}_2} \right) \\
& \leq \lambda^2 \left(e^{-\bar{a}n/4} 72\bar{\mathcal{C}}_{\mathbf{S}_2} \mathbb{E}[|\theta_0|^4] + 72\tilde{\mathcal{C}}_{\mathbf{S}_2} \right),
\end{aligned}$$

where the second last inequality holds by using Lemma D.11.

This completes the proof. \square

Definition D.13. Define $\mathfrak{M} = (\mathfrak{M}^{(i,j)})_{i,j=1,\dots,d} : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ by setting, for every $i, j = 1, \dots, d$,

$$\mathfrak{M}^{(i,j)}(\theta, \bar{\theta}) = \langle \nabla H^{(i,j)}(\bar{\theta}), \theta - \bar{\theta} \rangle, \quad \theta, \bar{\theta} \in \mathbb{R}^d.$$

Lemma D.14. Let Assumptions 4, 5, and 6 hold. Then, for any $0 < \lambda \leq \bar{\lambda}_{\max}$, $n \in \mathbb{N}_0$, $t \geq nT$, we obtain the following inequalities:

$$\begin{aligned}
& \mathbb{E} \left[\left| \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t \left(H(\bar{\Theta}_s^\lambda) - H(\bar{\Theta}_{[s]}^\lambda) - \mathfrak{M}(\bar{\Theta}_s^\lambda, \bar{\Theta}_{[s]}^\lambda) \right) dB_s^\lambda \right|^2 \right] \\
& \leq \lambda^{2+q} \left(e^{-\bar{a}n/2} \bar{\mathcal{C}}_{\mathbf{S}_2} \mathbb{E}[|\theta_0|^4] + \tilde{\mathcal{C}}_{\mathbf{S}_2} \right),
\end{aligned} \tag{156}$$

$$\mathbb{E} \left[\left| \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t \mathfrak{M}(\bar{\Theta}_s^\lambda, \bar{\Theta}_{[s]}^\lambda) dB_s^\lambda \right|^2 \right] \leq \lambda^2 \left(e^{-\bar{a}n/2} \bar{\mathcal{C}}_{\mathbf{S}_2} \mathbb{E}[|\theta_0|^4] + \tilde{\mathcal{C}}_{\mathbf{S}_2} \right), \tag{157}$$

$$\begin{aligned}
& \mathbb{E} \left[2\lambda\beta^{-1} \left\langle \int_{[t]}^t \int_{[s]}^s \left(H(\tilde{\zeta}_{[r]}^{\lambda,n}) - H(\bar{\Theta}_{[r]}^\lambda) \right) dB_r^\lambda ds, \int_{[t]}^t \mathfrak{M}(\bar{\Theta}_s^\lambda, \bar{\Theta}_{[s]}^\lambda) dB_s^\lambda \right\rangle \right] \\
& \leq \lambda^2 \left(e^{-\bar{a}n/4} 5\bar{\mathcal{C}}_{\mathbf{S}_2} \mathbb{E}[|\theta_0|^4] + 5\tilde{\mathcal{C}}_{\mathbf{S}_2} \right),
\end{aligned} \tag{158}$$

where $\bar{\mathcal{C}}_{\mathbf{S}_2}$ and $\tilde{\mathcal{C}}_{\mathbf{S}_2}$ are given in (153).

Proof. To show the inequalities hold, we follow the arguments below:

- (i) To establish the first inequality (156), by using Definition D.13, we observe that, for fixed $\theta, \bar{\theta} \in \mathbb{R}^d$,

$$\begin{aligned}
& |H(\theta) - H(\bar{\theta}) - \mathfrak{M}(\theta, \bar{\theta})|_{\mathbb{F}}^2 \\
& = \sum_{i,j=1}^d |H^{(i,j)}(\theta) - H^{(i,j)}(\bar{\theta}) - \mathfrak{M}^{(i,j)}(\theta, \bar{\theta})|^2 \\
& = \sum_{i,j=1}^d \left| \int_0^1 \langle \nabla H^{(i,j)}(\nu\theta + (1-\nu)\bar{\theta}), \theta - \bar{\theta} \rangle d\nu - \langle \nabla H^{(i,j)}(\bar{\theta}), \theta - \bar{\theta} \rangle \right|^2 \\
& \leq \int_0^1 \sum_{i,j=1}^d |\nabla H^{(i,j)}(\nu\theta + (1-\nu)\bar{\theta}) - \nabla H^{(i,j)}(\bar{\theta})|^2 d\nu |\theta - \bar{\theta}|^2
\end{aligned}$$

$$\begin{aligned}
&\leq \int_0^1 \sum_{i=1}^d d |\nabla^2 h^{(i)}(\nu\theta + (1-\nu)\bar{\theta}) - \nabla^2 h^{(i)}(\bar{\theta})|^2 d\nu |\theta - \bar{\theta}|^2 \\
&\leq d^2 \bar{L}_1^2 |\theta - \bar{\theta}|^{2+2q},
\end{aligned} \tag{159}$$

where the last inequality holds due to Assumption 5. By using (159), we obtain that

$$\begin{aligned}
&\mathbb{E} \left[\left| \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t \left(H(\bar{\Theta}_s^\lambda) - H(\bar{\Theta}_{[s]}^\lambda) - \mathfrak{M}(\bar{\Theta}_s^\lambda, \bar{\Theta}_{[s]}^\lambda) \right) dB_s^\lambda \right|^2 \right] \\
&= 2\lambda\beta^{-1} \int_{[t]}^t \mathbb{E} \left[\left| H(\bar{\Theta}_s^\lambda) - H(\bar{\Theta}_{[s]}^\lambda) - \mathfrak{M}(\bar{\Theta}_s^\lambda, \bar{\Theta}_{[s]}^\lambda) \right|_F^2 \right] ds \\
&\leq \lambda\beta^{-1} 2d^2 \bar{L}_1^2 \int_{[t]}^t \mathbb{E} \left[|\bar{\Theta}_s^\lambda - \bar{\Theta}_{[s]}^\lambda|^{2+2q} \right] ds \\
&\leq \lambda^{2+q} \beta^{-1} 2d^2 \bar{L}_1^2 \left(e^{-\lambda\bar{a}[t]} \bar{\mathcal{C}}_{\mathbf{S}0,1+q} \mathbb{E}[|\theta_0|^4] + \tilde{\mathcal{C}}_{\mathbf{S}0,1+q} \right) \\
&\leq \lambda^{2+q} \left(e^{-\bar{a}n/2} \bar{\mathcal{C}}_{\mathbf{S}2} \mathbb{E}[|\theta_0|^4] + \tilde{\mathcal{C}}_{\mathbf{S}2} \right),
\end{aligned}$$

where the third inequality holds due to Lemma D.3 and D.10, and $\bar{\mathcal{C}}_{\mathbf{S}2}$, $\tilde{\mathcal{C}}_{\mathbf{S}2}$ are given in (153).

