On stochastic gradient Langevin dynamics with dependent data streams

On stochastic gradient Langevin dynamics with dependent data streams: the fully non-convex case *

N. H. Chau [1], É. Moulines[2], M. Rásonyi [1], S. Sabanis[3, 4], and Y. Zhang[3]

[1]Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences, Hungary.
[2]Centre de Mathématiques Appliquées, UMR 7641, Ecole Polytechnique, France.
[3]School of Mathematics, The University of Edinburgh, UK.
[4]The Alan Turing Institute, UK.

May 31, 2019

Abstract

We consider the problem of sampling from a target distribution which is *not necessarily logconcave.* Non-asymptotic analysis results are established in a suitable Wasserstein-type distance of the Stochastic Gradient Langevin Dynamics (SGLD) algorithm, when the gradient is driven by even *dependent* data streams. Our estimates are sharper and *uniform* in the number of iterations, in contrast to those in previous studies.

1 Introduction

In this paper, the problem of sampling from a target distribution

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

is investigated, where $\theta \in \mathbb{R}^d$ and the function $U : \mathbb{R}^d \to \mathbb{R}$ satisfies Lipschitz continuity and a certain dissipativity condition. We establish non-asymptotic convergence rates for the Stochastic Gradient Langevin Dynamics (SGLD) algorithm, based on the stochastic differential equation

$$dx_t = -\beta\nabla U(x_t)dt + \sqrt{2\beta^{-1}}dB_t \tag{1}$$

where B is the standard Brownian motion in $\mathbb{R}^d$ and $\beta \in \mathbb{R}_+$ is the inverse temperature parameter.

Non-asymptotic convergence rates of Langevin dynamics based algorithms for approximate sampling of log-concave distributions have been intensively studied in recent years, starting with [7]. This was followed by [9], [13], [12], [5] amongst others.

Relaxing log-concavity is a more challenging problem. In [21], the log-concavity assumption is replaced by a logconcavity at infinity condition and L^1 and L^2-Wasserstein distances convergence rates are obtained. In a similar setting, [6] analyzes sampling errors in the L^1-Wasserstein distance for both overdamped and underdamped Langevin MCMC. In [23], only a dissipativity condition is assumed and convergence rates are obtained in the L^2-Wasserstein distance. Moreover, a clear and strong link between sampling via SGLD algorithms and non-convex optimization is highlighted. One can further consult [27], [8] and references therein.

In the present paper, we impose the dissipativity condition as in [23]. Using a different Wasserstein-type metric, we obtain shaper estimates and allow for possibly dependent data sequences. The key new idea is that we compare the SGLD algorithm to a suitable auxiliary continuous time processes inspired by (1) and we rely on contraction results developed in [15] for (1).

*All the authors were supported by The Alan Turing Institute, London under the EPSRC grant EP/N510129/1. N. H. C. and M. R. also enjoyed the support of the NKFIH (National Research, Development and Innovation Office, Hungary) grant KH 126505 and the "Lendület" grant LP 2015-6 of the Hungarian Academy of Sciences. Y. Z. was supported by The Maxwell Institute Graduate School in Analysis and its Applications, a Centre for Doctoral Training funded by the UK Engineering and Physical Sciences Research Council (grant EP/L016508/01), the Scottish Funding Council, Heriot-Watt University and the University of Edinburgh. We thank the Alan Turing Institute, London, UK; the Rényi Institute, Budapest, Hungary and the École Polytechnique, Palaiseau, France for hosting research meetings of the authors.

2 Main results

Let $(\Omega, \mathcal{F}, P)$ be a probability space. We denote by $\mathbb{E}[X]$ the expectation of a random variable X. For $1 \leq p < \infty$, L^p is used to denote the usual space of p-integrable real-valued random variables. Fix an integer $d \geq 1$. For an $\mathbb{R}^d$-valued random variable X, its law on $\mathcal{B}(\mathbb{R}^d)$ (the Borel sigma-algebra of $\mathbb{R}^d$) is denoted by $\mathcal{L}(X)$. 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). We fix a discrete-time filtration $\mathcal{G}_n := \sigma(\varepsilon_k,\ k \leq n,\ k \in \mathbb{Z})$, $n \in \mathbb{Z}$ where $(\varepsilon_n)_{n\in\mathbb{Z}}$ is an i.i.d. sequence with values in some Polish space. This represents the flow of past information. The notation $\mathcal{G}_\infty$ is self-explanatory. We also define the decreasing sequence of sigma-algebras $\mathcal{G}_n^+ := \sigma(\varepsilon_k,\ k > n)$, $n \in \mathbb{Z}$, representing future information at the respective time instants.

Fix an $\mathbb{R}^d$-valued random variable θ_0, representing the initial value of the procedure we consider. For each $\beta, \lambda > 0$, define the $\mathbb{R}^d$-valued random process θ_n^λ, $n \in \mathbb{N}$ by recursion:

$$\theta_0^\lambda := \theta_0, \quad \theta_{n+1}^\lambda := \theta_n^\lambda - \lambda H(\theta_n^\lambda, X_{n+1}) + \sqrt{\frac{2\lambda}{\beta}}\xi_{n+1},\ n \in \mathbb{N}, \tag{2}$$

where $H : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^d$ is a measurable function, X_n, $n \in \mathbb{N}$ is an $\mathbb{R}^m$-valued, $(\mathcal{G}_n)_{n\in\mathbb{N}}$-adapted process and ξ_n, $n \in \mathbb{N}$ is an independent sequence of standard d-dimensional Gaussian random variables.

We interpret X_n, $n \in \mathbb{N}$ as a stream of data and ξ_n, $n \in \mathbb{N}$ as an artificially generated noise sequence. We assume throughout the paper that θ_0, $\mathcal{G}_\infty$ and $(\xi_n)_{n\in\mathbb{N}}$ are independent.

Let $U : \mathbb{R}^d \to \mathbb{R}_+$ be continuously differentiable with derivative $h := \nabla U$. Let us define the probability

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

It is implicitly assumed that $\int_{\mathbb{R}^d} \mathrm{e}^{-\beta U(\theta)}\,\mathrm{d}\theta < \infty$ and this is indeed the case under Assumption 2.5 below, as easily seen. Our objective is to (approximately) sample from the distribution π_β using the scheme (2).

We now present our assumptions. First, the moments of the initial condition need to be controlled.

Assumption 2.1.

$$|\theta_0| \in \cap_{p\geq 1} L^p.$$

Next, we require joint Lipschitz-continuity of H.

Assumption 2.2. *There is $K_1 < \infty$ and $K_2 < \infty$ such that for all $\theta, \theta' \in \mathbb{R}^d$ and $x, x' \in \mathbb{R}^m$,*

$$|H(\theta, x) - H(\theta', x')| \leq K_1|\theta - \theta'| + K_2|x - x'|,$$

We set

$$H^* := |H(0, 0)|. \tag{3}$$

The data sequence X_n, $n \in \mathbb{N}$ need not be i.i.d., we require only a mixing property, defined in Section 3.1 below.

Assumption 2.3. *The process X_n, $n \in \mathbb{N}$ is conditionally L-mixing with respect to $(\mathcal{G}_n, \mathcal{G}_n^+)_{n\in\mathbb{N}}$. It satisfies*

$$\mathbb{E}[H(\theta, X_n)] = h(\theta),\ \theta \in \mathbb{R}^d,\ n \in \mathbb{N}. \tag{4}$$

Remark 2.4. Stationarity of the process X_n, $n \in \mathbb{N}$ would also be natural to assume but we need only the weaker property (4).

Finally, we present a dissipativity condition on H.

Assumption 2.5. *There exist a, $b > 0$ such that, for all $\theta \in \mathbb{R}^d$ and $x \in \mathbb{R}^m$,*

$$\langle H(\theta, x), \theta \rangle \geq a|\theta|^2 - b. \tag{5}$$

When $X_n = c$ for all $n \in \mathbb{N}$ for some $c \in \mathbb{R}^m$ (i.e. when $H(\theta, X_t)$ is replaced by $h(\theta)$ in (2)) then we arrive at the well-known unadjusted Langevin algorithm whose convergence properties have been amply analyzed, see e.g. [7, 12, 6, 21]. The case of i.i.d. X_n, $n \in \mathbb{N}$ has also been investigated in great detail, see e.g. [23, 27, 21].

In the present article, better estimates are obtained for the distance between $\mathcal{L}(\theta_n^\lambda)$ and π_β than those of [23] and [27]. Such rates have already been obtained in [2] for strongly convex U and in [21] for U that is convex outside a compact set. Here we make no convexity assumptions at all. This comes at the price of using the metric W_1 defined in (6) below while [23, 27, 21, 2] use Wasserstein distances with respect to the standard Euclidean metric, see (10) below.

Another novelty of our paper is that, just like in [2], we allow the data sample X_n, $n \in \mathbb{N}$ to be dependent. As observed data have no reason to be i.i.d., we believe that such a result is fundamental to assure the robustness of the sampling method based on the stochastic gradient Langevin dynamics (2).

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 $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, 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. Define

$$W_1(\mu, \nu) := \inf_{\zeta \in \mathcal{C}(\mu,\nu)} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} [|x-y| \wedge 1] \zeta(dx, dy), \tag{6}$$

which is the Wasserstein-1 distance associated to the bounded metric $|x-y| \wedge 1$, $x, y \in \mathbb{R}^d$.

Remark 2.6. In this work, the constants appearing are often denoted by C_j for some natural number $j \in \mathbb{N}$. Without further mention, these constants depend on θ_0, K_1, K_2, a, b, H^*, β, d and on the process X_n, $n \in \mathbb{N}$ through the quantities (13) below and, unless otherwise stated, they do not depend on anything else. In case of further dependencies (e.g. dependence on p, which is due to the drift condition, coming from Lemma 3.6 below), we signal these in parentheses, e.g. $C_6(p)$.

Our main contribution is summarized in the following result. Define

$$\lambda_{\max} = \min\{a/2K_1^2, 1/a\}\,. \tag{7}$$

Theorem 2.7. *Let Assumptions 2.1, 2.2, 2.3 and 2.5 be valid. Then there are finite constants $C_0 > 0$, C_1, C_2 such that, for every $0 < \lambda \leq \lambda_{\max}$.*

$$W_1(\mathcal{L}(\theta_n^\lambda), \pi) \leq C_1 \mathrm{e}^{-C_0 \lambda n} + C_2 \sqrt{\lambda}, \ n \in \mathbb{N} \tag{8}$$

Example 3.4 of [2] suggests that the best rate we can hope to get in (8) is $\sqrt{\lambda}$, even in the convex case. The above theorem achieves this rate. We remark that, although the statement of Theorem 2.7 concerns the discrete-time recursive scheme (2), its proof is carried out entirely in a continuous-time setting, in Section 3. It relies on techniques from [2] and [15]. The principal new idea is the introduction of the auxiliary process $\tilde{Y}_t^\lambda(\mathbf{x})$, $t \in \mathbb{R}_+$, see (25) below.

Consider now a strengthening of Assumption 2.5 by imposing convexity outside a compact set.

Assumption 2.8. *There exist $b, a > 0$ such that, for each θ, θ' satisfying $|\theta - \theta'| > b$,*

$$\langle H(\theta, x) - H(\theta', x), \theta - \theta' \rangle \geq a|\theta - \theta'|^2, \ x \in \mathbb{R}^m. \tag{9}$$

Then, we can recover analogous results to Theorem 2.7 by considering the L^1-Wasserstein distance. At this point, let us recall the definition of the familiar, "usual" Wasserstein-p (also know as L^p-Wasserstein) distance, for $p \geq 1$:

$$\tilde{W}_p(\mu, \nu) := \inf_{\zeta \in \mathcal{C}(\mu,\nu)} \left(\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |x-y|^p \zeta(dx, dy) \right)^{1/p}, \ \mu, \nu \in \mathcal{P}(\mathbb{R}^d). \tag{10}$$

Theorem 2.9. *Let Assumptions 2.1, 2.2, 2.3 and 2.8 be valid. Then there are constants $C_3, C_4, C_5 > 0$ such that, for every $0 < \lambda \leq \lambda_{\max}$,*

$$\tilde{W}_1(\mathcal{L}(\theta_n^\lambda), \pi) \leq C_4 \mathrm{e}^{-C_3 \lambda n} + C_5 \sqrt{\lambda}, \ n \in \mathbb{N}. \tag{11}$$

Strengthening the monotonicity condition (9) even guarantees convergence in $\tilde{W}_2$.

Assumption 2.10. *There exists $a > 0$ such that, for each θ, $\theta' \in \mathbb{R}^d$,*

$$\langle H(\theta, x) - H(\theta', x), \theta - \theta' \rangle \geq a[|\theta - \theta'|^2 + |H(\theta, x) - H(\theta', x)|^2], \ x \in \mathbb{R}^m.$$

Theorem 2.11. *Let Assumptions 2.1, 2.2, 2.3 and 2.10 be valid. Then there are constants $C_3', C_4', C_5' > 0$ such that*

$$\tilde{W}_2(\mathcal{L}(\theta_n^\lambda), \pi) \leq C_4' \mathrm{e}^{-C_3' \lambda n} + C_5' \sqrt{\lambda}, \ n \in \mathbb{N} \tag{12}$$

holds for every $0 < \lambda \leq \lambda_{\max}$.

2.1 Related work and our contributions

In the remarkable paper [23], a non-convex optimization problem is considered in the context of empirical risk minimization, which plays a central role in ML algorithms. The excess risk is decomposed into a sampling error resulting from the application of SGLD, a generalization error and a suboptimality error. Our aim is to improve the sampling error in the non-convex setting and provide sharper convergence estimates under more relaxed conditions. To this end, we focus on the comparison of our results with Proposition 3.3 of [23].