(ii) To establish the second inequality (157), we recall Definition D.13 and use Remark 3.2 to obtain

$$\begin{aligned}
&\mathbb{E} \left[\left| \sqrt{2\lambda\beta^{-1}} \int_{[t]}^t \mathfrak{M}(\bar{\Theta}_s^\lambda, \bar{\Theta}_{[s]}^\lambda) dB_s^\lambda \right|^2 \right] \\
&= 2\lambda\beta^{-1} \int_{[t]}^t \mathbb{E} \left[\sum_{i,j=1}^d \left| \langle \nabla H^{(i,j)}(\bar{\Theta}_{[s]}^\lambda), \bar{\Theta}_s^\lambda - \bar{\Theta}_{[s]}^\lambda \rangle \right|^2 \right] ds \\
&\leq 2^{3-2q} \lambda\beta^{-1} \int_{[t]}^t \mathbb{E} \left[d^2 \bar{\mathcal{K}}_0^2 (1 + |\bar{\Theta}_{[s]}^\lambda|)^{2q} |\bar{\Theta}_s^\lambda - \bar{\Theta}_{[s]}^\lambda|^2 \right] ds \\
&\leq \lambda\beta^{-1} 2^5 d^2 \bar{\mathcal{K}}_0^2 \int_{[t]}^t \left(\mathbb{E} [1 + |\bar{\Theta}_{[s]}^\lambda|^4] \right)^{1/2} \left(\mathbb{E} [|\bar{\Theta}_s^\lambda - \bar{\Theta}_{[s]}^\lambda|^4] \right)^{1/2} ds \\
&\leq \lambda\beta^{-1} 2^5 d^2 \bar{\mathcal{K}}_0^2 \left(e^{-\lambda\bar{a}[t]} \mathbb{E} [|\theta_0|^4] + \bar{\mathcal{C}}_2 (1 + 1/\bar{a}) + 1 \right)^{1/2} \\
&\quad \times \lambda \left(e^{-\lambda\bar{a}[t]} \bar{\mathcal{C}}_{\mathbf{S}0,2} \mathbb{E}[|\theta_0|^4] + \tilde{\mathcal{C}}_{\mathbf{S}0,2} \right)^{1/2} \\
&\leq \lambda^2 \left(e^{-\bar{a}n/2} \bar{\mathcal{C}}_{\mathbf{S}2} \mathbb{E}[|\theta_0|^4] + \tilde{\mathcal{C}}_{\mathbf{S}2} \right),
\end{aligned}$$

where the third inequality holds due to Lemma D.3 and D.10, and $\bar{\mathcal{C}}_{\mathbf{S}2}$, $\tilde{\mathcal{C}}_{\mathbf{S}2}$ are given in (153).

(iii) To obtain the third inequality (158), we use Definition D.13 and write the following

$$\mathbb{E} \left[2\lambda\beta^{-1} \left\langle \int_{[t]}^t \int_{[s]}^s \left(H(\tilde{\zeta}_{[r]}^{\lambda,n}) - H(\bar{\Theta}_{[r]}^\lambda) \right) dB_r^\lambda ds, \int_{[t]}^t \mathfrak{M}(\bar{\Theta}_s^\lambda, \bar{\Theta}_{[s]}^\lambda) dB_s^\lambda \right\rangle \right]$$

$$\begin{aligned}
&= 2\lambda\beta^{-1}\mathbb{E}\left[\left\langle\int_{[t]}^t\int_{[s]}^s\left(H(\tilde{\zeta}_{[r]}^{\lambda,n})-H(\bar{\Theta}_{[r]}^\lambda)\right)dB_r^\lambda ds,\right.\right. \\
&\quad\left.\left.\int_{[t]}^t\left\langle\nabla H(\bar{\Theta}_{[s]}^\lambda),\int_{[s]}^s-\lambda h(\bar{\Theta}_{[r]}^\lambda)dr\right\rangle dB_s^\lambda\right\rangle\right] \\
&+ 2\lambda\beta^{-1}\mathbb{E}\left[\left\langle\int_{[t]}^t\int_{[s]}^s\left(H(\tilde{\zeta}_{[r]}^{\lambda,n})-H(\bar{\Theta}_{[r]}^\lambda)\right)dB_r^\lambda ds,\right.\right. \\
&\quad\left.\left.\int_{[t]}^t\left\langle\nabla H(\bar{\Theta}_{[s]}^\lambda),\int_{[s]}^s\int_{[r]}^r\lambda^2 H(\bar{\Theta}_{[\nu]}^\lambda)h(\bar{\Theta}_{[\nu]}^\lambda)d\nu dr\right\rangle dB_s^\lambda\right\rangle\right] \\
&+ 2\lambda\beta^{-1}\mathbb{E}\left[\left\langle\int_{[t]}^t\int_{[s]}^s\left(H(\tilde{\zeta}_{[r]}^{\lambda,n})-H(\bar{\Theta}_{[r]}^\lambda)\right)dB_r^\lambda ds,\right.\right. \\
&\quad\left.\left.\int_{[t]}^t\left\langle\nabla H(\bar{\Theta}_{[s]}^\lambda),\int_{[s]}^s\int_{[r]}^r-\lambda^2\beta^{-1}\Upsilon(\bar{\Theta}_{[\nu]}^\lambda)d\nu dr\right\rangle dB_s^\lambda\right\rangle\right] \tag{160} \\
&+ 2\lambda\beta^{-1}\mathbb{E}\left[\left\langle\int_{[t]}^t\int_{[s]}^s\left(H(\tilde{\zeta}_{[r]}^{\lambda,n})-H(\bar{\Theta}_{[r]}^\lambda)\right)dB_r^\lambda ds,\right.\right. \\
&\quad\left.\left.\int_{[t]}^t\left\langle\nabla H(\bar{\Theta}_{[s]}^\lambda),\int_{[s]}^s-\lambda\sqrt{2\lambda\beta^{-1}}\int_{[r]}^r H(\bar{\Theta}_{[\nu]}^\lambda)dB_\nu^\lambda dr\right\rangle dB_s^\lambda\right\rangle\right] \\
&+ 2\lambda\beta^{-1}\mathbb{E}\left[\left\langle\int_{[t]}^t\int_{[s]}^s\left(H(\tilde{\zeta}_{[r]}^{\lambda,n})-H(\bar{\Theta}_{[r]}^\lambda)\right)dB_r^\lambda ds,\right.\right. \\
&\quad\left.\left.\int_{[t]}^t\left\langle\nabla H(\bar{\Theta}_{[s]}^\lambda),\sqrt{2\lambda\beta^{-1}}\int_{[s]}^s dB_r^\lambda\right\rangle dB_s^\lambda\right\rangle\right].
\end{aligned}$$

Recall that $(\mathcal{F}_t^\lambda)_{t\geq 0}$ is the completed natural filtration of $(B_t^\lambda)_{t\geq 0}$. Then, we note that, for any $i, j, k = 1, \dots, d$, it holds that

$$\begin{aligned}
&\mathbb{E}\left[\left(\int_{[t]}^t\int_{[s]}^s\left(H^{(i,j)}(\tilde{\zeta}_{[r]}^{\lambda,n})-H^{(i,j)}(\bar{\Theta}_{[r]}^\lambda)\right)d(B_r^\lambda)^{(j)}ds\right)\right. \\
&\quad\left.\times\left(\int_{[t]}^t\left(\partial_{\theta^{(k)}}H^{(i,j)}(\bar{\Theta}_{[s]}^\lambda)\int_{[s]}^s d(B_r^\lambda)^{(k)}\right)d(B_s^\lambda)^{(j)}\right)\right] \\
&= \mathbb{E}\left[\left(H^{(i,j)}(\tilde{\zeta}_{[t]}^{\lambda,n})-H^{(i,j)}(\bar{\Theta}_{[t]}^\lambda)\right)\partial_{\theta^{(k)}}H^{(i,j)}(\bar{\Theta}_{[t]}^\lambda)\right. \\
&\quad\left.\times\mathbb{E}\left[\int_{[t]}^t\int_{[s]}^s d(B_r^\lambda)^{(j)}ds\int_{[t]}^t\int_{[s]}^s d(B_r^\lambda)^{(k)}d(B_s^\lambda)^{(j)}\middle|\mathcal{F}_{[t]}^\lambda\right]\right] \\
&= \mathbb{E}\left[\left(H^{(i,j)}(\tilde{\zeta}_{[t]}^{\lambda,n})-H^{(i,j)}(\bar{\Theta}_{[t]}^\lambda)\right)\partial_{\theta^{(k)}}H^{(i,j)}(\bar{\Theta}_{[t]}^\lambda)\right. \\
&\quad\left.\times\mathbb{E}\left[\int_{[t]}^t(t-s)d(B_s^\lambda)^{(j)}\int_{[t]}^t\int_{[s]}^s d(B_r^\lambda)^{(k)}d(B_s^\lambda)^{(j)}\middle|\mathcal{F}_{[t]}^\lambda\right]\right] \\
&= \mathbb{E}\left[\left(H^{(i,j)}(\tilde{\zeta}_{[t]}^{\lambda,n})-H^{(i,j)}(\bar{\Theta}_{[t]}^\lambda)\right)\partial_{\theta^{(k)}}H^{(i,j)}(\bar{\Theta}_{[t]}^\lambda)\right. \\
&\quad\left.\times\left(t\int_{[t]}^t\mathbb{E}\left[\int_{[t]}^s d(B_r^\lambda)^{(k)}\middle|\mathcal{F}_{[t]}^\lambda\right]ds-\int_{[t]}^t s\mathbb{E}\left[\int_{[t]}^s d(B_r^\lambda)^{(k)}\middle|\mathcal{F}_{[t]}^\lambda\right]ds\right)\right] \\
&= 0.
\end{aligned}$$