Condition $(A.5)$ of [23] is (much) stronger than Assumption 2.1 above. Assumption 2.5 is identical to $(A.3)$ in [23]. Condition $(A.2)$ in [23] corresponds to Lipschitz-continuity of H in its first variable with a Lipschitz-constant independent from its second variable and $(A.1)$ there means that $H(0,\cdot)$, $u(0,\cdot)$ are bounded where $U(\theta) = \mathbb{E}[u(\theta, X_0)]$ and $H(\cdot,\cdot) = \partial_\theta u(\cdot,\cdot)$. Hence Assumption 2.2 here is neither stronger nor weaker than $(A.2)$ of [23], they are incomparable conditions. In any case, Assumption 2.2 does not seem to be restrictive for practical purposes. Condition $(A.4)$ in [23] is implied by Assumptions 2.2 and 2.3.

We obtain stronger rates (which we believe to be optimal) than those of [23]. More precisely, we obtain a rate $\lambda^{1/2}$ in (8) for the W_1 distance while [23] only obtains $\lambda^{5/4}n$ (which depends on n) but in the $\tilde{W}_2$ distance. Furthermore, we allow a possibly dependent data sequence. In other words, [23] is applicable only if $X_n, n \in \mathbb{N}$ is i.i.d. while Assumption 2.3 suffices for the derivation of our results.

Now let us turn to [21]. The comparison is made only in the presence of convexity (outside a compact set) for U as it is a requirement for the results in [21]. Their Assumption 1.1 is precisely Assumptions 2.2 and 2.8 combined, however this is stipulated for h in [21] while we need it for $H(\cdot, x)$, for all x, as we allow dependent data streams. Furthermore, Assumption 1.3 in [21] requires that the variance of $H(\theta, X_0)$ is controlled by a power of the step size λ while we do not need such an assumption. The second conclusion of their Theorem 1.4 (with $\alpha = 1$, using their notation α) is the same as our Theorem 2.9.

Remark 2.12. In the particular case where X_n, $n \in \mathbb{N}$ are i.i.d., one can replace Theorem 3.2 below by Doob's inequality in the arguments for proving Theorem 2.7. The full power of Assumption 2.2 is used only in Lemma 3.14. When X_n are i.i.d. then Lemma 3.14 is trivial and it is enough to assume only (A.2) of [23] instead of Assumption 2.2.

3 Proofs

3.1 Conditional L-mixing

L-mixing processes and random fields were introduced in [16]. In [4], the closely related concept of *conditional* L-mixing was created. We define this concept below and recall some related results. This section is an almost exact replica of Section 2 in [2].

We assume that the probability space is equipped with a discrete-time filtration $\mathcal{R}_n$, $n \in \mathbb{N}$ as well as with a decreasing sequence of sigma-fields $\mathcal{R}_n^+$, $n \in \mathbb{N}$ such that $\mathcal{R}_n$ is independent of $\mathcal{R}_n^+$, for all n.

Fix an integer $q \geq 1$ and let $D \subset \mathbb{R}^q$ be a set of parameters. A measurable function $U : \mathbb{N} \times D \times \Omega \to \mathbb{R}^m$ is called a random field. We drop the dependence on $\omega \in \Omega$ in the notation henceforth and write $(U_n(\theta))_{n\in\mathbb{N},\theta\in D}$. A random process $(U_n)_{n\in\mathbb{N}}$ corresponds to a random field where D is a singleton. A random field is L^r-*bounded* for some $r \geq 1$ if

$$\sup_{n\in\mathbb{N}} \sup_{\theta\in D} \mathbb{E}^{1/r}[|U_n(\theta)|^r] < \infty.$$

Now we define conditional L-mixing. Recall that, for any family Z_i, $i \in I$ of real-valued random variables, $\operatorname{ess\,sup}_{i\in I} Z_i$ denotes a random variable that is an almost sure upper bound for each Z_i and it is, almost surely, smaller than or equal to any other such bound.

Let $(U_n(\theta))_{n\in\mathbb{N},\theta\in D}$ be L^r-bounded for each $r \geq 1$. Define, for each $n \in \mathbb{N}$, and for $\tau \geq 0$,

$$\begin{aligned} M_r^n(U) &:= \operatorname*{ess\,sup}_{\theta\in D} \sup_{m\in\mathbb{N}} \mathbb{E}^{1/r}\left[|U_{n+m}(\theta)|^r \,|\, \mathcal{R}_n\right] \\ \gamma_r^n(\tau, U) &:= \operatorname*{ess\,sup}_{\theta\in D} \sup_{m\geq\tau} \mathbb{E}^{1/r}\left[|U_{n+m}(\theta) - \mathbb{E}\left[U_{n+m}(\theta) \,|\, \mathcal{R}_{n+m-\tau}^+ \vee \mathcal{R}_n\right]|^r \,\middle|\, \mathcal{R}_n\right], \\ \Gamma_r^n(U) &:= \sum_{\tau=0}^{\infty} \gamma_r^n(\tau, U). \end{aligned} \tag{13}$$

When necessary, $M_r^n(U,D)$, $\gamma_r^n(\tau,U,D)$ and $\Gamma_r^n(U,D)$ are used to emphasize dependence of these quantities on the domain D which may vary.

Definition 3.1 (Conditional L-mixing)**.** *We call* $(U_n(\theta))_{n\in\mathbb{N},\theta\in D}$ uniformly conditionally L-mixing *(UCLM) with respect to* $(\mathcal{R}_n, \mathcal{R}_n^+)_{n\in\mathbb{N}}$ *if* $(U_n(\theta))_{n\in\mathbb{N}}$ *is adapted to* $(\mathcal{R}_n)_{n\in\mathbb{N}}$ *for all* $\theta \in D$*; for all* $r \geq 1$*, it is* L^r*-bounded; and the sequences* $(M_r^n(U))_{n\in\mathbb{N}}$*,* $(\Gamma_r^n(U))_{n\in\mathbb{N}}$ *are also* L^r*-bounded for all* $r \geq 1$*. In the case of stochastic processes (when* D *is a singleton) the terminology "conditionally* L*-mixing process" is used.*

Conditionally L-mixing encompasses a broad class of processes (linear processes, functionals of Markov processes, etc.), see Example 2.1 in [2]. The following maximal inequality is pivotal for our arguments.

Theorem 3.2. *Assume that* $\mathcal{R}_k := \sigma(\epsilon_j,\ j \leq k)$ *for some Polish space-valued independent random variables* ϵ_j*,* $j \leq k$*,* $j \in \mathbb{Z}$*. Fix* $r > 2$ *and* $k \in \mathbb{N}$*. Let* W_n*,* $n \in \mathbb{N}$ *be a conditionally* L*-mixing process w.r.t.* $(\mathcal{R}_n, \mathcal{R}_n^+)$*,* *satisfying* $\mathbb{E}[W_n|\mathcal{R}_k] = 0$ *a.s. for all* $n \geq k$. *Let* $m > k$ *and let* b_j, $k < j \leq m$ *be deterministic numbers. Then we have*

$$\mathbb{E}^{1/r}\left[\max_{k<j\leq m}\left|\sum_{i=k+1}^{j} b_i W_i\right|^r \middle| \mathcal{R}_k\right] \leq C(r)\left(\sum_{i=k+1}^{m} b_i^2\right)^{1/2}\sqrt{M_r^k(W)\Gamma_r^k(W)}, \tag{14}$$

almost surely, where $C(r)$ *is a deterministic constant depending only on* r *but independent of* k, m.

Proof. See Theorem 2.6 of [4] (there, ϵ_j, $j \in \mathbb{N}$, are assumed to be i.i.d.; the proof, though, trivially works for a merely independent sequence, too). □

Remark 3.3. We will apply Theorem 3.2 with the choice $r = 3$. In that case it is known that $C(3) \leq 20$, see Theorem A.1 of [2].

Lemma 3.4. *Let* X_t, $t \in \mathbb{N}$ *be conditionally L-mixing. Let Assumption 2.2 hold true. Then, for each* $i \in \mathbb{N}$, *the random field* $H(\theta, X_t)$, $t \in \mathbb{N}$, $\theta \in B(i)$, *the closed ball of radius* i *centered at* 0, *is uniformly conditionally L-mixing with*

$$M_r^n(H(\theta, X), B(i)) \leq K_1 i + K_2 M_r^n(X) + H^* \tag{15}$$

and

$$\Gamma_r^n(H(\theta, X), B(i)) \leq 2K_2\Gamma_r^n(X). \tag{16}$$

Proof. See Lemma 6.4 and Example 2.4 of [2]. □

Lemma 3.5. *Let* $D \subset \mathbb{R}^d$ *be bounded. Fix* $n \in \mathbb{N}$ *and let* $\psi_t, t \geq n$ *be a sequence of* $\mathcal{R}_n$ *measurable random variables. Let* $X_t(\theta)$ *be conditionally L-mixing and Lipschitz in* $\theta \in D$. *Define the process* $Y_t = X_t(\psi_t), t \geq n$. *Then*

$$M_p^n(Y) \leq M_p^n(X), \qquad \Gamma_p^n(Y) \leq \Gamma_p^n(X).$$

Proof. The proof is identical to that of Lemma 6.3 of [4], noting the Lipschitz continuity. □

3.2 Further notation and introduction of auxiliary processses

Throughout this section we assume that the hypotheses of Theorem 2.7 are valid. Note that Assumption 2.2 implies

$$|h(\theta) - h(\theta')| \leq K_1|\theta - \theta'|,\ \theta, \theta' \in \mathbb{R}^d, \tag{17}$$

Assumption 2.5 implies

$$\langle h(\theta), \theta\rangle \geq a|\theta|^2 - b,\ \theta \in \mathbb{R}^d. \tag{18}$$

Also, Assumption 2.2 implies

$$|H(\theta, x)| \leq K_1|\theta| + K_2|x| + H^*, \tag{19}$$

with the constant H^* defined in (3). We will employ a family of Lyapunov-functions in the sequel. For this purpose, let us define, for each $p \geq 2$, $\mathrm{v}_p(x) := (1 + x^2)^{p/2}$, for any real $x \geq 0$, and similarly

$$V_p(\theta) := (1 + |\theta|^2)^{p/2},\ \theta \in \mathbb{R}^d.$$

Notice that these functions are twice continuously differentiable and

$$\lim_{|\theta|\to\infty}\frac{\nabla V_p(\theta)}{V_p(\theta)} = 0. \tag{20}$$

Let $\mathcal{P}_{V_p}$ denote the set of $\mu \in \mathcal{P}(\mathbb{R}^d)$ satisfying

$$\int_{\mathbb{R}^d} V_p(\theta)\,\mu(\mathrm{d}\theta) < \infty.$$

For $\mu \in \mathcal{P}(\mathbb{R}^d)$ and for a non-negative measurable $f : \mathbb{R}^d \to \mathbb{R}$, the notation

$$\mu(f) := \int_{\mathbb{R}^d} f(\theta)\mu(\mathrm{d}\theta)$$

is used. The following functional is pivotal in our arguments as it is used to measure the distance between probability measures. We define, for any $p \geq 1$ and $\mu, \nu \in \mathcal{P}_{V_p}$,

$$w_{1,p}(\mu, \nu) := \inf_{\zeta\in\mathcal{C}(\mu,\nu)}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}[1 \wedge |\theta - \theta'|](1 + V_p(\theta) + V_p(\theta'))\zeta(\mathrm{d}\theta\mathrm{d}\theta'), \tag{21}$$

Though $w_{1,p}$ is not a metric, it satisfies trivially

$$W_1(\mu, \nu) \leq w_{1,p}(\mu, \nu). \tag{22}$$

In the sequel we will need the case $p = 2$, that is, $w_{1,2}$.

Our estimations are carried out in a *continuous-time* setting, so we define and discuss a number of auxiliary continuous-time processes below. First, consider L_t, $t \in \mathbb{R}_+$ defined by the stochastic differential equation (SDE)

$$\mathrm{d}L_t = -h(L_t)\,\mathrm{d}t + \sqrt{\frac{2}{\beta}}\mathrm{d}B_t, \quad L_0 := \theta_0, \tag{23}$$

where B is standard Brownian motion on $(\Omega, \mathcal{F}, P)$, independent of $\mathcal{G}_\infty \vee \sigma(\theta_0)$. Its natural filtration is denoted by $\mathcal{F}_t$, $t \in \mathbb{R}_+$ henceforth. The meaning of $\mathcal{F}_\infty$ is clear. Equation (23) has a unique solution on $\mathbb{R}_+$ adapted to $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ since h is Lipschitz-continuous by (17). We proceed by defining, for each $\lambda > 0$,

$$L_t^\lambda := L_{\lambda t}, \ t \in \mathbb{R}_+.$$

Notice that $\tilde{B}_t^\lambda := B_{\lambda t}/\sqrt{\lambda}$, $t \in \mathbb{R}_+$ is also a Brownian motion and

$$\mathrm{d}L_t^\lambda = -\lambda h(L_t^\lambda)\,\mathrm{d}t + \sqrt{\frac{2\lambda}{\beta}}\mathrm{d}\tilde{B}_t^\lambda, \ L_0^\lambda = \theta_0. \tag{24}$$

Define $\mathcal{F}_t^\lambda := \mathcal{F}_{\lambda t}$, $t \in \mathbb{R}_+$, the natural filtration of $\tilde{B}_t^\lambda$, $t \in \mathbb{R}_+$.

Let us also introduce, for each $\lambda > 0$ and for each $\mathbf{x} = (x_0, x_1, \ldots) \in (\mathbb{R}^m)^\mathbb{N}$, the process $\tilde{Y}^\lambda(\mathbf{x})$, $t \in \mathbb{R}_+$ satisfying

$$\mathrm{d}\tilde{Y}_t^\lambda(\mathbf{x}) = -\lambda H(\tilde{Y}_t^\lambda(\mathbf{x}), x_{\lfloor t\rfloor})\,\mathrm{d}t + \sqrt{\frac{2\lambda}{\beta}}\mathrm{d}\tilde{B}_t^\lambda, \tag{25}$$

with initial condition $\tilde{Y}_0^\lambda(\mathbf{x}) = \theta_0$. Due to Assumption 2.2, there is a unique solution to (25) which is adapted to $(\mathcal{F}_t^\lambda)_{t\in\mathbb{R}_+}$. Moreover, for any given $s \geq 0$ and $t \geq s$, consider the following auxiliary process, which plays an important role in the derivation of our results,

$$\mathrm{d}\tilde{\zeta}^\lambda(t, s; \mathbf{x}, \theta) = -\lambda H(\tilde{\zeta}^\lambda(t, s; \mathbf{x}, \theta), x_{\lfloor t\rfloor})\mathrm{d}t + \sqrt{\frac{2\lambda}{\beta}}\mathrm{d}\tilde{B}_t^\lambda, \qquad \text{for } t > s, \tag{26}$$

with initial condition $\tilde{\zeta}^\lambda(s, s; \mathbf{x}, \theta) = \theta$, and notice that $\tilde{\zeta}^\lambda(t, s; \mathbf{x}, \tilde{Y}_s^\lambda(\mathbf{x})) = \tilde{Y}_t^\lambda(\mathbf{x})$.