This implies that the last term in (160) is zero. Indeed, we have that

$$\begin{aligned}
& 2\lambda\beta^{-1}\mathbb{E}\left[\left\langle\int_{[t]}^t\int_{[s]}^s\left(H(\tilde{\zeta}_{[r]}^{\lambda,n})-H(\bar{\Theta}_{[r]}^\lambda)\right)dB_r^\lambda ds,\right.\right. \\
& \quad \left.\left.\int_{[t]}^t\left\langle\nabla H(\bar{\Theta}_{[s]}^\lambda),\sqrt{2\lambda\beta^{-1}}\int_{[s]}^s dB_r^\lambda\right\rangle dB_s^\lambda\right\rangle\right] \\
&= (2\lambda\beta^{-1})^{3/2}\mathbb{E}\left[\sum_{i=1}^d\left(\int_{[t]}^t\sum_{j=1}^d\int_{[s]}^s\left(H^{(i,j)}(\tilde{\zeta}_{[r]}^{\lambda,n})-H^{(i,j)}(\bar{\Theta}_{[r]}^\lambda)\right)d(B_r^\lambda)^{(j)} ds\right)\right. \\
& \quad \left.\times\left(\sum_{j=1}^d\int_{[t]}^t\left(\sum_{k=1}^d\partial_{\theta^{(k)}}H^{(i,j)}(\bar{\Theta}_{[s]}^\lambda)\int_{[s]}^s d(B_r^\lambda)^{(k)}\right)d(B_s^\lambda)^{(j)}\right)\right]=0.
\end{aligned}$$

Then, by using Remark 3.2 and (160) with the result above, we obtain that

$$\begin{aligned}
& \mathbb{E}\left[2\lambda\beta^{-1}\left\langle\int_{[t]}^t\int_{[s]}^s\left(H(\tilde{\zeta}_{[r]}^{\lambda,n})-H(\bar{\Theta}_{[r]}^\lambda)\right)dB_r^\lambda ds,\int_{[t]}^t\mathfrak{M}(\bar{\Theta}_s^\lambda,\bar{\Theta}_{[s]}^\lambda)dB_s^\lambda\right\rangle\right] \\
& \leq 4\lambda^2\beta^{-2}\mathbb{E}\left[\left|\int_{[t]}^t\int_{[s]}^s\left(H(\tilde{\zeta}_{[r]}^{\lambda,n})-H(\bar{\Theta}_{[r]}^\lambda)\right)dB_r^\lambda ds\right|^2\right] \\
& \quad + \lambda^2\mathbb{E}\left[\left|\int_{[t]}^t\left\langle\nabla H(\bar{\Theta}_{[s]}^\lambda),\int_{[s]}^s-h(\bar{\Theta}_{[r]}^\lambda)dr\right\rangle dB_s^\lambda\right|^2\right] \\
& \quad + \lambda^4\mathbb{E}\left[\left|\int_{[t]}^t\left\langle\nabla H(\bar{\Theta}_{[s]}^\lambda),\int_{[s]}^s\int_{[r]}^r H(\bar{\Theta}_{[\nu]}^\lambda)h(\bar{\Theta}_{[\nu]}^\lambda)d\nu dr\right\rangle dB_s^\lambda\right|^2\right] \\
& \quad + \lambda^4\beta^{-2}\mathbb{E}\left[\left|\int_{[t]}^t\left\langle\nabla H(\bar{\Theta}_{[s]}^\lambda),\int_{[s]}^s\int_{[r]}^r-\Upsilon(\bar{\Theta}_{[\nu]}^\lambda)d\nu dr\right\rangle dB_s^\lambda\right|^2\right] \\
& \quad + 2\lambda^3\beta^{-1}\mathbb{E}\left[\left|\int_{[t]}^t\left\langle\nabla H(\bar{\Theta}_{[s]}^\lambda),\int_{[s]}^s-\int_{[r]}^r H(\bar{\Theta}_{[\nu]}^\lambda)dB_\nu^\lambda dr\right\rangle dB_s^\lambda\right|^2\right] \\
& \leq 4\lambda^2\beta^{-2}\int_{[t]}^t\int_{[s]}^s\mathbb{E}\left[|H(\tilde{\zeta}_{[r]}^{\lambda,n})-H(\bar{\Theta}_{[r]}^\lambda)|_{\mathbb{F}}^2\right]dr ds \\
& \quad + \lambda^2\int_{[t]}^t\mathbb{E}\left[\sum_{i,j=1}^d\left|\left\langle\nabla H^{(i,j)}(\bar{\Theta}_{[s]}^\lambda),h(\bar{\Theta}_{[s]}^\lambda)\right\rangle\right|^2\right]ds \\
& \quad + \lambda^4\int_{[t]}^t\mathbb{E}\left[\sum_{i,j=1}^d\left|\left\langle\nabla H^{(i,j)}(\bar{\Theta}_{[s]}^\lambda),H(\bar{\Theta}_{[s]}^\lambda)h(\bar{\Theta}_{[s]}^\lambda)\right\rangle\right|^2\right]ds \\
& \quad + \lambda^4\beta^{-2}\int_{[t]}^t\mathbb{E}\left[\sum_{i,j=1}^d\left|\left\langle\nabla H^{(i,j)}(\bar{\Theta}_{[s]}^\lambda),\Upsilon(\bar{\Theta}_{[s]}^\lambda)\right\rangle\right|^2\right]ds \\
& \quad + 2\lambda^3\beta^{-1}\int_{[t]}^t\mathbb{E}\left[\sum_{i,j=1}^d\left|\left\langle\nabla H^{(i,j)}(\bar{\Theta}_{[s]}^\lambda),H(\bar{\Theta}_{[s]}^\lambda)\int_{[s]}^s\int_{[r]}^r dB_\nu^\lambda dr\right\rangle\right|^2\right]ds \\
& \leq 8\lambda^2\beta^{-2}d\bar{L}_2^2\int_{[t]}^t\mathbb{E}\left[|\tilde{\zeta}_{[s]}^{\lambda,n}|^2+|\bar{\Theta}_{[s]}^\lambda|^2\right]ds \\
& \quad + \lambda^2\int_{[t]}^t\mathbb{E}\left[d^2\bar{K}_0^2(1+|\bar{\Theta}_{[s]}^\lambda|^q)^2\bar{K}_1^2(1+|\bar{\Theta}_{[s]}^\lambda|^2)\right]ds
\end{aligned}$$

$$\begin{aligned}
& + \lambda^4 \int_{[t]}^t \mathbb{E} \left[d^2 \bar{\mathcal{K}}_0^2 (1 + |\bar{\Theta}_{[s]}^\lambda|^q)^2 \bar{L}_3^2 \bar{\mathcal{K}}_1^2 (1 + |\bar{\Theta}_{[s]}^\lambda|)^2 \right] ds \\
& + \lambda^4 \beta^{-2} \int_{[t]}^t \mathbb{E} \left[d^2 \bar{\mathcal{K}}_0^2 (1 + |\bar{\Theta}_{[s]}^\lambda|^q)^2 d^2 \bar{L}_2^2 \right] ds \\
& + 2\lambda^3 \beta^{-1} \int_{[t]}^t \mathbb{E} \left[d^2 \bar{\mathcal{K}}_0^2 (1 + |\bar{\Theta}_{[s]}^\lambda|^q)^2 \bar{L}_3^2 \left| \int_{[s]}^s \int_{[r]}^r dB_\nu^\lambda dr \right|^2 \right] ds \\
& \leq 8\lambda^2 \beta^{-2} d \bar{L}_2^2 \int_{[t]}^t \mathbb{E} \left[V_4(\tilde{\zeta}_{[s]}^{\lambda,n}) + 1 + |\bar{\Theta}_{[s]}^\lambda|^4 \right] ds \\
& \quad + 4\lambda^2 (\bar{\mathcal{K}}_1^2 + \bar{L}_3^2 \bar{\mathcal{K}}_1^2 + d^2 \beta^{-2} \bar{L}_2^2) d^2 \bar{\mathcal{K}}_0^2 \int_{[t]}^t \mathbb{E} \left[1 + |\bar{\Theta}_{[s]}^\lambda|^4 \right] ds \\
& \quad + 8\lambda^2 \beta^{-1} d^2 \bar{\mathcal{K}}_0^2 \bar{L}_3^2 \int_{[t]}^t \left(\mathbb{E} \left[1 + |\bar{\Theta}_{[s]}^\lambda|^4 \right] \right)^{1/2} (3(d+4)) ds \\
& \leq 24\lambda^2 \beta^{-2} d \bar{L}_2^2 \left(e^{-\lambda \bar{a} |t|/2} \mathbb{E}[|\theta_0|^4] + (\bar{c}_2 (1 + 1/\bar{a}) + 1) + v_4(\bar{M}_V(4)) \right) \\
& \quad + 2^7 \lambda^2 (\bar{\mathcal{K}}_1^2 + \bar{L}_3^2 \bar{\mathcal{K}}_1^2 + \beta^{-2} \bar{L}_2^2 + \beta^{-1} \bar{L}_3^2) d^4 \bar{\mathcal{K}}_0^2 \\
& \quad \times \left(e^{-\lambda \bar{a} |t|/2} \mathbb{E}[|\theta_0|^4] + \bar{c}_2 (1 + 1/\bar{a}) + 1 \right) \\
& \leq \lambda^2 \left(e^{-\bar{a} n/4} 5 \bar{C}_{S_2} \mathbb{E}[|\theta_0|^4] + 5 \tilde{C}_{S_2} \right),
\end{aligned}$$

where the fifth inequality holds due to Lemma D.3 and D.5, and where $\bar{C}_{S_2}, \tilde{C}_{S_2}$ are given in (153).

This completes the proof. \square

Proof of Lemma D.6. By using the definitions of $\bar{\Theta}_t^\lambda$ in (111) and $\tilde{\zeta}_t^{\lambda,n}$ in Definition D.2, and by applying Itô's formula, we obtain, for any $n \in \mathbb{N}_0, t \in (nT, (n+1)T)$,

$$\begin{aligned}
& W_2^2(\mathcal{L}(\bar{\Theta}_t^\lambda), \mathcal{L}(\tilde{\zeta}_t^{\lambda,n})) \\
& \leq \mathbb{E} \left[\left| \bar{\Theta}_t^\lambda - \tilde{\zeta}_t^{\lambda,n} \right|^2 \right] \\
& = -2\lambda \mathbb{E} \left[\int_{nT}^t \left\langle \bar{\Theta}_s^\lambda - \tilde{\zeta}_s^{\lambda,n}, h(\bar{\Theta}_{[s]}^\lambda) - h(\tilde{\zeta}_s^{\lambda,n}) \right\rangle ds \right] \\
& \quad + 2\lambda \mathbb{E} \left[\int_{nT}^t \left\langle \bar{\Theta}_s^\lambda - \tilde{\zeta}_s^{\lambda,n}, \lambda \int_{[s]}^s \left(H(\bar{\Theta}_{[r]}^\lambda) h(\bar{\Theta}_{[r]}^\lambda) - \beta^{-1} \Upsilon(\bar{\Theta}_{[r]}^\lambda) \right) dr \right\rangle ds \right] \\
& \quad - 2\lambda \mathbb{E} \left[\int_{nT}^t \left\langle \bar{\Theta}_s^\lambda - \tilde{\zeta}_s^{\lambda,n}, \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s H(\bar{\Theta}_{[r]}^\lambda) dB_r^\lambda \right\rangle ds \right] \\
& = -2\lambda \mathbb{E} \left[\int_{nT}^t \left\langle \bar{\Theta}_s^\lambda - \tilde{\zeta}_s^{\lambda,n}, h(\bar{\Theta}_s^\lambda) - h(\tilde{\zeta}_s^{\lambda,n}) \right\rangle ds \right] \\
& \quad - 2\lambda \mathbb{E} \left[\int_{nT}^t \left\langle \bar{\Theta}_s^\lambda - \tilde{\zeta}_s^{\lambda,n}, h(\bar{\Theta}_{[s]}^\lambda) - h(\bar{\Theta}_s^\lambda) \right\rangle ds \right] \\
& \quad + 2\lambda \mathbb{E} \left[\int_{nT}^t \left\langle \bar{\Theta}_s^\lambda - \tilde{\zeta}_s^{\lambda,n}, \lambda \int_{[s]}^s \left(H(\bar{\Theta}_{[r]}^\lambda) h(\bar{\Theta}_{[r]}^\lambda) - \beta^{-1} \Upsilon(\bar{\Theta}_{[r]}^\lambda) \right) dr \right\rangle ds \right] \\
& \quad - 2\lambda \mathbb{E} \left[\int_{nT}^t \left\langle \bar{\Theta}_s^\lambda - \tilde{\zeta}_s^{\lambda,n}, \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s H(\bar{\Theta}_{[r]}^\lambda) dB_r^\lambda \right\rangle ds \right].
\end{aligned} \tag{161}$$

By applying Itô's formula to $h(\bar{\Theta}_s^\lambda)$, we obtain (154). Substituting (154) into (161), using Assumption 5 and Young's inequality yield