Let us now define the continuously interpolated Euler-Maruyama approximation of $\tilde{Y}_t^\lambda(\mathbf{x})$, $t \in \mathbb{R}_+$ via

$$\mathrm{d}Y_t^\lambda(\mathbf{x}) = -\lambda H(Y_{\lfloor t\rfloor}^\lambda(\mathbf{x}), x_{\lfloor t\rfloor})\,\mathrm{d}t + \sqrt{\frac{2\lambda}{\beta}}\mathrm{d}\tilde{B}_t^\lambda, \tag{27}$$

with initial condition $Y_0^\lambda(\mathbf{x}) = \theta_0$. Notice at this point that (27), can be solved by a simple recursion. In addition, if one considers $Y_t^\lambda(\mathbf{X})$, $t \in \mathbb{R}_+$, where $\mathbf{X}$ is a random element in $(\mathbb{R}^m)^\mathbb{N}$ defined by $\mathbf{X}_i := X_i$, $i \in \mathbb{N}$, then for each integer $n \in \mathbb{N}$,

$$\mathcal{L}(Y_n^\lambda(\mathbf{X})) = \mathcal{L}(\theta_n^\lambda). \tag{28}$$

3.3 Layout of the proof

In view of (28), the main objective is to bound $W_1(\mathcal{L}(Y_t^\lambda(\mathbf{X})), \pi_\beta)$, which is decomposed as follows

$$W_1(\mathcal{L}(Y_t^\lambda(\mathbf{X})), \pi_\beta) \leq W_1(\mathcal{L}(Y_t^\lambda(\mathbf{X})), \mathcal{L}(\tilde{Y}_t^\lambda(\mathbf{X}))) + W_1(\mathcal{L}(\tilde{Y}_t^\lambda(\mathbf{X})), \mathcal{L}(L_t^\lambda)) + W_1(\mathcal{L}(L_t^\lambda), \pi_\beta).$$

The last term is controlled using the drift condition (29) below, due to the dissipativity Assumption 2.5, and Lipschitzness of the mean field h, see (17). The second term is controlled uniformly in t by a quantity which is proportional to $\sqrt{\lambda}$. For that purpose, we use novel results by [15], which give us a contraction in $w_{1,2}$, see Proposition 3.12 and, in particular, (49). To obtain this result, the mixing condition also plays a crucial role, see Lemma 3.19. Finally, the first term is controlled uniformly in t by a quantity which is also proportional to $\sqrt{\lambda}$, see Corollary 3.25. This is based on Kullback-Leibler distance estimates which go back to [10].

3.4 Crucial estimates

The next lemma shows that the SDEs (25) and (23) satisfy standard drift conditions involving the functions V_p. Note that, on the left-hand side of (29) below, the infinitesimal generator of the diffusion process L appears which is applied to the function V_p.

Lemma 3.6. *Let Assumption 2.5 hold. For each* $p \geq 2$,

$$\frac{\Delta V_p(\theta)}{\beta} - \langle h(\theta), \nabla V_p(\theta) \rangle \leq -C_6(p) V_p(\theta) + C_7(p), \;\; \theta \in \mathbb{R}^d, \tag{29}$$

and, for all $x \in \mathbb{R}^m$,

$$\frac{\Delta V_p(\theta)}{\beta} - \langle H(\theta, x), \nabla V_p(\theta) \rangle \leq -C_6(p) V_p(\theta) + C_7(p), \;\; \theta \in \mathbb{R}^d, \tag{30}$$

where $C_6(p) = ap/4$, $C_7(p) = (3/4)ap\mathrm{v}_p(\overline{M}(p))$ *with*

$$\overline{M}(p) = \sqrt{1/3 + 4b/(3a) + 4d/(3a\beta) + 4(p-2)/(3a\beta)}\,. \tag{31}$$

Proof. By direct calculation, the left-hand side of (29) equals

$$\frac{dp(|\theta|^2+1)^{(p-2)/2}}{\beta} + \frac{p(p-2)(|\theta|^2+1)^{(p-4)/2}|\theta|^2}{\beta} - p\langle h(\theta), (|\theta|^2+1)^{(p-2)/2}\theta \rangle. \tag{32}$$

By Assumption 2.5, see also (18), the third term of (32) is dominated by

$$-pa|\theta|^2(|\theta|^2+1)^{(p-2)/2} + pb(|\theta|^2+1)^{(p-2)/2}. \tag{33}$$

Then, for $|\theta| > \overline{M}(p)$, one observes that

$$\frac{\Delta V_p(\theta)}{\beta} - \langle h(\theta), \nabla V_p(\theta) \rangle \leq -\frac{ap}{4} V_p(\theta).$$

As for $|\theta| \leq \overline{M}(p)$, one obtains

$$\frac{\Delta V_p(\theta)}{\beta} - \langle h(\theta), \nabla V_p(\theta) \rangle + \frac{ap}{4} V_p(\theta) \leq \frac{3}{4} ap\mathrm{v}_p(\overline{M}(p)).$$

Take into consideration of the two cases, we have for all $\theta \in \mathbb{R}^d$,

$$\frac{\Delta V_p(\theta)}{\beta} - \langle h(\theta), \nabla V_p(\theta) \rangle \leq -C_6(p) V_p(\theta) + C_7(p).$$

The statement (30) follows in an identical way, noting that the constants which appear do not depend on x. ☐

Now, we proceed with the required moment estimates which play a crucial role in the derivation of the main results as given in Theorems 2.7, 2.9 and 2.11.

Lemma 3.7. *Let Assumptions 2.1, 2.2 and 2.5 hold. Let* $p \geq 2$. *For* $0 \leq s \leq t$, *let* $\tilde{\zeta}^\lambda(t, s; \mathbf{x}, \tilde{\theta})$ *be the solution of* (26) *with an initial condition* $\tilde{\theta} \in L^{2p-2}$. *Then, for any* $t > s \geq 0$,

$$\sup_{\mathbf{x} \in (\mathbb{R}^m)^{\mathbb{N}}} \sup_{t \geq s} \mathbb{E}[V_p(\tilde{\zeta}^\lambda(t, s; \mathbf{x}, \tilde{\theta}))] \leq \mathbb{E}[V_p(\tilde{\theta})] + 3\mathrm{v}_p(\overline{M}(p)) \tag{34}$$

where $\bar{M}(p)$ *is defined in* (31).

Proof. We note that $2p-2 \geq p$ for $p \geq 2$, hence $\mathbb{E}[V_p(\tilde{\theta})] < \infty$. For any fixed sequence $\mathbf{x} \in (\mathbb{R}^m)^{\mathbb{N}}$ and $t > s \geq 0$, by Itô's formula, one obtains almost surely,

$$\mathrm{d}V_p(\tilde{\zeta}^\lambda(t, s; \mathbf{x}, \tilde{\theta})) = \left[\lambda \frac{\Delta V_p(\tilde{\zeta}^\lambda(t, s; \mathbf{x}, \tilde{\theta}))}{\beta} - \lambda \langle H(\tilde{\zeta}^\lambda(t, s; \mathbf{x}, \tilde{\theta}), x_{\lfloor t \rfloor}), \nabla V_p(\tilde{\zeta}^\lambda(t, s; \mathbf{x}, \tilde{\theta})) \rangle \right] \mathrm{d}t$$
$$+ \sqrt{\frac{2\lambda}{\beta}} \langle \nabla V_p(\tilde{\zeta}^\lambda(t, s; \mathbf{x}, \tilde{\theta})), \mathrm{d}\tilde{B}_t^\lambda \rangle,$$

which implies

$$\mathbb{E}[V_p(\tilde{\zeta}^\lambda(t,s;\mathbf{x},\tilde{\theta}))] = \mathbb{E}[V_p(\tilde{\theta})] + \int_s^t \mathbb{E}\left[\lambda\frac{\Delta V_p(\tilde{\zeta}^\lambda(u,s;\mathbf{x},\tilde{\theta}))}{\beta} - \lambda\langle H(\tilde{\zeta}^\lambda(u,s;\mathbf{x},\tilde{\theta}),x_{\lfloor u\rfloor}),\nabla V_p(\tilde{\zeta}^\lambda(u,s;\mathbf{x},\tilde{\theta}))\rangle\right]\mathrm{d}u,$$

where the expectation of the stochastic integral disappears since $\sup_{0\le s\le t} E(\nabla V_p)^2(\tilde{\zeta}^\lambda(t,s;\mathbf{x},\tilde{\theta})) < \infty$ by $\tilde{\theta} \in L^{2p-2}$. Differentiating both sides and using Lemma 3.6, one obtains

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[V_p(\tilde{\zeta}^\lambda(t,s;\mathbf{x},\tilde{\theta}))] &= \mathbb{E}\left[\lambda\frac{\Delta V_p(\tilde{\zeta}^\lambda(t,s;\mathbf{x},\tilde{\theta}))}{\beta} - \lambda\langle H(\tilde{\zeta}^\lambda(t,s;\mathbf{x},\tilde{\theta}),x_{\lfloor t\rfloor}),\nabla V_p(\tilde{\zeta}^\lambda(t,s;\mathbf{x},\tilde{\theta}))\rangle\right]\\ &\le -\lambda C_6(p)\mathbb{E}[V_p(\tilde{\zeta}^\lambda(t,s;\mathbf{x},\tilde{\theta}))] + \lambda C_7(p),\end{aligned}$$

which yields

$$\mathbb{E}[V_p(\tilde{\zeta}^\lambda(t,s;\mathbf{x},\tilde{\theta}))] \le \mathrm{e}^{-\lambda C_6(p)(t-s)}\mathbb{E}[V_p(\tilde{\theta})] + \frac{C_7(p)}{C_6(p)}\left(1-\mathrm{e}^{-\lambda C_6(p)(t-s)}\right) \le \mathbb{E}[V_p(\tilde{\theta})] + \frac{C_7(p)}{C_6(p)}, \tag{35}$$

where $C_6(p)$, $C_7(p)$ and $\overline{M}(p)$ are defined in Lemma 3.6. Taking supremum over t and $\mathbf{x}$ on both sides yield (34). □

Corollary 3.8. *Let Assumptions 2.1, 2.2 and 2.5 hold. For any integer $p \ge 2$,*

$$\sup_{\mathbf{x}\in(\mathbb{R}^m)^{\mathbb{N}}}\sup_{t\in\mathbb{R}_+}\mathbb{E}[V_p(\tilde{Y}_t^\lambda(\mathbf{x}))] \le \mathbb{E}[V_p(\theta_0)] + 3\mathrm{v}_p(\overline{M}(p)), \tag{36}$$

where $\overline{M}(p)$ is defined in (31)*.*

Proof. By noting that $\tilde{Y}_t^\lambda(\mathbf{x}) = \tilde{\zeta}^\lambda(t,0;\mathbf{x},\theta_0)$, one immediately recovers the desired result from Lemma 3.7. □

Corollary 3.9. *Let Assumptions 2.1, 2.2 and 2.5 hold. For any integer $p \ge 2$,*

$$\sup_{t\in\mathbb{R}_+}\mathbb{E}[V_p(\tilde{Y}_t^\lambda(\mathbf{X}))] \le \mathbb{E}[V_p(\theta_0)] + 3\mathrm{v}_p(\overline{M}(p)), \tag{37}$$

where the constant $\overline{M}(p)$ is given explicitly in the proof of Lemma 3.6.