$$\begin{aligned}
& \mathbb{E} \left[\left| \bar{\Theta}_t^\lambda - \tilde{\zeta}_t^{\lambda,n} \right|^2 \right] \\
& \leq 2\lambda \bar{L}_3 \int_{nT}^t \mathbb{E} \left[\left| \bar{\Theta}_s^\lambda - \tilde{\zeta}_s^{\lambda,n} \right|^2 \right] ds \\
& \quad + 2\lambda \mathbb{E} \left[\int_{nT}^t \left\langle \bar{\Theta}_s^\lambda - \tilde{\zeta}_s^{\lambda,n}, -\lambda \int_{[s]}^s \left(H(\bar{\Theta}_r^\lambda) - H(\bar{\Theta}_{[r]}) \right) h(\bar{\Theta}_{[r]}) dr \right\rangle ds \right] \\
& \quad + 2\lambda \mathbb{E} \left[\int_{nT}^t \left\langle \bar{\Theta}_s^\lambda - \tilde{\zeta}_s^{\lambda,n}, \lambda^2 \int_{[s]}^s H(\bar{\Theta}_r^\lambda) \int_{[r]}^r H(\bar{\Theta}_{[\nu]}) h(\bar{\Theta}_{[\nu]}) d\nu dr \right\rangle ds \right] \\
& \quad + 2\lambda \mathbb{E} \left[\int_{nT}^t \left\langle \bar{\Theta}_s^\lambda - \tilde{\zeta}_s^{\lambda,n}, -\lambda^2 \beta^{-1} \int_{[s]}^s H(\bar{\Theta}_r^\lambda) \int_{[r]}^r \Upsilon(\bar{\Theta}_{[\nu]}) d\nu dr \right\rangle ds \right] \\
& \quad + 2\lambda \mathbb{E} \left[\int_{nT}^t \left\langle \bar{\Theta}_s^\lambda - \tilde{\zeta}_s^{\lambda,n}, -\lambda \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s H(\bar{\Theta}_r^\lambda) \int_{[r]}^r H(\bar{\Theta}_{[\nu]}) dB_\nu^\lambda dr \right\rangle ds \right] \\
& \quad + 2\lambda \mathbb{E} \left[\int_{nT}^t \left\langle \bar{\Theta}_s^\lambda - \tilde{\zeta}_s^{\lambda,n}, \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s \left(H(\bar{\Theta}_r^\lambda) - H(\bar{\Theta}_{[r]}) \right) dB_r^\lambda \right\rangle ds \right] \\
& \quad + 2\lambda \mathbb{E} \left[\int_{nT}^t \left\langle \bar{\Theta}_s^\lambda - \tilde{\zeta}_s^{\lambda,n}, \lambda\beta^{-1} \int_{[s]}^s \left(\Upsilon(\bar{\Theta}_r^\lambda) - \Upsilon(\bar{\Theta}_{[r]}) \right) dr \right\rangle ds \right] \\
& \leq \lambda(2\bar{L}_3 + 5) \int_{nT}^t \mathbb{E} \left[\left| \bar{\Theta}_s^\lambda - \tilde{\zeta}_s^{\lambda,n} \right|^2 \right] ds \\
& \quad + \lambda \int_{nT}^t \mathbb{E} \left[\left| -\lambda \int_{[s]}^s \left(H(\bar{\Theta}_r^\lambda) - H(\bar{\Theta}_{[r]}) \right) h(\bar{\Theta}_{[r]}) dr \right|^2 \right] ds \\
& \quad + \lambda \int_{nT}^t \mathbb{E} \left[\left| \lambda^2 \int_{[s]}^s H(\bar{\Theta}_r^\lambda) \int_{[r]}^r H(\bar{\Theta}_{[\nu]}) h(\bar{\Theta}_{[\nu]}) d\nu dr \right|^2 \right] ds \\
& \quad + \lambda \int_{nT}^t \mathbb{E} \left[\left| -\lambda^2 \beta^{-1} \int_{[s]}^s H(\bar{\Theta}_r^\lambda) \int_{[r]}^r \Upsilon(\bar{\Theta}_{[\nu]}) d\nu dr \right|^2 \right] ds \\
& \quad + \lambda \int_{nT}^t \mathbb{E} \left[\left| -\lambda \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s H(\bar{\Theta}_r^\lambda) \int_{[r]}^r H(\bar{\Theta}_{[\nu]}) dB_\nu^\lambda dr \right|^2 \right] ds \\
& \quad + \lambda \int_{nT}^t \mathbb{E} \left[\left| \lambda\beta^{-1} \int_{[s]}^s \left(\Upsilon(\bar{\Theta}_r^\lambda) - \Upsilon(\bar{\Theta}_{[r]}) \right) dr \right|^2 \right] ds \\
& \quad + 2\lambda \mathbb{E} \left[\int_{nT}^t \left\langle \bar{\Theta}_s^\lambda - \tilde{\zeta}_s^{\lambda,n}, \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s \left(H(\bar{\Theta}_r^\lambda) - H(\bar{\Theta}_{[r]}) \right) dB_r^\lambda \right\rangle ds \right].
\end{aligned} \tag{162}$$

By using Lemma D.11 and (162) becomes

$$\begin{aligned}
\mathbb{E} \left[\left| \bar{\Theta}_t^\lambda - \tilde{\zeta}_t^{\lambda,n} \right|^2 \right] & \leq \lambda(2\bar{L}_3 + 5) \int_{nT}^t \mathbb{E} \left[\left| \bar{\Theta}_s^\lambda - \tilde{\zeta}_s^{\lambda,n} \right|^2 \right] ds \\
& \quad + 5\lambda^{2+q} \left(e^{-\bar{a}n/2} \bar{\mathcal{C}}_{S2} \mathbb{E}[|\theta_0|^4] + \tilde{\mathcal{C}}_{S2} \right) \\
& \quad + \bar{\mathfrak{J}}_1 + \bar{\mathfrak{J}}_2,
\end{aligned} \tag{163}$$

where

$$\begin{aligned}\bar{\mathfrak{J}}_1^\lambda(t) &:= 2\lambda \mathbb{E} \left[\int_{nT}^t \left\langle \bar{\Theta}_s^\lambda - \tilde{\zeta}_s^{\lambda,n}, \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s \left(H(\bar{\Theta}_r^\lambda) - H_\lambda(\bar{\Theta}_{[r]}^\lambda) - \mathfrak{M}(\bar{\Theta}_r^\lambda, \bar{\Theta}_{[r]}^\lambda) \right) dB_r^\lambda \right\rangle ds \right], \\ \bar{\mathfrak{J}}_2^\lambda(t) &:= 2\lambda \mathbb{E} \left[\int_{nT}^t \left\langle \bar{\Theta}_s^\lambda - \tilde{\zeta}_s^{\lambda,n}, \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s \mathfrak{M}(\bar{\Theta}_r^\lambda, \bar{\Theta}_{[r]}^\lambda) dB_r^\lambda \right\rangle ds \right]\end{aligned}$$

with \mathfrak{M} defined in Definition D.13. By using Young's inequality and Lemma D.14, we have that

$$\bar{\mathfrak{J}}_1^\lambda(t) \leq \lambda \int_{nT}^t \mathbb{E} \left[|\bar{\Theta}_s^\lambda - \tilde{\zeta}_s^{\lambda,n}|^2 \right] ds + \lambda^{2+q} \left(e^{-\bar{a}n/2} \bar{\mathfrak{C}}_{\mathbf{S}2} \mathbb{E}[|\theta_0|^4] + \tilde{\mathfrak{C}}_{\mathbf{S}2} \right). \quad (164)$$

To establish an upper bound for $\bar{\mathfrak{J}}_2^\lambda(t)$, we recall the definitions of $(\bar{\Theta}_t^\lambda)_{t \geq 0}$ and $(\tilde{\zeta}_t^{\lambda,n})_{t \geq 0}$ given in (111) and Definition D.2, respectively, and consider the following splitting:

$$\begin{aligned}\bar{\mathfrak{J}}_2^\lambda(t) &= 2\lambda \int_{nT}^t \mathbb{E} \left[\left\langle \bar{\Theta}_s^\lambda - \bar{\Theta}_{[s]}^\lambda - (\tilde{\zeta}_s^{\lambda,n} - \tilde{\zeta}_{[s]}^{\lambda,n}), \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s \mathfrak{M}(\bar{\Theta}_r^\lambda, \bar{\Theta}_{[r]}^\lambda) dB_r^\lambda \right\rangle ds \right. \\ &= 2\lambda \int_{nT}^t \mathbb{E} \left[\left\langle \int_{[s]}^s \left(\lambda(h(\tilde{\zeta}_r^{\lambda,n}) - h(\bar{\Theta}_{[r]}^\lambda)) + \lambda^2 \int_{[r]}^r \left(H(\bar{\Theta}_{[v]}^\lambda) h(\bar{\Theta}_{[v]}^\lambda) - \beta^{-1} \Upsilon(\bar{\Theta}_{[v]}^\lambda) \right) dv \right. \right. \\ &\quad \left. \left. - \lambda \sqrt{2\lambda\beta^{-1}} \int_{[r]}^r H(\bar{\Theta}_{[v]}^\lambda) dB_v^\lambda \right) dr, \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s \mathfrak{M}(\bar{\Theta}_r^\lambda, \bar{\Theta}_{[r]}^\lambda) dB_r^\lambda \right\rangle ds \right. \\ &= 2\lambda \int_{nT}^t \mathbb{E} \left[\left\langle \int_{[s]}^s \left(\lambda(h(\bar{\Theta}_r^\lambda) - h(\bar{\Theta}_{[r]}^\lambda)) + \lambda^2 \int_{[r]}^r \left(H(\bar{\Theta}_{[v]}^\lambda) h(\bar{\Theta}_{[v]}^\lambda) - \beta^{-1} \Upsilon(\bar{\Theta}_{[v]}^\lambda) \right) dv \right. \right. \\ &\quad \left. \left. - \lambda \sqrt{2\lambda\beta^{-1}} \int_{[r]}^r H(\bar{\Theta}_{[v]}^\lambda) dB_v^\lambda \right) dr, \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s \mathfrak{M}(\bar{\Theta}_r^\lambda, \bar{\Theta}_{[r]}^\lambda) dB_r^\lambda \right\rangle ds + \bar{\mathfrak{J}}_{2,1}^\lambda(t), \end{aligned}$$

where the first equality holds due to the following:

$$\begin{aligned}\mathbb{E} \left[\left\langle \bar{\Theta}_{[s]}^\lambda + \tilde{\zeta}_{[s]}^{\lambda,n}, \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s \mathfrak{M}(\bar{\Theta}_r^\lambda, \bar{\Theta}_{[r]}^\lambda) dB_r^\lambda \right\rangle \right] \\ = \mathbb{E} \left[\left\langle \bar{\Theta}_{[s]}^\lambda + \tilde{\zeta}_{[s]}^{\lambda,n}, \mathbb{E} \left[\sqrt{2\lambda\beta^{-1}} \int_{[s]}^s \mathfrak{M}(\bar{\Theta}_r^\lambda, \bar{\Theta}_{[r]}^\lambda) dB_r^\lambda \middle| \mathcal{F}_{[s]}^\lambda \right] \right\rangle \right] = 0, \end{aligned} \quad (165)$$

and where

$$\bar{\mathfrak{J}}_{2,1}^\lambda(t) = 2\lambda \int_{nT}^t \mathbb{E} \left[\left\langle \int_{[s]}^s \lambda(h(\tilde{\zeta}_r^{\lambda,n}) - h(\bar{\Theta}_r^\lambda)) dr, \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s \mathfrak{M}(\bar{\Theta}_r^\lambda, \bar{\Theta}_{[r]}^\lambda) dB_r^\lambda \right\rangle ds \right].$$

By applying Cauchy-Schwarz inequality, Corollary D.12 and Lemma D.14, we further obtain that

$$\begin{aligned}\bar{\mathfrak{J}}_2^\lambda(t) &\leq 2\lambda \int_{nT}^t \left(\mathbb{E} \left[\left| \int_{[s]}^s \lambda \left(h(\bar{\Theta}_r^\lambda) - h(\bar{\Theta}_{[r]}^\lambda) + \lambda \int_{[r]}^r \left(H(\bar{\Theta}_{[v]}^\lambda) h(\bar{\Theta}_{[v]}^\lambda) - \beta^{-1} \Upsilon(\bar{\Theta}_{[v]}^\lambda) \right) dv \right. \right. \right. \right. \\ &\quad \left. \left. \left. - \sqrt{2\lambda\beta^{-1}} \int_{[r]}^r H(\bar{\Theta}_{[v]}^\lambda) dB_v^\lambda \right) dr \right|^2 \right] \right)^{1/2} \\ &\quad \times \left(\mathbb{E} \left[\left| \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s \mathfrak{M}(\bar{\Theta}_r^\lambda, \bar{\Theta}_{[r]}^\lambda) dB_r^\lambda \right|^2 \right] \right)^{1/2} ds + \bar{\mathfrak{J}}_{2,1}^\lambda(t) \\ &\leq 2\lambda \int_{nT}^t \left(\lambda^4 \left(e^{-\bar{a}n/2} 36 \bar{\mathfrak{C}}_{\mathbf{S}2} \mathbb{E}[|\theta_0|^4] + 36 \tilde{\mathfrak{C}}_{\mathbf{S}2} \right) \right)^{1/2} \\ &\quad \times \left(\lambda^2 \left(e^{-\bar{a}n/2} \bar{\mathfrak{C}}_{\mathbf{S}2} \mathbb{E}[|\theta_0|^4] + \tilde{\mathfrak{C}}_{\mathbf{S}2} \right) \right)^{1/2} ds + \bar{\mathfrak{J}}_{2,1}^\lambda(t) \\ &\leq 12\lambda^3 \left(e^{-\bar{a}n/2} \bar{\mathfrak{C}}_{\mathbf{S}2} \mathbb{E}[|\theta_0|^4] + \tilde{\mathfrak{C}}_{\mathbf{S}2} \right) + \bar{\mathfrak{J}}_{2,1}^\lambda(t). \end{aligned} \quad (166)$$

At this stage, the task reduces to upper bound $\bar{\mathfrak{J}}_{2,1}^\lambda(t)$. To achieve this, we apply Cauchy-Schwarz inequality, Corollary D.12 and Lemma D.14 to obtain