Proof. Due to the fact that the dissipativity condition 2.5 is uniform in x, all estimates are independent of x and therefore the result follows immediately from Corollary 3.8. □

Lemma 3.10. *Let Assumptions 2.1, 2.2 and 2.5 hold. For any $\lambda < \lambda_{\max}$ (see* (7)*), $n \in \mathbb{N}$ and $t \in (n, n+1]$ and any sequence $\mathbf{x} \in (\mathbb{R}^m)^{\mathbb{N}}$,*

$$\mathbb{E}[|Y_t^\lambda(\mathbf{x})|^{2p}] \le \mathbb{E}[|\theta_0|^{2p}] + \lambda a M(p,d)|x_n|^{2p} + \lambda a M(p,d)\sum_{j=0}^{n-1}(1-a\lambda)^j\,|x_{n-j}|^{2p} + \hat{M}(p,d), \tag{38}$$

where the constants $M(p,d)$ and $\hat{M}(p,d)$ are given by

$$\begin{aligned} M(p,d) &= (2\lambda_{\max}+4/a)^{p-1}\left[1/a + d\tilde{M}^2(p)\right]c_0^p\\ \hat{M}(p,d) &= (2\lambda_{\max}+4/a)^{p-1}\left[1/a + d\tilde{M}^2(p)\right]c_2^p + \tilde{M}^2(p)\,(\lambda_{\max}+2/a)^{p-1}\left(d + (1/\beta)^{p-1}\,(2dp(2p-1))^p\right)\end{aligned} \tag{39}$$

and

$$\tilde{M}(p) := 2^p\sqrt{p(2p-1)/(a\beta)}. \tag{40}$$

In particular,

$$\mathbb{E}|Y_t^\lambda(\mathbf{x})|^2 \le \mathbb{E}|\theta_0|^2 + \lambda c_0|x_n|^2 + \lambda c_0\sum_{j=0}^{n-1}(1-a\lambda)^j|x_{n-j}|^2 + c_1\,, \tag{41}$$

where c_0 and c_1 are defined by

$$c_0 = 8K_2^2\lambda_{\max},\ c_1 = 2a^{-1}(b + 4\lambda_{\max}(H^*)^2 + \beta^{-1}) \quad \textit{and} \quad c_2 = 2b + 8\lambda_{\max}(H^*)^2\,. \tag{42}$$

Proof. For any $n \in \mathbb{N}$ and $t \in (n, n+1]$, define

$$\Delta_n(\mathbf{x}, t) = Y_n^\lambda(\mathbf{x}) - \lambda H(Y_n^\lambda(\mathbf{x}), x_n)(t-n).$$

Consider initially only the calculations around the square of the norm of $Y_t^\lambda(\mathbf{x})$, i.e.

$$\begin{aligned}\mathbb{E}[|Y_t^\lambda(\mathbf{x})|^2|Y_n^\lambda(\mathbf{x})] &= \mathbb{E}\left[|\Delta_n(\mathbf{x},t)|^2 + (2\lambda/\beta)|\tilde{B}_t^\lambda - \tilde{B}_n^\lambda|^2 + 2\left\langle \Delta_n(\mathbf{x},t), \sqrt{(2\lambda/\beta)}(\tilde{B}_t^\lambda - \tilde{B}_n^\lambda)\right\rangle \Big| Y_n^\lambda(\mathbf{x})\right] \\ &= \mathbb{E}|\Delta_n(\mathbf{x},t)|^2 + (2\lambda/\beta)(t-n).\end{aligned}$$

Using Assumptions 2.2 and 2.5, one obtains for all $\lambda \le \lambda_{\max}$,

$$\begin{aligned}|\Delta_n(\mathbf{x},t)|^2 &= |Y_n^\lambda(\mathbf{x})|^2 - 2\lambda(t-n)\langle Y_n^\lambda(\mathbf{x}), H(Y_n^\lambda(\mathbf{x}), x_n)\rangle + \lambda^2|H(Y_n^\lambda(\mathbf{x}), x_n)(t-n)|^2 \\ &\le (1-2a\lambda(t-n))|Y_n^\lambda(\mathbf{x})|^2 + 2b\lambda(t-n) + 2\lambda^2(t-n)^2\{K_1^2|Y_n^\lambda(\mathbf{x})|^2 + 4K_2^2|x_n|^2 + 4(H^*)^2\} \\ &\le (1-a\lambda(t-n))|Y_n^\lambda(\mathbf{x})|^2 + \lambda(t-n)(c_0|x_n|^2 + 2b + 8\lambda_{\max}(H^*)^2). \end{aligned} \tag{43}$$

The desired result (41) follows from an easy induction. For higher moments, the calculation is somewhat more involved. To this end, one calculates

$$\begin{aligned}\mathbb{E}[|Y_t^\lambda(\mathbf{x})|^{2p}|Y_n^\lambda(\mathbf{x})] &= \mathbb{E}\left[\left(|\Delta_n(\mathbf{x},t)|^2 + \frac{2\lambda}{\beta}|\tilde{B}_t^\lambda - \tilde{B}_n^\lambda|^2 + 2\left\langle \Delta_n(\mathbf{x},t), \sqrt{\frac{2\lambda}{\beta}}(\tilde{B}_t^\lambda - \tilde{B}_n^\lambda)\right\rangle\right)^p \Bigg| Y_n^\lambda(\mathbf{x})\right] \\ &= \sum_{k_1+k_2+k_3=p} \frac{p!}{k_1!k_2!k_3!}\mathbb{E}\left[|\Delta_n(\mathbf{x},t)|^{2k_1}\left|\sqrt{\frac{2\lambda}{\beta}}(\tilde{B}_t^\lambda - \tilde{B}_n^\lambda)\right|^{2k_2}\right. \\ &\qquad \left. \times \left(2\left\langle \Delta_n(\mathbf{x},t), \sqrt{\frac{2\lambda}{\beta}}(\tilde{B}_t^\lambda - \tilde{B}_n^\lambda)\right\rangle\right)^{k_3} \Bigg| Y_n^\lambda(\mathbf{x})\right] \\ &\le |\Delta_n(\mathbf{x},t)|^{2p} + 2p\mathbb{E}\left[|\Delta_n(\mathbf{x},t)|^{2p-2}\left\langle \Delta_n(\mathbf{x},t), \sqrt{\frac{2\lambda}{\beta}}(\tilde{B}_t^\lambda - \tilde{B}_n^\lambda)\right\rangle \Bigg| Y_n^\lambda(\mathbf{x})\right] \\ &\qquad + \sum_{k=2}^{2p}\binom{2p}{k}\mathbb{E}\left[|\Delta_n(\mathbf{x},t)|^{2p-k}\left|\sqrt{\frac{2\lambda}{\beta}}(\tilde{B}_t^\lambda - \tilde{B}_n^\lambda)\right|^k \Bigg| Y_n^\lambda(\mathbf{x})\right],\end{aligned}$$

where the last inequality is due to Lemma 4.4. The following inequality is used in the subsequent analysis

$$(r+s)^p \le (1+\epsilon)^{p-1}r^p + (1+\epsilon^{-1})^{p-1}s^p$$

where $p \ge 2$, r, $s \ge 0$ and $\epsilon > 0$. We continue as follows

$$\begin{aligned}\mathbb{E}[|Y_t^\lambda(\mathbf{x})|^{2p}|Y_n^\lambda(\mathbf{x})] &\le |\Delta_n(\mathbf{x},t)|^{2p} + \mathbb{E}\Bigg[\left(\sum_{l=0}^{2(p-1)}\binom{2p}{l+2}|\Delta_n(\mathbf{x},t)|^{2(p-1)-l}\left|\sqrt{\frac{2\lambda}{\beta}}(\tilde{B}_t^\lambda - \tilde{B}_n^\lambda)\right|^l\right) \\ &\qquad \times \left(\frac{2\lambda}{\beta}|\tilde{B}_t^\lambda - \tilde{B}_n^\lambda|^2\right) \Bigg| Y_n^\lambda(\mathbf{x})\Bigg] \\ &\le |\Delta_n(\mathbf{x},t)|^{2p} + \mathbb{E}\Bigg[\binom{2p}{2}\left(\sum_{l=0}^{2(p-1)}\binom{2(p-1)}{l}|\Delta_n(\mathbf{x},t)|^{2(p-1)-l}\left|\sqrt{\frac{2\lambda}{\beta}}(\tilde{B}_t^\lambda - \tilde{B}_n^\lambda)\right|^l\right) \\ &\qquad \times \left(\frac{2\lambda}{\beta}|\tilde{B}_t^\lambda - \tilde{B}_n^\lambda|^2\right) \Bigg| Y_n^\lambda(\mathbf{x})\Bigg] \\ &= |\Delta_n(\mathbf{x},t)|^{2p} + \frac{2\lambda p(2p-1)}{\beta}\mathbb{E}\Bigg[\left(|\Delta_n(\mathbf{x},t)| + \sqrt{\frac{2\lambda}{\beta}}|\tilde{B}_t^\lambda - \tilde{B}_n^\lambda|\right)^{2p-2}|\tilde{B}_t^\lambda - \tilde{B}_n^\lambda|^2 \Bigg| Y_n^\lambda(\mathbf{x})\Bigg] \\ &\le |\Delta_n(\mathbf{x},t)|^{2p} + 2^{2p-3}p(2p-1)|\Delta_n(\mathbf{x},t)|^{2p-2}\frac{2\lambda}{\beta}\mathbb{E}[|\tilde{B}_t^\lambda - \tilde{B}_n^\lambda|^2|Y_n^\lambda(\mathbf{x})] \\ &\qquad + 2^{2p-3}p(2p-1)\left(\frac{2\lambda}{\beta}\right)^p\mathbb{E}[|\tilde{B}_t^\lambda - \tilde{B}_n^\lambda|^{2p}|Y_n^\lambda(\mathbf{x})]\end{aligned}$$

$$= |\Delta_n(\mathbf{x},t)|^{2p} + \lambda(t-n)2^{2p-2}p(2p-1)\frac{d}{\beta}|\Delta_n(\mathbf{x},t)|^{2p-2}$$
$$+ 2^{2p-3}p(2p-1)\left(\frac{2\lambda}{\beta}\right)^p \mathbb{E}\left[\left|\int_n^t 1\, d\tilde{B}_s^\lambda\right|^{2p}\right]$$

which yields

$$\mathbb{E}[|Y_t^\lambda(\mathbf{x})|^{2p}|Y_n^\lambda(\mathbf{x})] \leq |\Delta_n(\mathbf{x},t)|^{2p} + \lambda(t-n)2^{2p-2}p(2p-1)\frac{d}{\beta}|\Delta_n(\mathbf{x},t)|^{2p-2}$$
$$+ 2^{3p-3}\left(\lambda(t-n)\right)^p \left(p(2p-1)\right)^{p+1}\left(\frac{d}{\beta}\right)^p \tag{44}$$

where the moment estimates of stochastic integrals are given in Theorem 7.1 in Chapter 1 of [22]. Using the notations in (42) and the inequality (43), one calculates

$$\begin{aligned}
|\Delta_n(\mathbf{x},t)|^{2p} &\leq \{(1-a\lambda(t-n))|Y_n^\lambda(\mathbf{x})|^2 + \lambda(t-n)(c_0|x_n|^2+c_2)\}^p \\
&\leq (1+\frac{a\lambda(t-n)}{2})^{p-1}(1-a\lambda(t-n))^p|Y_n^\lambda(\mathbf{x})|^{2p} + (1+\frac{2}{a\lambda(t-n)})^{p-1}\lambda^p(t-n)^p(c_0|x_n|^2+c_2)^p \\
&\leq (1-\frac{a\lambda(t-n)}{2})^{p-1}(1-a\lambda(t-n))|Y_n^\lambda(\mathbf{x})|^{2p} + (\lambda(t-n)+\frac{2}{a})^{p-1}\lambda(t-n)(c_0|x_n|^2+c_2)^p. \quad (45)
\end{aligned}$$

Substituting (45) into (44) yields

$$\begin{aligned}
\mathbb{E}[|Y_t^\lambda(\mathbf{x})|^{2p}|Y_n^\lambda(\mathbf{x})] \leq &(1-\frac{a\lambda(t-n)}{2})^{p-1}(1-a\lambda(t-n))|Y_n^\lambda(\mathbf{x})|^{2p} + (\lambda(t-n)+\frac{2}{a})^{p-1}\lambda(t-n)(c_0|x_n|^2+c_2)^p \\
&+ \lambda(t-n)2^{2p-2}p(2p-1)\frac{d}{\beta}\Big[(1-\frac{a\lambda(t-n)}{2})^{p-2}(1-a\lambda(t-n))|Y_n^\lambda(\mathbf{x})|^{2(p-1)} \\
&\qquad + (\lambda(t-n)+\frac{2}{a})^{p-2}\lambda(t-n)(c_0|x_n|^2+c_2)^{p-1}\Big] \\
&+ 2^{3p-3}\left(\lambda(t-n)\right)^p\left(p(2p-1)\right)^{p+1}\left(\frac{d}{\beta}\right)^p. \quad (46)
\end{aligned}$$

Define $\tilde{M}(p)$ as in (40) and observe that for $|Y_n^\lambda(\mathbf{x})| \geq \sqrt{d}\tilde{M}(p)$

$$\frac{a\lambda(t-n)}{4}|Y_n^\lambda(\mathbf{x})|^{2p} \geq \lambda(t-n)2^{2p}p(2p-1)\frac{d}{4\beta}|Y_n^\lambda(\mathbf{x})|^{2(p-1)}.$$

Consequently, on $\{|Y_n^\lambda(\mathbf{x})| \geq \sqrt{d}\tilde{M}(p)\}$ the inequality (46) yields

$$\begin{aligned}
\mathbb{E}[|Y_t^\lambda(\mathbf{x})|^{2p}|Y_n^\lambda(\mathbf{x})] \leq &(1-\frac{a\lambda(t-n)}{4})(1-\frac{a\lambda(t-n)}{2})^{p-2}(1-a\lambda(t-n))|Y_n^\lambda(\mathbf{x})|^{2p} \\
&+ \lambda(t-n)\left[\lambda(t-n)+\frac{2}{a}\right]^{p-1}(c_0|x_n|^2+c_2)^p \\
&+ \lambda^2(t-n)^2\left[\lambda(t-n)+\frac{2}{a}\right]^{p-2}2^{2p-2}p(2p-1)\frac{d}{\beta}(c_0|x_n|^2+c_2)^{p-1} \\
&+ \lambda^p(t-n)^p 2^{3p-3}(p(2p-1))^{p+1}\left(\frac{d}{\beta}\right)^p \\
\leq &(1-a\lambda(t-n))|Y_n^\lambda(\mathbf{x})|^{2p} + \lambda(t-n)\left(\lambda_{\max}+\frac{2}{a}\right)^{p-1}(1+ad\tilde{M}^2(p))(c_0|x_n|^2+c_2)^p \\
&+ \lambda(t-n)ad\tilde{M}^2(p)\left(\lambda_{\max}+\frac{2}{a}\right)^{p-1} + \lambda(t-n)(8\lambda_{\max})^{p-1}(p(2p-1))^{p+1}\left(\frac{d}{\beta}\right)^p \\
\leq &(1-a\lambda(t-n))|Y_n^\lambda(\mathbf{x})|^{2p} + \lambda(t-n)a\left(M(p,d)|x_n|^{2p}+\hat{M}(p,d)\right), \quad (47)
\end{aligned}$$

where the constants $M(p,d)$ and $\hat{M}(p,d)$ are defined in (39). Moreover, on $\{|Y_n^\lambda(\mathbf{x})| < \sqrt{d}\tilde{M}(p)\}$ the inequality (46) yields again

$$\mathbb{E}[|Y_t^\lambda(\mathbf{x})|^{2p}|Y_n^\lambda(\mathbf{x})] \leq (1-a\lambda(t-n))|Y_n^\lambda(\mathbf{x})|^{2p} + \lambda(t-n)a\left(M(p,d)|x_n|^{2p}+\hat{M}(p,d)\right) \tag{48}$$

Due to (48) and (47), one obtains by induction

$$\mathbb{E}[|Y_t^\lambda(\mathbf{x})|^{2p}] \leq (1-a\lambda)^n \mathbb{E}[|Y_0|^{2p}] + \hat{M}(p,d) + \lambda a M(p,d)|x_n|^{2p} + \lambda a M(p,d) \sum_{j=0}^{n-1} (1-a\lambda)^j \, |x_{n-j}|^{2p}.$$

The desired result (38) follows immediately. □

Remark 3.11. One notes here that $\left(\mathbb{E}[|Y_t^\lambda(\mathbf{x})|^{2p}]\right)^{1/(2p)}$ is of order $\sqrt{d}$, where d denotes the dimension of the problem.

A crucial contraction property is formulated in the next theorem.

Proposition 3.12. *Let L_t', $t \in \mathbb{R}_+$ be the solution of* (23) *with initial condition $L_0' = \theta_0'$ which is independent of $\mathcal{F}_\infty$ and satisfies $|\theta_0'| \in L^2$. Then,*

$$w_{1,2}(\mathcal{L}(L_t), \mathcal{L}(L_t')) \leq C_9 \mathrm{e}^{-C_8 t} w_{1,2}(\mathcal{L}(\theta_0), \mathcal{L}(\theta_0')), \ t \in \mathbb{R}_+, \tag{49}$$

where the constants C_8 and C_9 are given explicitly in Lemma 3.26. Fix a positive integer m. Suppose, for any $t > m$, $\tilde{\zeta}^\lambda(t,m;\mathbf{x},\tilde{\theta})$ and $\tilde{\zeta}^\lambda(t,m;\mathbf{x},\tilde{\theta}')$ are the solutions of (26) *with initial conditions $\tilde{\theta}$, $\tilde{\theta}' \in L^2$, which are independent of $\mathcal{F}_\infty$. Then,*

$$w_{1,2}(\mathcal{L}(\tilde{\zeta}^\lambda(t,m;\mathbf{x},\tilde{\theta}), \mathcal{L}(\tilde{\zeta}^\lambda(t,m;\mathbf{x},\tilde{\theta}')) \leq C_9 e^{-C_8 \lambda(t-m)} w_{1,2}(\mathcal{L}(\tilde{\theta}), \mathcal{L}(\tilde{\theta}')), \qquad \text{for any } t > m. \tag{50}$$

Proof. We first treat L_t, L_t'. Assumption 2.1 of [15] holds with κ constant (and equal to K_1) due to Assumption 2.2. Assumption 2.5 of the same paper is valid due to (20) and, moreover, Assumption 2.2 of [15] holds with $V = V_2$ due to Lemma 3.6 (note that in that paper the diffusion coefficient is assumed to be 1 while in our case it is $\sqrt{2/\beta}$ but this does not affect the validity of the arguments, only the values of the constants). Thus, in view of Corollary 2.3 of [15],

$$\mathcal{W}_{\rho_2}(\mathcal{L}(L_t), \mathcal{L}(L_t')) \leq \mathrm{e}^{-C_8 t} \mathcal{W}_{\rho_2}(\mathcal{L}(\theta_0), \mathcal{L}(\theta_0')), \ t \in \mathbb{R}_+,$$

where C_8 is given in Lemma 3.26 and the functional $\mathcal{W}_{\rho_2}$ comes from [15] with the choice $V := V_2$, i.e.

$$\mathcal{W}_{\rho_2}(\mu,\nu) := \inf_{\zeta \in \mathcal{C}(\mu,\nu)} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(|\theta-\theta'|)(1+\epsilon V_2(\theta) + \epsilon V_2(\theta'))\zeta(\mathrm{d}\theta, \mathrm{d}\theta'), \text{ where } \mu,\nu \in \mathcal{P}_{V_2},$$

where f is given in Lemma 3.26. Note that f is a concave, bounded and non-decreasing continuous function and ϵ is a positive constant, for more details see Theorem 2.2, Section 5 of [15]. Consequently, by using the definition of $\mathcal{W}_{\rho_2}$, one obtains

$$C_{10} w_{1,2}(\mu,\nu) \leq \mathcal{W}_{\rho_2}(\mu,\nu) \leq C_{11} w_{1,2}(\mu,\nu), \ \mu,\nu \in \mathcal{P}_{V_2},$$

where C_{10}, C_{11} can be found in Lemma 3.26. Statement (49) follows with $C_9 = C_{11}/C_{10}$.

The same approach is used for $\tilde{\zeta}^\lambda(t,m;\mathbf{x},\tilde{\theta})$ and $\tilde{\zeta}^\lambda(t,m;\mathbf{x},\tilde{\theta}')$, with the only difference being that we derive first the contraction on an interval of length at most one, since the contribution from the data sequence, through $x_{\lfloor t \rfloor}$, remains constant and thus, the drift coefficient remains autonomous for such an interval. More concretely, Assumption 2.1 of [15] holds in this case too with κ constant and equal to K_1 due to Assumption 2.2. Assumption 2.2 of [15] is true with $V = V_2$ due to Lemma 3.6. Note that the statements in these Assumptions are uniform in x (and thus identical for different values of $x_{\lfloor t \rfloor}$). Finally, Assumption 2.5 of [15] is also true due to (20). Thus, the results of Corollary 2.3 of [15] apply in this case, too, and one concludes that

$$\begin{aligned}
\mathcal{W}_{\rho_2}(\mathcal{L}(\tilde{\zeta}^\lambda(t,m;\mathbf{x},\tilde{\theta})), \mathcal{L}(\tilde{\zeta}^\lambda(t,m;\mathbf{x},\tilde{\theta}'))) =& \mathcal{W}_{\rho_2}(\mathcal{L}(\tilde{\zeta}^\lambda(t,\lfloor t \rfloor;\mathbf{x},\tilde{\zeta}^\lambda(\lfloor t \rfloor,m;\mathbf{x},\tilde{\theta}))), \mathcal{L}(\tilde{\zeta}^\lambda(t,\lfloor t \rfloor;\mathbf{x},\tilde{\zeta}^\lambda(\lfloor t \rfloor,m;\mathbf{x},\tilde{\theta}')))) \\
\leq& \mathrm{e}^{-C_8\lambda(t-\lfloor t \rfloor)} \mathcal{W}_{\rho_2}(\mathcal{L}(\tilde{\zeta}^\lambda(\lfloor t \rfloor,m;\mathbf{x},\tilde{\theta})), \mathcal{L}(\tilde{\zeta}^\lambda(\lfloor t \rfloor,m;\mathbf{x},\tilde{\theta}'))) \\
\leq& \mathrm{e}^{-C_8\lambda(t-(\lfloor t \rfloor-1))} \mathcal{W}_{\rho_2}(\mathcal{L}(\tilde{\zeta}^\lambda(\lfloor t \rfloor-1,m;\mathbf{x},\tilde{\theta})), \mathcal{L}(\tilde{\zeta}^\lambda(\lfloor t \rfloor-1,m;\mathbf{x},\tilde{\theta}'))) \\
\leq& e^{-C_8\lambda(t-m)} \mathcal{W}_{\rho_2}(\mathcal{L}(\tilde{\theta}), \mathcal{L}(\tilde{\theta}')).
\end{aligned} \tag{51}$$

Observing as above that $\mathcal{W}_{\rho_2}$ is controlled from above and below by multiples of $w_{1,2}$, (51) yields the result. □

Define a continuous-time filtration $\mathcal{H}_t := \mathcal{F}_\infty \vee \mathcal{G}_{\lfloor t \rfloor}$, $t \in \mathbb{R}_+$ and the corresponding decreasing family of sigma-algebras $\mathcal{H}_t^+ := \mathcal{G}_{\lfloor t \rfloor}^+$, $t \in \mathbb{R}_+$. Moreover, let $T := \lfloor 1/\lambda \rfloor$, which is used for the creation of a suitable set of grid points.

Lemma 3.13. *For each $n \in \mathbb{N}$, there exists a measurable function*

$$h_{\cdot,nT} : \Omega \times [nT, \infty) \times \mathbb{R}^d \to \mathbb{R}^d$$

such that, for each $t \geq nT$ and $\theta \in \mathbb{R}^d$, $h_{t,nT}(\theta)(\omega)$ is a version of $\mathbb{E}[H(\theta, X_{\lfloor t \rfloor})|\mathcal{H}^\lambda_{nT}]$ and, for almost every $\omega \in \Omega$, $\theta \to h_{t,nT}(\theta)(\omega)$ is continuous.

Proof. As $h_{t,nT}$, $t \in [k, k+1)$ can be assumed constant for each $k \in \mathbb{N}$, it suffices to prove the existence of a measurable $h_{k,nT} : \Omega \times \mathbb{R}^d \to \mathbb{R}^d$ which is continuous in its second variable, for each fixed k. This follows from Lemma 8.5 of [2]. □

Lemma 3.14. *There exist random variables Ξ_n, $n \in \mathbb{N}$ such that, for all $\theta \in \mathbb{R}^d$,*

$$\int_{nT}^{\infty} |h_{t,nT}(\theta) - h(\theta)| \, \mathrm{d}t \leq \Xi_n,$$

and for each $p \geq 1$ there exist $C_{18}(p)$ such that

$$\sup_{n \in \mathbb{N}} \mathbb{E}[\Xi_n^p] \leq C_{18}(p).$$

Proof. Notice that, for any integer $k \geq nT$,

$$\mathbb{E}[H(\theta, \mathbb{E}[X_k|\mathcal{H}^+_{nT}])|\mathcal{H}_{nT}] = \mathbb{E}[H(\theta, \mathbb{E}[X_k|\mathcal{H}^+_{nT}])],$$

since $\mathcal{H}^+_{nT}$ is independent of $\mathcal{H}_{nT}$. This implies

$$\begin{aligned}|h_{k,nT}(\theta) - h(\theta)| \leq & \left|\mathbb{E}[H(\theta, X_k)|\mathcal{H}_{nT}] - \mathbb{E}[H(\theta, \mathbb{E}[X_k|\mathcal{H}^+_{nT}])|\mathcal{H}_{nT}]\right| + \left|\mathbb{E}[H(\theta, \mathbb{E}[X_k|\mathcal{H}^+_{nT}])] - \mathbb{E}[H(\theta, X_k)]\right| \\ \leq & K_2 \mathbb{E}[|X_k - \mathbb{E}[X_k|\mathcal{H}^+_{nT}]||\mathcal{H}_{nT}] + K_2 \mathbb{E}[|X_k - \mathbb{E}[X_k|\mathcal{H}^+_{nT}]|] \\ \leq & K_2[\gamma_1^{nT}(k - nT, X) + \mathbb{E}\gamma_1^{nT}(k - nT, X)].\end{aligned}$$

Hence, noting that $h_{\cdot,nT}(\theta)$ is constant on each interval $[k, k+1)$, $k \in \mathbb{N}$, $k \geq nT$,

$$\int_{nT}^{\infty} |h_{t,nT}(\theta) - h(\theta)| \, \mathrm{d}t \leq K_2 \left(\Gamma_1^{nT}(X) + \mathbb{E}[\Gamma_1^{nT}(X)] \right)$$

Since X is conditionally L-mixing, $\sup_{n \in \mathbb{N}} \mathbb{E}[(\Gamma_1^{nT}(X))^p] < \infty$ for every $p \geq 1$. The statement follows. □

We recall that $\mathbf{X}$ refers to the $(\mathbb{R}^m)^{\mathbb{N}}$-valued random variable that has coordinates $\mathbf{X}_i = X_i$, $i \in \mathbb{N}$. Let $Z^\lambda(t, s, \vartheta)$, $t \geq s$ denote the solution of the SDE

$$\mathrm{d}Z^\lambda(t, s, \vartheta) = -\lambda h(Z^\lambda(t, s, \vartheta)) \, \mathrm{d}t + \sqrt{2\lambda/\beta} \mathrm{d}\tilde{B}^\lambda_t,$$

with initial condition $Z^\lambda(s, s, \vartheta) := \vartheta$ for some $\mathcal{H}^\lambda_s$-measurable random variable ϑ. Note that $L^\lambda_t = Z^\lambda(t, 0, \theta_0)$.

Definition 3.15. *Fix $n \in \mathbb{N}$ and define*

$$\overline{Z}^{\lambda,n}_t = Z^\lambda(t, nT, \tilde{Y}^\lambda_{nT}(\mathbf{X})), \qquad \text{for any } t \in [nT, \infty).$$

It should be emphasized that for different n, the process $\overline{Z}^{\lambda,n}$ is redefined accordingly and $\overline{Z}^{\lambda,n}_t$ is $\mathcal{H}_{nT}$-measurable for all $t \geq nT$.

Lemma 3.16. *Let Assumptions 2.1, 2.2 and 2.5 hold. For any integers $p \geq 2$ and n,*

$$\sup_{t \geq nT} \mathbb{E}[V_p(\overline{Z}^{\lambda,n}_t)] \leq \mathbb{E}[V_p(\theta_0)] + 6\mathrm{v}_p(\overline{M}(p)), \tag{52}$$

where the constant $\overline{M}(p)$ is given explicitly in the proof of Lemma 3.6.