$$\begin{aligned}
\bar{\mathfrak{J}}_{2,1}^\lambda(t) &= 2\lambda \int_{nT}^t \mathbb{E} \left[\left\langle \int_{[s]}^s \lambda \left(h(\tilde{\zeta}_r^{\lambda,n}) - h(\tilde{\zeta}_{[r]}^{\lambda,n}) - \sqrt{2\lambda\beta^{-1}} \int_{[r]}^r H(\tilde{\zeta}_{[\nu]}^{\lambda,n}) dB_\nu^\lambda - \left(h(\bar{\Theta}_r^\lambda) \right. \right. \right. \\
&\quad \left. \left. \left. - h(\bar{\Theta}_{[r]}^\lambda) - \sqrt{2\lambda\beta^{-1}} \int_{[r]}^r H(\bar{\Theta}_{[\nu]}^\lambda) dB_\nu^\lambda \right) dr, \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s \mathfrak{M}(\bar{\Theta}_r^\lambda, \bar{\Theta}_{[r]}^\lambda) dB_r^\lambda \right\rangle \right] ds \\
&\quad + 2\lambda \int_{nT}^t \mathbb{E} \left[\left\langle \int_{[s]}^s \lambda \sqrt{2\lambda\beta^{-1}} \int_{[r]}^r (H(\tilde{\zeta}_{[\nu]}^{\lambda,n}) - H(\bar{\Theta}_{[\nu]}^\lambda)) dB_\nu^\lambda dr, \right. \right. \\
&\quad \left. \left. \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s \mathfrak{M}(\bar{\Theta}_r^\lambda, \bar{\Theta}_{[r]}^\lambda) dB_r^\lambda \right\rangle \right] ds \\
&\leq 2\lambda \int_{nT}^t \left(\mathbb{E} \left[\left| \int_{[s]}^s \lambda \left(h(\tilde{\zeta}_r^{\lambda,n}) - h(\tilde{\zeta}_{[r]}^{\lambda,n}) - \sqrt{2\lambda\beta^{-1}} \int_{[r]}^r H(\tilde{\zeta}_{[\nu]}^{\lambda,n}) dB_\nu^\lambda - \left(h(\bar{\Theta}_r^\lambda) \right. \right. \right. \right. \right. \\
&\quad \left. \left. \left. - h(\bar{\Theta}_{[r]}^\lambda) - \sqrt{2\lambda\beta^{-1}} \int_{[r]}^r H(\bar{\Theta}_{[\nu]}^\lambda) dB_\nu^\lambda \right) dr \right|^2 \right] \right)^{1/2} \\
&\quad \times \left(\mathbb{E} \left[\left| \sqrt{2\lambda\beta^{-1}} \int_{[s]}^s \mathfrak{M}(\bar{\Theta}_r^\lambda, \bar{\Theta}_{[r]}^\lambda) dB_r^\lambda \right|^2 \right] \right)^{1/2} ds \\
&\quad + 2\lambda^3 \left(e^{-\bar{a}n/4} 5\bar{\mathfrak{C}}_{\mathbf{S}2} \mathbb{E}[|\theta_0|^4] + 5\tilde{\mathfrak{C}}_{\mathbf{S}2} \right) \\
&\leq 2\lambda \int_{nT}^t \left(\lambda^4 \left(e^{-\bar{a}n/4} 72\bar{\mathfrak{C}}_{\mathbf{S}2} \mathbb{E}[|\theta_0|^4] + 72\tilde{\mathfrak{C}}_{\mathbf{S}2} \right) \right)^{1/2} \left(\lambda^2 \left(e^{-\bar{a}n/2} \bar{\mathfrak{C}}_{\mathbf{S}2} \mathbb{E}[|\theta_0|^4] + \tilde{\mathfrak{C}}_{\mathbf{S}2} \right) \right)^{1/2} ds \\
&\quad + 10\lambda^3 \left(e^{-\bar{a}n/4} \bar{\mathfrak{C}}_{\mathbf{S}2} \mathbb{E}[|\theta_0|^4] + \tilde{\mathfrak{C}}_{\mathbf{S}2} \right) \\
&\leq 34\lambda^3 \left(e^{-\bar{a}n/4} \bar{\mathfrak{C}}_{\mathbf{S}2} \mathbb{E}[|\theta_0|^4] + \tilde{\mathfrak{C}}_{\mathbf{S}2} \right). \tag{167}
\end{aligned}$$

Substituting (167) into (166) yields

$$\bar{\mathfrak{J}}_2^\lambda(t) \leq 46\lambda^3 \left(e^{-\bar{a}n/4} \bar{\mathfrak{C}}_{\mathbf{S}2} \mathbb{E}[|\theta_0|^4] + \tilde{\mathfrak{C}}_{\mathbf{S}2} \right). \tag{168}$$

By applying (164) and (168) to (163), we obtain that

$$\mathbb{E} \left[\left| \bar{\Theta}_t^\lambda - \tilde{\zeta}_t^{\lambda,n} \right|^2 \right] \leq \lambda(2\bar{L}_3 + 6) \int_{nT}^t \mathbb{E} \left[\left| \bar{\Theta}_s^\lambda - \tilde{\zeta}_s^{\lambda,n} \right|^2 \right] ds + 52\lambda^{2+q} \left(e^{-\bar{a}n/4} \bar{\mathfrak{C}}_{\mathbf{S}2} \mathbb{E}[|\theta_0|^4] + \tilde{\mathfrak{C}}_{\mathbf{S}2} \right),$$

which, by applying Grönwall's lemma, yields

$$W_2^2(\mathcal{L}(\bar{\Theta}_t^\lambda), \mathcal{L}(\tilde{\zeta}_t^{\lambda,n})) \leq \mathbb{E} \left[\left| \bar{\Theta}_t^\lambda - \tilde{\zeta}_t^{\lambda,n} \right|^2 \right] \leq \lambda^{2+q} \left(e^{-\bar{a}n/4} \mathfrak{C}_{\text{Lin},0} \mathbb{E}[|\theta_0|^4] + \mathfrak{C}_{\text{Lin},1} \right),$$

where

$$\mathfrak{C}_{\text{Lin},0} := 52e^{2\bar{L}_3+6} \bar{\mathfrak{C}}_{\mathbf{S}2}, \quad \mathfrak{C}_{\text{Lin},1} := 52e^{2\bar{L}_3+6} \tilde{\mathfrak{C}}_{\mathbf{S}2} \tag{169}$$

with $\bar{\mathfrak{C}}_{\mathbf{S}2}, \tilde{\mathfrak{C}}_{\mathbf{S}2}$ given in (153). \square

REFERENCES

- [1] O Deniz Akyildiz and Sotirios Sabanis. Nonasymptotic analysis of Stochastic Gradient Hamiltonian Monte Carlo under local conditions for nonconvex optimization. *Journal of Machine Learning Research*, 25(113):1–34, 2024.
- [2] Mathias Barkhagen, Ngoc Huy Chau, Éric Moulines, Miklós Rásonyi, Sotirios Sabanis, and Ying Zhang. On stochastic gradient Langevin dynamics with dependent data streams in the logconcave case. *Bernoulli*, 27(1):1–33, 2021.

- [3] Nicolas Brosse, Alain Durmus, and Eric Moulines. The promises and pitfalls of stochastic gradient Langevin dynamics. In *Advances in Neural Information Processing Systems*, pages 8268–8278, 2018.
- [4] Nicolas Brosse, Alain Durmus, Éric Moulines, and Sotirios Sabanis. The Tamed Unadjusted Langevin Algorithm. *Stochastic Processes and their Applications*, 129(10):3638–3663, 2019.
- [5] Huy N Chau and Miklos Rasonyi. Stochastic Gradient Hamiltonian Monte Carlo for Non-convex Learning. *Stochastic Processes and their Applications*, 149:341–368, 2022.
- [6] Ngoc Huy Chau, Éric Moulines, Miklos Rásonyi, Sotirios Sabanis, and Ying Zhang. On stochastic gradient Langevin dynamics with dependent data streams: The fully nonconvex case. *SIAM Journal on Mathematics of Data Science*, 3(3):959–986, 2021.
- [7] Xiang Cheng, Niladri S Chatterji, Yasin Abbasi-Yadkori, Peter L Bartlett, and Michael I Jordan. Sharp convergence rates for Langevin dynamics in the nonconvex setting. *arXiv preprint arXiv:1805.01648*, 2018.
- [8] Xiang Cheng, Niladri S Chatterji, Peter L Bartlett, and Michael I Jordan. Underdamped langevin mcmc: A non-asymptotic analysis. In *Conference on learning theory*, pages 300–323. PMLR, 2018.
- [9] Sinho Chewi, Murat A Erdogdu, Mufan Li, Ruoqi Shen, and Matthew S Zhang. Analysis of langevin monte carlo from poincare to log-sobolev. *Foundations of Computational Mathematics*, pages 1–51, 2024.
- [10] Arnak S Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3): 651–676, 2017.
- [11] Arnak S Dalalyan and Avetik Karagulyan. User-friendly guarantees for the Langevin Monte Carlo with inaccurate gradient. *Stochastic Processes and their Applications*, 2019.
- [12] Arnak S Dalalyan and Lionel Riou-Durand. On sampling from a log-concave density using kinetic Langevin diffusions. *Bernoulli*, 26(3):1956–1988, 2020.
- [13] Alain Durmus and Eric Moulines. High-dimensional Bayesian inference via the unadjusted Langevin algorithm. *Bernoulli*, 25(4A):2854–2882, 2019.
- [14] Alain Durmus, Szymon Majewski, and Błażej Miasojedow. Analysis of Langevin Monte Carlo via convex optimization. *The Journal of Machine Learning Research*, 20(1):2666–2711, 2019.
- [15] Andreas Eberle. Reflection couplings and contraction rates for diffusions. *Probability theory and related fields*, 166(3-4):851–886, 2016.
- [16] Andreas Eberle, Arnaud Guillin, and Raphael Zimmer. Quantitative Harris-type theorems for diffusions and McKean–Vlasov processes. *Transactions of the American Mathematical Society*, 371(10):7135–7173, 2019.
- [17] Murat A Erdogdu and Rasa Hosseinzadeh. On the convergence of langevin monte carlo: The interplay between tail growth and smoothness. In *Conference on Learning Theory*, pages 1776–1822. PMLR, 2021.
- [18] Wenjiang J Fu. Penalized regressions: the bridge versus the lasso. *Journal of computational and graphical statistics*, 7(3):397–416, 1998.
- [19] Xuefeng Gao, Mert Gürbüzbalaban, and Lingjiong Zhu. Global Convergence of Stochastic Gradient Hamiltonian Monte Carlo for Nonconvex Stochastic Optimization: Nonasymptotic Performance Bounds and Momentum-Based Acceleration. *Operations Research*, 70(5):2931–2947, 2022.
- [20] Martin Hutzenthaler, Arnulf Jentzen, and Peter E Kloeden. Strong and weak divergence in finite time of Euler’s method for stochastic differential equations with non-globally Lipschitz continuous coefficients. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 467(2130):1563–1576, 2011.
- [21] Martin Hutzenthaler, Arnulf Jentzen, and Peter E Kloeden. Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients. *The Annals of Applied Probability*, 22(4):1611–1641, 2012.
- [22] Peter E Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations*, volume 23. Springer Science & Business Media, 2013.
- [23] N. V. Krylov. A Simple Proof of the Existence of a Solution of Itô’s Equation with Monotone Coefficients. *Theory of Probability & Its Applications*, 35(3):583–587, 1991. doi: 10.1137/1135082.
- [24] Luxu Liang, Ariel Neufeld, and Ying Zhang. Non-asymptotic convergence analysis of the stochastic gradient hamiltonian monte carlo algorithm with discontinuous stochastic gradient with applications