Proof. By application of Ito's formula, one obtains almost surely, for any $t \in [nT, \infty)$,

$$\mathrm{d}V_p(\overline{Z}^{\lambda,n}_t) = \left[\lambda \frac{\Delta V_p(\overline{Z}^{\lambda,n}_t)}{\beta} - \lambda \langle h(\overline{Z}^{\lambda,n}_t), \nabla V_p(\overline{Z}^{\lambda,n}_t) \rangle \right] \mathrm{d}t + \sqrt{\frac{2\lambda}{\beta}} \langle \nabla V_p(\overline{Z}^{\lambda,n}_t), \mathrm{d}\tilde{B}^\lambda_t \rangle,$$

which implies, noting that the expectation of the stochastic integral vanishes by standard arguments,

$$\mathbb{E}[V_p(\overline{Z}^{\lambda,n}_t)] = \mathbb{E}[V_p(\tilde{Y}^\lambda_{nT}(\mathbf{X}))] + \int_{nT}^{t} \mathbb{E}\left[\lambda \frac{\Delta V_p(\overline{Z}^{\lambda,n}_s)}{\beta} - \lambda \langle h(\overline{Z}^{\lambda,n}_s), \nabla V_p(\overline{Z}^{\lambda,n}_s) \rangle \right] \mathrm{d}s,$$

Differentiating both sides and using Lemma 3.6, one obtains for any $t \geq nT$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[V_p(\overline{Z}_t^{\lambda,n})] = \mathbb{E}\left[\lambda\frac{\Delta V_p(\overline{Z}_t^{\lambda,n})}{\beta} - \lambda\langle h(\overline{Z}_t^{\lambda,n}), \nabla V_p(\overline{Z}_t^{\lambda,n})\rangle\right] \leq -\lambda C_6(p)\mathbb{E}[V_p(\overline{Z}_t^{\lambda,n})] + \lambda C_7(p),$$

which yields

$$\mathbb{E}[V_p(\overline{Z}_t^{\lambda,n})] \leq \mathrm{e}^{-\lambda C_6(p)(t-nT)}\mathbb{E}[V_p(\tilde{Y}_{nT}^{\lambda}(\mathbf{X}))] + \frac{C_7(p)}{C_6(p)}\left(1 - \mathrm{e}^{-\lambda C_6(p)(t-nT)}\right) \leq \mathbb{E}[V_p(\tilde{Y}_{nT}^{\lambda}(\mathbf{X}))] + \frac{C_7(p)}{C_6(p)},$$

and thus, in view of (37),

$$\mathbb{E}[V_p(\overline{Z}_t^{\lambda,n})] \leq \mathbb{E}[V_p(\theta_0)] + 2\frac{C_7(p)}{C_6(p)}.$$

Finally, since $C_6(p) = ap/4$, $C_7(p) = (3/4)ap\mathrm{v}_p(\overline{M}(p))$ according to the proof of Lemma 3.6, the desired result (52) follows. □

Corollary 3.17. *Assume 2.1, 2.2 and 2.5 hold. For any integer $p \geq 2$, and $T = \lfloor 1/\lambda \rfloor$,*

$$\mathbb{E}[\sup_{nT \leq t \leq (n+1)T} V_p(\overline{Z}_t^{\lambda,n})] \leq 3\sqrt{\mathbb{E}[V_{2p}(\theta_0)] + 3(1+ap)\mathrm{v}_{2p}(\overline{M}(2p))} \tag{53}$$

where the constant $\overline{M}(2p)$ is given explicitly in the proof of Lemma 3.6.

Proof. For any $n \in \mathbb{N}$, $q \geq 2$ and any bounded stopping time $\tau_n \geq nT$ (a.s.), one obtains by application of Ito's formula that, almost surely,

$$V_q(\overline{Z}_{\tau_n}^{\lambda,n}) = V_q(\tilde{Y}_{nT}^{\lambda}(\mathbf{X})) + \int_{nT}^{\tau_n}\left[\lambda\frac{\Delta V_q(\overline{Z}_s^{\lambda,n})}{\beta} - \lambda\langle h(\overline{Z}_s^{\lambda,n}), \nabla V_q(\overline{Z}_t^{\lambda,n})\rangle\right]\mathrm{d}s + \sqrt{\frac{2\lambda}{\beta}}\int_{nT}^{\tau_n}\langle\nabla V_q(\overline{Z}_s^{\lambda,n}), \mathrm{d}\tilde{B}_s^{\lambda}\rangle.$$

Due to Lemmata 3.6 and 3.16,

$$\begin{aligned}\mathbb{E}[V_q(\overline{Z}_{\tau_n}^{\lambda,n})] &\leq \mathbb{E}[V_q(\tilde{Y}_{nT}^{\lambda}(\mathbf{X}))] + \mathbb{E}\left[\int_{nT}^{\tau_n}\left(-\lambda C_6(p)V_q(\overline{Z}_s^{\lambda,n}) + \lambda C_7(p)\right)\mathrm{d}s\right] \\ &\leq \mathbb{E}[V_q(\tilde{Y}_{nT}^{\lambda}(\mathbf{X}))] + \lambda C_7(p)\mathbb{E}[(\tau_n - nT)]\end{aligned}$$

Then, according to Lenglart's domination inequality [18], see also Proposition 4.7 of [24], for any $k \in (0,1)$

$$\mathbb{E}\left[\left(\sup_{nT \leq t \leq (n+1)T} V_q(\overline{Z}_t^{\lambda,n})\right)^k\right] \leq \frac{2-k}{1-k}\left(\mathbb{E}[V_q(\tilde{Y}_{nT}^{\lambda}(\mathbf{X}))] + \lambda C_7(p)T\right)^k.$$

Consequently, for $k = 1/2$ and $q = 2p$ and in view of Corollary 3.9, the desired result holds. □

Lemma 3.18. *Let $X_k, k \in \mathbb{N}$ be conditionally L-mixing. Recall $T = \lfloor 1/\lambda \rfloor$ and choose $n, N \in \mathbb{N}$. We define the filtrations $\mathcal{R}_j := \mathcal{H}_{\lfloor nT+j/N \rfloor}$, $\mathcal{R}_j^+ := \mathcal{H}_{\lfloor nT+j/N \rfloor}^+$, $j \in \mathbb{N}$ and $G_j := X_{\lfloor nT+j/N \rfloor}$, $j \in \mathbb{N}$. Then, it holds that*

$$M_p^0(G) = M_p^{nT}(X) \text{ and } \Gamma_p^0(G) \leq 2N\Gamma_p^{nT}(X). \tag{54}$$

Proof. Note that $M_p^0(G)$, $\Gamma_p^0(G)$ are calculated with respect to $(\mathcal{R}_j, \mathcal{R}_j^+)_{j\in\mathbb{N}}$ while the quantities $M_p^{nT}(X)$ and $\Gamma_p^{nT}(G)$ are calculated with respect to $(\mathcal{G}_j, \mathcal{G}_j^+)_{j\in\mathbb{N}}$. The equality in (54) is trivial.

Notice that, if $\tau \leq j$, $mN \leq \tau < (m+1)N$ for some $m \in \mathbb{N}$ then

$$\begin{aligned}&E^{1/p}[|X_{\lfloor nT+j/N \rfloor} - E[X_{\lfloor nT+j/N \rfloor}|\mathcal{H}_{\lfloor nT+(j-\tau)/N \rfloor}^+ \vee \mathcal{H}_{nT}]|^p|\mathcal{H}_{nT}] \\ \leq\ &\max\{\gamma_p^{nT}(m, X), \gamma_p^{nT}(m+1, X)\} \\ \leq\ &\max\{\gamma_p^{nT}(m, X), 2\gamma_p^{nT}(m, X)\} = 2\gamma_p^{nT}(m, X)\end{aligned}$$

using Lemma A.1 of [4], whence

$$\gamma_p^0(\tau, G) \leq 2\gamma_p^{nT}(m, X)$$

for each $mN \leq \tau < (m+1)N$. The inequality in (54) follows. □

Lemma 3.19. *Assume 2.1, 2.2 and 2.5 hold. There is C_{19} such that, for each $0 < \lambda \leq \lambda_{\max}$,*

$$W_1(\mathcal{L}(L_t^{\lambda}), \mathcal{L}(\tilde{Y}_t^{\lambda}(\mathbf{X}))) \leq C_{19}\sqrt{\lambda}.$$

Proof. Fix $t \in [nT, (n+1)T]$. Let us estimate, using Assumption 2.2,

$$\begin{aligned}
\left|\tilde{Y}_t^\lambda(\mathbf{X}) - \overline{Z}_t^{\lambda,n}\right| \leq & \lambda \left| \int_{nT}^t \left[H(\tilde{Y}_s^\lambda(\mathbf{X}), X_{\lfloor s \rfloor}) - h(\overline{Z}_s^{\lambda,n}) \right] \mathrm{d}s \right| \\
\leq & \lambda \int_{nT}^t \left| H(\tilde{Y}_s^\lambda(\mathbf{X}), X_{\lfloor s \rfloor}) - H(\overline{Z}_s^{\lambda,n}, X_{\lfloor s \rfloor}) \right| \mathrm{d}s + \lambda \left| \int_{nT}^t \left[H(\overline{Z}_s^{\lambda,n}, X_{\lfloor s \rfloor}) - h_{s,nT}(\overline{Z}_s^{\lambda,n}) \right] \mathrm{d}s \right| \\
& + \lambda \int_{nT}^t \left| h_{s,nT}(\overline{Z}_s^{\lambda,n}) - h(\overline{Z}_s^{\lambda,n}) \right| \mathrm{d}s \\
\leq & \lambda K_1 \int_{nT}^t \left| \tilde{Y}_s^\lambda(\mathbf{X}) - \overline{Z}_s^{\lambda,n} \right| \mathrm{d}s + \lambda \sup_{u \in [nT,(n+1)T]} \left| \int_{nT}^u \left[H(\overline{Z}_s^{\lambda,n}, X_{\lfloor s \rfloor}) - h_{s,nT}(\overline{Z}_s^{\lambda,n}) \right] \mathrm{d}s \right| \\
& + \lambda \int_{nT}^\infty \left| h_{s,nT}(\overline{Z}_s^{\lambda,n}) - h(\overline{Z}_s^{\lambda,n}) \right| \mathrm{d}s,
\end{aligned}$$

where $h_{s,nT}$ is defined in Lemma 3.13. Now let us apply Grönwall's lemma and take the square of both sides. Using the elementary $(x+y)^2 \leq 2(x^2+y^2)$, $x, y \geq 0$, we arrive at

$$\begin{aligned}
\left|\tilde{Y}_t^\lambda(\mathbf{X}) - \overline{Z}_t^{\lambda,n}\right|^2 \leq 2\lambda^2 \mathrm{e}^{2K_1\lambda T} \Bigg[& \sup_{u \in [nT,(n+1)T]} \left| \int_{nT}^u \left[H(\overline{Z}_s^{\lambda,n}, X_{\lfloor s \rfloor}) - h_{s,nT}(\overline{Z}_s^{\lambda,n}) \right] \mathrm{d}s \right|^2 \\
& + \left(\int_{nT}^\infty \left| h_{s,nT}(\overline{Z}_s^{\lambda,n}) - h(\overline{Z}_s^{\lambda,n}) \right| \mathrm{d}s \right)^2 \Bigg]. \quad (55)
\end{aligned}$$

As $s \to h_{s,nT}(\overline{Z}_s^{\lambda,n})$ and $s \to H(\overline{Z}_s^{\lambda,n}, X_{\lfloor s \rfloor})$ are a.s. piecewise continuous (see Lemma 3.13), Lemma 4.1 guarantees that

$$\begin{aligned}
& \sup_{u \in [nT,(n+1)T]} \left| \int_{nT}^u \left[H(\overline{Z}_s^{\lambda,n}, X_{\lfloor s \rfloor}) - h_{s,nT}(\overline{Z}_s^{\lambda,n}) \right] \mathrm{d}s \right| \\
& \qquad = \lim_{N \to \infty} \frac{1}{N} \max_{1 \leq k \leq \lfloor TN \rfloor} \left| \sum_{j=0}^k \left[H(\overline{Z}_{nT+j/N}^{\lambda,n}, X_{\lfloor nT+j/N \rfloor}) - h_{nT+j/N,nT}(\overline{Z}_{nT+j/N}^{\lambda,n}) \right] \right|, \quad (56)
\end{aligned}$$

a.s., with N ranging over integers. Defining $\mathcal{R}_j$ and $\mathcal{R}_j^+$ as in Lemma 3.18, the process

$$G_j := X_{\lfloor nT+j/N \rfloor}, \quad j \in \mathbb{N}$$

satisfies

$$M_p^0(G) = M_p^{nT}(\mathbf{X}) \text{ and } \Gamma_p^0(G) \leq 2N\Gamma_p^{nT}(\mathbf{X}), \quad (57)$$

for $p \geq 1$. Introduce the events $F_i := \{i \leq \sup_{s \in [nT,(n+1)T]} |\overline{Z}_s^{\lambda,n}| < i+1\}$, $i \in \mathbb{N}$, which are $\mathcal{H}_{nT}$ measurable, and fix i for the moment. Next, we apply Theorem 3.2 to the process

$$W_j := \left(H(\overline{Z}_{nT+j/N}^{\lambda,n}, X_{\lfloor nT+j/N \rfloor}) - h_{nT+j/N,nT}(\overline{Z}_{nT+j/N}^{\lambda,n}) \right) \mathbb{1}_{F_i}, \qquad j \in \mathbb{N}.$$

Clearly, $\mathbb{E}[W_j|\mathcal{R}_0] = \mathbb{E}[W_j|\mathcal{H}_{nT}] = 0$. From Lemma 3.4, we know that the estimates (15) (resp. (16)) hold for $M_p^{nT}(H(\theta,G), B(i))$ (resp. $\Gamma_p^{nT}(H(\theta,G), B(i))$). Noting that $\overline{Z}_{nT+j/N}^{\lambda,n}$ is $\mathcal{H}_{nT}$-measurable for $j \in \mathbb{N}$ and H is Lipschitz-continuous, Lemma 3.5 implies that the process

$$\bar{W}_j := H(\overline{Z}_{nT+j/N}^{\lambda,n}, X_{\lfloor nT+j/N \rfloor}) \mathbb{1}_{F_i}, \mathrm{for} j \in \mathbb{N} \quad (58)$$

satisfies

$$M_p^0(\bar{W}) \leq C^*[M_p^{nT}(\mathbf{X}) + i + 1], \text{ and } \qquad \Gamma_p^0(\bar{W}) \leq 4K_2N\Gamma_p^{nT}(\mathbf{X}).$$

where $C^* := \max\{K_1, K_2, H^*\}$. Then, by Remark 6.4 of [4],

$$M_p^0(W) \leq 2C^*[M_p^{nT}(\mathbf{X}) + i + 1], \; \Gamma_p^0(W) \leq 4K_2N\Gamma_p^{nT}(\mathbf{X}).$$

Apply Theorem 3.2 with $r := 3$ at $k = 0$. We obtain

$$\mathbb{E}^{1/2}\left[\max_{1 \leq k \leq \lfloor TN \rfloor} \left| \sum_{j=1}^k [H(\overline{Z}_{nT+j/N}^{\lambda,n}, X_{\lfloor nT+j/N \rfloor}) - h_{s,nT}(\overline{Z}_{nT+j/N}^{\lambda,n})] \right|^2 \mathbb{1}_{F_i} | \mathcal{H}_{nT} \right]$$

$$\leq \quad \mathbb{E}^{1/3}\left[\max_{1\leq k\leq\lfloor TN\rfloor}\left|\sum_{j=1}^{k}[H(\overline{Z}^{\lambda,n}_{nT+j/N},X_{\lfloor nT+j/N\rfloor})-h_{s,nT}(\overline{Z}^{\lambda,n}_{nT+j/N})]\right|^3\mathbb{1}_{F_i}|\mathcal{H}_{nT}\right]$$
$$\leq C(3)\sqrt{TN}\sqrt{2C^*[M_3^{nT}(\mathbf{X})+i+1]}\sqrt{4K_2N\Gamma_3^{nT}(\mathbf{X})}1_{F_i},$$

whence, by the conditional Fatou lemma and (56),

$$\mathbb{E}^{1/2}\left[\sup_{u\in[nT,(n+1)T]}\left|\int_{nT}^{u}[H(\overline{Z}^{\lambda,n}_s,X_{\lfloor s\rfloor})-h_{s,nT}(\overline{Z}^{\lambda,n}_s)]\,\mathrm{d}s\right|^2\mathbb{1}_{F_i}|\mathcal{H}_{nT}\right]$$
$$\leq C(3)\sqrt{T}\sqrt{2C^*[M_3^{nT}(\mathbf{X})+i+1]}\sqrt{4K_2\Gamma_3^{nT}(\mathbf{X})}1_{F_i}.$$

Fix $p_1\geq 1$, to be chosen later. We can then estimate, using Cauchy's inequality (twice) and Lemma 3.17,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{u\in[nT,(n+1)T]}\left|\int_{nT}^{u}[H(\overline{Z}^{\lambda,n}_s,X_{\lfloor s\rfloor})-h_{s,nT}(\overline{Z}^{\lambda,n}_s)]\,\mathrm{d}s\right|^2\right]\\
&\leq 8K_2C^2(3)C^*T\sum_{i=0}^{\infty}\mathbb{E}\left[\mathbb{1}_{F_i}[M_3^{nT}(\mathbf{X})+i+1]\Gamma_3^{nT}(\mathbf{X})\right]\\
&\leq 8K_2C^2(3)C^*T\sum_{i=0}^{\infty}\mathbb{P}^{1/2}(F_i)\sqrt{\mathbb{E}\left[([(M_3^{nT}(\mathbf{X})+i+1]\Gamma_3^{nT}(\mathbf{X}))^2\right]}\\
&\leq 8K_2C^2(3)C^*T\sum_{i=0}^{\infty}\mathbb{P}^{1/2}(F_i)\sqrt[4]{\mathbb{E}[(M_3^{nT}(\mathbf{X})+i+1)^4]}\sqrt[4]{\mathbb{E}[(\Gamma_3^{nT}(\mathbf{X}))^4]}\\
&\leq 8K_2C^2(3)C^*T\sum_{i=0}^{\infty}\sqrt{\frac{\mathbb{E}\left[\sup_{s\in[nT,(n+1)T]}(|\overline{Z}^{\lambda,n}_s|+1)^{p_1}\right]}{(i+1)^{p_1}}}\sqrt[4]{\mathbb{E}[(M_3^{nT}(\mathbf{X})+i+1)^4]}\sqrt[4]{\mathbb{E}[(\Gamma_3^{nT}(\mathbf{X}))^4]}\\
&\leq 8K_2C^2(3)C^*T\sqrt{\mathbb{E}[V_{2p_1}(\theta_0)]+3(1+ap_1)\mathrm{v}_{2p_1}(\overline{M}(p_1))}\\
&\qquad\times\sum_{i=0}^{\infty}\sqrt{\frac{1}{(i+1)^{p_1}}}\left[\sqrt[4]{\mathbb{E}[(M_3^{nT}(\mathbf{X}))^4]}+(i+1)\right]\sqrt[4]{\mathbb{E}[(\Gamma_3^{nT}(\mathbf{X}))^4]}\\
&\leq C_{20}\frac{1}{\lambda}
\end{aligned}\tag{59}$$

where

$$C_{20}=8K_2C^2(3)C^*\sqrt{\mathbb{E}[V_{2p_1}(\theta_0)]+3(1+ap_1)\mathrm{v}_{2p_1}(\overline{M}(2p_1))}\left[1+\sqrt[4]{\mathbb{E}[(M_3^{nT}(\mathbf{X}))^4]}\right]\sqrt[4]{\mathbb{E}[(\Gamma_3^{nT}(\mathbf{X}))^4]}\sum_{i=0}^{\infty}\frac{1}{(i+1)^{p_1/2-1}}.$$

Hence we set $p_1:=5$ (any $p_1>4$ would do) and obtain from (55) finally, for any $t\in[nT,(n+1)T]$, due to (59) and Lemma 3.14,

$$\begin{aligned}
\tilde{W}_2(\mathcal{L}(\tilde{Y}^{\lambda}_t(\mathbf{X})),\mathcal{L}(\overline{Z}^{\lambda,n}_t)) &\leq \mathbb{E}^{1/2}\left|\tilde{Y}^{\lambda}_t(\mathbf{X})-\overline{Z}^{\lambda,n}_t\right|^2\\
&\leq \sqrt{2}\mathrm{e}^{K_1}\sqrt{C_{20}\frac{1}{\lambda}\lambda^2+\lambda^2\mathbb{E}[\Xi_n^2]}\\
&\leq C_{21}\sqrt{\lambda},
\end{aligned}\tag{60}$$

where

$$C_{21}=\mathrm{e}^{K_1}\sqrt{2[C_{20}+C_{18}(2)]}.$$

We now turn to estimating $W_1(\mathcal{L}(\overline{Z}^{\lambda,n}_t),\mathcal{L}(L^{\lambda}_t))$. Recall that $n\in\mathbb{N}$ and $t\in[nT,(n+1)T)$ are fixed. Now we may write, using the triangle inequality, the definition of $\overline{Z}^{\lambda,n}$, (22), Lemma 3.16, (60), and Proposition 3.12

$$\begin{aligned}
&W_1(\mathcal{L}(\overline{Z}^{\lambda,n}_t),\mathcal{L}(L^{\lambda}_t)) \qquad (61)\\
&\qquad\leq \sum_{k=1}^{n}W_1(\mathcal{L}(\overline{Z}^{\lambda,k}_t),\mathcal{L}(\overline{Z}^{\lambda,k-1}_t))
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=1}^{n} W_1(\mathcal{L}(Z^\lambda(t,kT,\tilde{Y}^\lambda_{kT}(\mathbf{X}))),\mathcal{L}(Z^\lambda(t,(k-1)T,\tilde{Y}^\lambda_{(k-1)T}(\mathbf{X})))\\
&\le \sum_{k=1}^{n} w_{1,2}(\mathcal{L}(Z^\lambda(t,kT,\tilde{Y}^\lambda_{kT}(\mathbf{X}))),\mathcal{L}(Z^\lambda(t,kT,Z^\lambda(kT,(k-1)T,\tilde{Y}^\lambda_{(k-1)T}(\mathbf{X})))))\\
&\le C_9\sum_{k=1}^{n}\exp\left(-C_8(n-k)\right) w_{1,2}(\mathcal{L}(\tilde{Y}^\lambda_{kT}(\mathbf{X})),\mathcal{L}(Z^\lambda(kT,(k-1)T,\tilde{Y}^\lambda_{(k-1)T}(\mathbf{X}))))\\
&= C_9\sum_{k=1}^{n}\exp\left(-C_8(n-k)\right) w_{1,2}(\mathcal{L}(\tilde{Y}^\lambda_{kT}(\mathbf{X})),\mathcal{L}(\overline{Z}^{\lambda,k-1}_{kT}))\\
&\le \frac{C_9}{1-\exp\left(-C_8\right)}\max_{1\le k\le n} w_{1,2}(\mathcal{L}(\tilde{Y}^\lambda_{kT}(\mathbf{X})),\mathcal{L}(\overline{Z}^{\lambda,k-1}_{kT}))\\
&\le \frac{C_9}{1-\exp\left(-C_8\right)}\max_{1\le k\le n} \tilde{W}_2(\mathcal{L}(\tilde{Y}^\lambda_{kT}(\mathbf{X})),\mathcal{L}(\overline{Z}^{\lambda,k-1}_{kT}))\\
&\qquad\qquad \times\left[1+\{\mathbb{E}[V_4((\tilde{Y}^\lambda_{kT}(\mathbf{X}))]\}^{1/2}+\{\mathbb{E}[V_4(\overline{Z}^{\lambda,k-1}_{kT})]\}^{1/2}\right]\\
&\le C_{22}\sqrt{\lambda},
\end{aligned}
$$

where the penultimate inequality is due to Lemma 3.28 and

$$
\begin{aligned}
C_{22} = {}& \frac{C_9}{1-\exp\left(-\frac{C_8}{2}\right)}\sqrt{2}\mathrm{e}^{K_1}\sqrt{C_{20}+C_{18}(2)}\\
&\times\left[1+\sqrt{\mathbb{E}[V_4(\theta_0)]+3\mathrm{v}_4(\overline{M}(p))}+\sqrt{\mathbb{E}[V_4(\theta_0)]+6\mathrm{v}_4(\overline{M}(p))}\right],
\end{aligned}
$$

by Corollaries 3.8, 3.16. Now, putting together our estimations, we arrive at

$$
\begin{aligned}
W_1(\mathcal{L}(\tilde{Y}^\lambda_t(\mathbf{X})),\mathcal{L}(L^\lambda_t)) &\le W_1(\mathcal{L}(\tilde{Y}^\lambda_t(\mathbf{X})),\mathcal{L}(\overline{Z}^{\lambda,n}_t))+W_1(\mathcal{L}(\overline{Z}^{\lambda,n}_t),\mathcal{L}(L^\lambda_t))\\
&\le C_{19}\lambda^{1/2},
\end{aligned}
$$

where $C_{19}=C_{21}+C_{22}$, which finishes the proof. □

Remark 3.20. Our assumptions can be somewhat weakened. Indeed, the above arguments go through if we assume only that the sequences $M^n_3(X),\Gamma^n_3(X), n\in\mathbb{N}$ are bounded in L^4 and $|\theta_0|\in L^{10}$. The former propriety is called "conditional L-mixing of order $(3,4)$", see [4].

Corollary 3.21. *Let Assumptions 2.1, 2.2 and 2.5 hold. For each $0<\lambda\le\lambda_{\max}$ and $0\le s\le t$ let $\tilde{\zeta}^\lambda(t,s;\mathbf{x},\tilde{\theta})$ be the solution of* (26)*. with an initial condition $\tilde{\theta}$. Then for each $k\ge1$,*

$$
\begin{aligned}
&\mathbb{E}[V_4(\tilde{\zeta}^\lambda(kT,(k-1)T;\mathbf{x},Y^\lambda_{(k-1)T}(\mathbf{x})))]\\
&\le 2\left(\mathbb{E}[|\theta_0|^4]+a\lambda M(2,d)\sum_{j=0}^{(k-1)T-1}(1-a\lambda)^j\,|x_{(k-1)T-j}|^4+\hat{M}(2,d)\right)+3\mathrm{v}_4(\overline{M}(4))+2,
\end{aligned}
\tag{62}
$$

and

$$
\mathbb{E}[V_4(Y^\lambda_{(k-1)T})]\le 2+2\left(\mathbb{E}[|\theta_0|^4]+a\lambda M(2,d)\sum_{j=0}^{(k-1)T-1}(1-a\lambda)^j\,|x_{(k-1)T-j}|^4+\hat{M}(2,d)\right),\tag{63}
$$

where the constants $M(2,d)$ and $\hat{M}(2,d)$ are given by (39) *and* (40) *with $p=2$.*

Proof. A direct consequence of Lemma 3.7 and (38). □

Lemma 3.22. *Let Assumptions 2.1, 2.2 and 2.5 hold. For each $0<\lambda\le\lambda_{\max}$ and $n\in\mathbb{N}$ we have, for all $t\in(nT,(n+1)T]$,*

$$
W_1(\mathcal{L}(\tilde{Y}^\lambda_t(\mathbf{x})),\mathcal{L}(Y^\lambda_t(\mathbf{x})))\le C_{23}(\mathbf{x},n)\lambda^{1/2},\ \mathbf{x}\in(\mathbb{R}^m)^{\mathbb{N}}.
$$

where $C_{23}(\mathbf{x},n)$ is given in (69).