- to training of relu neural networks. *arXiv preprint arXiv:2409.17107*, 2024.
- [25] Dong-Young Lim and Sotirios Sabanis. Polygonal Unadjusted Langevin Algorithms: Creating stable and efficient adaptive algorithms for neural networks. *Journal of Machine Learning Research*, 2024.
- [26] Dong-Young Lim, Ariel Neufeld, Sotirios Sabanis, and Ying Zhang. Langevin dynamics based algorithm e-TH ϵ O POULA for stochastic optimization problems with discontinuous stochastic gradient. *arXiv preprint arXiv:2210.13193*, 2022.
- [27] Dong-Young Lim, Ariel Neufeld, Sotirios Sabanis, and Ying Zhang. Non-asymptotic estimates for TUSLA algorithm for non-convex learning with applications to neural networks with ReLU activation function. *IMA Journal of Numerical Analysis*, 2023.
- [28] Wei Liu and Xuerong Mao. Strong convergence of the stopped Euler–Maruyama method for nonlinear stochastic differential equations. *Applied Mathematics and Computation*, 223:389–400, 2013.
- [29] Attila Lovas, Iosif Lytras, Miklós Rásonyi, and Sotirios Sabanis. Taming neural networks with tusla: Nonconvex learning via adaptive stochastic gradient langevin algorithms. *SIAM Journal on Mathematics of Data Science*, 5(2):323–345, 2023.
- [30] Xuerong Mao. The truncated Euler–Maruyama method for stochastic differential equations. *Journal of Computational and Applied Mathematics*, 290:370–384, 2015.
- [31] Wenlong Mou, Yi-An Ma, Martin J Wainwright, Peter L Bartlett, and Michael I Jordan. High-order Langevin diffusion yields an accelerated MCMC algorithm. *Journal of Machine Learning Research*, 22(42):1–41, 2021.
- [32] Wenlong Mou, Nicolas Flammarion, Martin J Wainwright, and Peter L Bartlett. Improved bounds for discretization of Langevin diffusions: Near-optimal rates without convexity. *Bernoulli*, 28(3):1577–1601, 2022.
- [33] Ariel Neufeld, Matthew Ng Cheng En, and Ying Zhang. Non-asymptotic convergence bounds for modified tamed unadjusted Langevin algorithm in non-convex setting. *arXiv preprint arXiv:2207.02600*, 2022.
- [34] Ariel Neufeld, Matthew Ng Cheng En, and Ying Zhang. Robust SGLD algorithm for solving non-convex distributionally robust optimisation problems. *arXiv preprint arXiv:2403.09532*, 2024.
- [35] Dao Nguyen. Unadjusted langevin algorithm for sampling a mixture of weakly smooth potentials. *Brazilian Journal of Probability and Statistics*, 36(3):504–539, 2022.
- [36] François Orieux, Olivier Féron, and J-F Giovannelli. Sampling high-dimensional gaussian distributions for general linear inverse problems. *IEEE Signal Processing Letters*, 19(5):251–254, 2012.
- [37] Eckhard Platen and Wolfgang Wagner. On a Taylor formula for a class of Itô processes. *Probability and Mathematical Statistics*, 3(1):37–51, 1982.
- [38] Maxim Raginsky, Alexander Rakhlin, and Matus Telgarsky. Non-convex learning via Stochastic Gradient Langevin Dynamics: a nonasymptotic analysis. In *Proceedings of the 2017 Conference on Learning Theory*, volume 65 of *Proceedings of Machine Learning Research*, pages 1674–1703, Amsterdam, Netherlands, 07–10 Jul 2017. PMLR.
- [39] Sotirios Sabanis. A note on tamed Euler approximations. *Electronic Communications in Probability*, 18:1–10, 2013.
- [40] Sotirios Sabanis and Ying Zhang. Higher order Langevin Monte Carlo algorithm. *Electronic Journal of Statistics*, 13(2):3805 – 3850, 2019. doi: 10.1214/19-EJS1615. URL <https://doi.org/10.1214/19-EJS1615>.
- [41] Sotirios Sabanis and Ying Zhang. On explicit order 1.5 approximations with varying coefficients: the case of super-linear diffusion coefficients. *Journal of Complexity*, 50:84–115, 2019.
- [42] Sotirios Sabanis and Ying Zhang. A fully data-driven approach to minimizing CVaR for portfolio of assets via SGLD with discontinuous updating. *arXiv preprint arXiv:2007.01672*, 2020.
- [43] Santosh Vempala and Andre Wibisono. Rapid convergence of the unadjusted langevin algorithm: Isoperimetry suffices. *Advances in neural information processing systems*, 32, 2019.
- [44] Maxime Vono, Nicolas Dobigeon, and Pierre Chainais. High-dimensional gaussian sampling: a review and a unifying approach based on a stochastic proximal point algorithm. *SIAM Review*, 64(1):3–56, 2022.

- [45] Pan Xu, Jinghui Chen, Difan Zou, and Quanquan Gu. Global convergence of Langevin dynamics based algorithms for nonconvex optimization. In *Advances in Neural Information Processing Systems*, pages 3122–3133, 2018.
- [46] Ying Zhang, Ömer Deniz Akyildiz, Theodoros Damoulas, and Sotirios Sabanis. Nonasymptotic estimates for stochastic gradient Langevin dynamics under local conditions in nonconvex optimization. *Applied Mathematics & Optimization*, 87(2), 2023.

DIVISION OF MATHEMATICAL SCIENCES, NANYANG TECHNOLOGICAL UNIVERSITY, 21 NANYANG LINK, 637371 SINGAPORE

Email address: ariel.neufeld@ntu.edu.sg

FINANCIAL TECHNOLOGY THRUST, SOCIETY HUB, THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU), GUANGZHOU, CHINA

Email address: yingzhang@hkust-gz.edu.cn