Proof. Recall (26) and observes that $\tilde{Y}^\lambda_t(\mathbf{x})=\tilde{\zeta}^\lambda(t,0;\mathbf{x},\theta_0)$. Then, one calculates for $t\in(nT,(n+1)T]$,

$$
W_1(\mathcal{L}(\tilde{Y}^\lambda_t(\mathbf{x})),\mathcal{L}(Y^\lambda_t(\mathbf{x})))=W_1(\mathcal{L}(\tilde{\zeta}^\lambda(t,0;\mathbf{x},\theta_0)),\mathcal{L}(Y^\lambda_t(\mathbf{x})))
$$

$$
\begin{aligned}
&\le \sum_{k=1}^{n} W_1(\mathcal{L}(\tilde{\zeta}^\lambda(t,kT;\mathbf{x},Y^\lambda_{kT}(\mathbf{x}))),\mathcal{L}(\tilde{\zeta}^\lambda(t,(k-1)T;\mathbf{x},Y^\lambda_{(k-1)T}(\mathbf{x}))))\\
&\quad + W_1(\mathcal{L}(\tilde{\zeta}^\lambda(t,nT;\mathbf{x},Y^\lambda_{nT}(\mathbf{x}))),\mathcal{L}(Y^\lambda_t(\mathbf{x})))\\
&\le \sum_{k=1}^{n} W_1(\mathcal{L}(\tilde{\zeta}^\lambda(t,kT;\mathbf{x},Y^\lambda_{kT}(\mathbf{x}))),\mathcal{L}(\tilde{\zeta}^\lambda(t,kT;\mathbf{x},\tilde{\zeta}^\lambda(kT,(k-1)T;\mathbf{x},Y^\lambda_{(k-1)T}(\mathbf{x}))))\\
&\quad + W_1(\mathcal{L}(\tilde{\zeta}^\lambda(t,nT;\mathbf{x},Y^\lambda_{nT}(\mathbf{x}))),\mathcal{L}(Y^\lambda_t(\mathbf{x}))),
\end{aligned}
$$

and thus, due to the domination of W_1 by $w_{1,2}$, see (22), and Proposition 3.12, see (50), one obtains

$$
\begin{aligned}
W_1(\mathcal{L}(\tilde{Y}^\lambda_t(\mathbf{x})),\mathcal{L}(Y^\lambda_t(\mathbf{x}))) \le& \sum_{k=1}^{n} w_{1,2}(\mathcal{L}(\tilde{\zeta}^\lambda(t,kT;\mathbf{x},Y^\lambda_{kT}(\mathbf{x}))),\mathcal{L}(\tilde{\zeta}^\lambda(t,kT;\mathbf{x},\tilde{\zeta}^\lambda(kT,(k-1)T;\mathbf{x},Y^\lambda_{(k-1)T}(\mathbf{x}))))\\
&+ w_{1,2}(\mathcal{L}(\tilde{\zeta}^\lambda(t,nT;\mathbf{x},Y^\lambda_{nT}(\mathbf{x}))),\mathcal{L}(Y^\lambda_t(\mathbf{x})))\\
\le& C_9 \sum_{k=1}^{n} e^{-C_8(n-k)} w_{1,2}(\mathcal{L}(Y^\lambda_{kT}(\mathbf{x})),\mathcal{L}(\tilde{\zeta}^\lambda(kT,(k-1)T;\mathbf{x},Y^\lambda_{(k-1)T}(\mathbf{x})))\\
&+ w_{1,2}(\mathcal{L}(\tilde{\zeta}^\lambda(t,nT;\mathbf{x},Y^\lambda_{nT}(\mathbf{x}))),\mathcal{L}(Y^\lambda_t(\mathbf{x}))).
\end{aligned}
\tag{64}
$$

At this point, one notes that due to Lemma 4.3, for any two probability measures μ and ν on $\mathbb{R}^d$,

$$
w_{1,2}(\mu,\nu) \le \sqrt{2}\left\{1+[\mu(V_4)]^{1/2}+[\nu(V_4)]^{1/2}\right\}\{\mathrm{KL}(\mu,\nu)\}^{1/2} .
$$

where $\mathrm{KL}(\mu,\nu)$ denotes the Kullback-Leibler distance of the two measures. Thus

$$
\begin{aligned}
& w_{1,2}(\mathcal{L}(Y^\lambda_{kT}(\mathbf{x})),\mathcal{L}(\tilde{\zeta}^\lambda(kT,(k-1)T;\mathbf{x},Y^\lambda_{(k-1)T}(\mathbf{x})))) && (65)\\
\le\ & \sqrt{2}\left\{1+\mathbb{E}^{1/2}[V_4(Y^\lambda_{kT}(\mathbf{x}))]+\mathbb{E}^{1/2}[V_4(\tilde{\zeta}^\lambda(kT,(k-1)T;\mathbf{x},Y^\lambda_{(k-1)T}(\mathbf{x})))]\right\} && (66)\\
\times\ & \left\{\mathrm{KL}(\mathcal{L}(Y^\lambda_{kT}(\mathbf{x})),\mathcal{L}(\tilde{\zeta}^\lambda(kT,(k-1)T;\mathbf{x},Y^\lambda_{(k-1)T}(\mathbf{x}))))\right\}^{1/2}.
\end{aligned}
$$

For $a<b$, $\mathbf{C}[a,b]$ denotes the Banach space of $\mathbb{R}^d$-valued continuous functions on the interval $[a,b]$. Let $\hat{\mathcal{Q}}$ denote the law of the process $\tilde{\zeta}^\lambda(s,(k-1)T;\mathbf{x},Y^\lambda_{(k-1)T}(\mathbf{x}))$, $s\in[(k-1)T,kT]$ on $\mathbf{C}[(k-1)T,kT]$. Similarly, let $\mathcal{Q}$ denote the law of $Y^\lambda_s(\mathbf{x})$, $s\in[(k-1)T,kT]$. Lemma 4.2 implies that these two probability laws are equivalent. Thus, in view of (75), one then calculates

$$
\begin{aligned}
\mathrm{KL}(\hat{\mathcal{Q}}\|\mathcal{Q}) &= -\mathbb{E}[\ln(\mathrm{d}\mathcal{Q}/\mathrm{d}\hat{\mathcal{Q}}(Y)]\\
&= \frac{\lambda^2}{2}\frac{\beta}{2\lambda}\int_{(k-1)T}^{kT}\mathbb{E}|H(Y^\lambda_{\lfloor s\rfloor}(\mathbf{x}),x_{\lfloor s\rfloor})-H(Y^\lambda_s(\mathbf{x}),x_{\lfloor s\rfloor})|^2\,\mathrm{d}s\\
&\le \frac{\lambda\beta K_1^2}{4}\int_{(k-1)T}^{kT}\mathbb{E}|Y^\lambda_{\lfloor s\rfloor}(\mathbf{x})-Y^\lambda_s(\mathbf{x})|^2\mathrm{d}s\\
&= \frac{\lambda\beta K_1^2}{4}\sum_{j=(k-1)T}^{kT-1}\int_j^{j+1}\mathbb{E}|-\lambda H(Y^\lambda_j(\mathbf{x}),x_j)(s-j)+\sqrt{2\lambda/\beta}(\tilde{B}^\lambda_s-\tilde{B}^\lambda_j)|^2\mathrm{d}s\\
&= \frac{\lambda\beta K_1^2}{4}\sum_{j=(k-1)T}^{kT-1}\left\{2\lambda^2\mathbb{E}|H(Y^\lambda_j(\mathbf{x}),x_j)|^2+\frac{4\lambda}{\beta}\right\}\\
&\le \frac{\lambda\beta K_1^2}{4}\sum_{j=(k-1)T}^{kT-1}\left\{6\lambda^2((H^*)^2+K_1^2\mathbb{E}|Y^\lambda_j(\mathbf{x})|^2+K_2^2|x_j|^2)+\frac{4\lambda}{\beta}\right\}\\
&= \frac{3}{2}\lambda\beta K_1^2\left\{\lambda(H^*)^2+\lambda^2K_1^2\sum_{j=(k-1)T}^{kT-1}\mathbb{E}|Y^\lambda_j(\mathbf{x}))|^2+\lambda^2K_2^2\sum_{j=(k-1)T}^{kT-1}|x_j|^2+\frac{1}{\beta}\right\}.
\end{aligned}
\tag{67}
$$

Notice that

$$
\frac{3}{2}\beta K_1^2\left(\lambda(H^*)^2+\lambda^2K_1^2\sum_{j=(k-1)T}^{kT-1}\left[\mathbb{E}|\theta_0|^2+\lambda c_0\sum_{l=0}^{j-1}(1-a\lambda)^l|x_{j-l}|^2+c_1\right]+\lambda^2K_2^2\sum_{l=(k-1)T}^{kT-1}|x_l|^2+\frac{1}{\beta}\right)
$$

$$
\begin{aligned}
&\leq \frac{3}{2}\beta K_1^2\lambda_{\max}\left((H^*)^2 + K_1^2\left[\mathbb{E}|\theta_0|^2 + c_1\right] + \lambda K_1^2 \sum_{j=(k-1)T}^{kT} \lambda c_0 \sum_{l=0}^{j-1}(1-a\lambda)^l |x_{j-l}|^2 + \lambda K_2^2 \sum_{l=(k-1)T}^{kT} |x_l|^2\right) \\
&+ \frac{3}{2}K_1^2\lambda_{\max} \qquad (68)\\
&=: \bar{C}(\mathbf{x},k).
\end{aligned}
$$

It follows that, for $k = 1, \ldots, n$,

$$
\begin{aligned}
& w_{1,2}(\mathcal{L}(Y^\lambda_{kT}(\mathbf{x})), \mathcal{L}(\tilde{\zeta}^\lambda(kT,(k-1)T;\mathbf{x},Y^\lambda_{(k-1)T}(\mathbf{x}))) \\
\leq\ & \sqrt{\lambda}\sqrt{2}\sqrt{\bar{C}(\mathbf{x},k)} \\
\times\ & \left(1 + 2\sqrt{\mathbb{E}[|\theta_0|^4] + a\lambda M(2,d)\sum_{j=0}^{(k-1)T-1}(1-a\lambda)^j\,|x_{(k-1)T-j}|^4 + \hat{M}(2,d) + 3\mathrm{v}_4(\overline{M}(4))}\right) \\
=:\ & \hat{C}(\mathbf{x},k)\sqrt{\lambda}.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
& w_{1,2}(\mathcal{L}(\tilde{\zeta}^\lambda(t,nT;\mathbf{x},Y^\lambda_{nT}(\mathbf{x}))), \mathcal{L}(Y^\lambda_t(\mathbf{x}))) \\
\leq\ & \sqrt{\lambda}\sqrt{2}\sqrt{\bar{C}(\mathbf{x},n+1)} \\
\times\ & \left(1 + 2\sqrt{\mathbb{E}[|\theta_0|^4] + a\lambda M(2,d)\sum_{j=0}^{nT-1}(1-a\lambda)^j\,|x_{nT-j}|^4 + \hat{M}(2,d) + 3\mathrm{v}_4(\overline{M}(4))}\right) \\
=:\ & \hat{C}(\mathbf{x},n+1)\sqrt{\lambda}.
\end{aligned}
$$

Note that $\hat{C}(\mathbf{x},k) \leq \hat{C}(\mathbf{x},n+1)$, $k = 1,\ldots,n$. Thus, due to Corollaries 3.9, 3.21 as well as (41) and (67), one concludes that the statement holds with

$$
C_{23}(\mathbf{x},n) := \left[1 + \sum_{k=1}^{n} C_9 e^{-C_8(n-k)}\right]\hat{C}(\mathbf{x},n+1). \qquad (69)
$$

□

Recall that $\mathcal{P}(\mathbb{R}^q)$ is the set of probability measures on $\mathcal{B}(\mathbb{R}^q)$ equipped with topology of weak convergence. It is known (see Section 8.3 of [3]) that $\mathcal{P}(\mathbb{R}^q)$ can be equipped with the structure of a complete separable metric space such that the generated topolology coincides with the topology of weak convergence.

Lemma 3.23. *The mappings $\tilde{\mu}: \mathbf{x} \to \mathcal{L}(\tilde{Y}^\lambda_t(\mathbf{x}))$ and $\mu: \mathbf{x} \to \mathcal{L}(Y^\lambda_t(\mathbf{x}))$ are $\mathcal{X}/\mathcal{B}(\mathcal{P}(\mathbb{R}^d))$-measurable for all $\lambda \leq \lambda_{\max}$.*

Proof. Recall that $\mathbf{x}^n \to \mathbf{x}$, $n \to \infty$ if and only if $\mathbf{x}^n_i \to \mathbf{x}_i$ for each coordinate $i \in \mathbb{N}$. We will show, by induction on $j \in \mathbb{N}$ that

$$
Y^\lambda_t(\mathbf{x}^n) \to Y^\lambda_t(\mathbf{x}) \qquad (70)
$$

for all $t \in (j, j+1]$ almost surely, $n \to \infty$. Note that (70) is trivial for $t = 0$.

Now notice that

$$
Y^\lambda_t(\mathbf{x}^n) = \lambda(t-j)H(Y^\lambda_j(\mathbf{x}^n), \mathbf{x}^n_j) + \sqrt{2\lambda}[\tilde{B}^\lambda_t - \tilde{B}^\lambda_j],
$$

so this tends a.s. to $Y^\lambda_t(\mathbf{x})$ as $n \to \infty$, by continuity of $H(\cdot,\cdot)$ and by the induction hypothesis. Since almost sure convergence entails convergence in law, this shows that μ is, in fact, a continuous functional of $\mathbf{x}$.

Now we turn our attention to $\tilde{\mu}$. For each $\mathbf{x} \in \mathcal{R}$, we define a recursive (Picard-type) iteration:

$$
D^0_s(\mathbf{x}) := \theta_0,\ 0 \leq s \leq t,\ D^{k+1}_s(\mathbf{x}) := \theta_0 + \lambda\int_0^s H(D^k_u(\mathbf{x}), \mathbf{x}_{\lfloor u \rfloor})\,du + \sqrt{2\lambda}\tilde{B}^\lambda_s,\ k \in \mathbb{N}.
$$

Define $\Phi_k(\mathbf{x}) := \mathcal{L}(D^k_t(\mathbf{x}))$, $\mathbf{x} \in \mathcal{R}$, $k \in \mathbb{N}$.

We now establish for each $k \in \mathbb{N}$ that, when $\mathbf{x}^n \to \mathbf{x}$, $n \to \infty$, we have $D^k_s(\mathbf{x}^n) \to D^k_s(\mathbf{x})$ in L^1 (hence also in law). We will verify by induction on k that

$$
\sup_{0 \leq s \leq t} \mathbb{E}|D^k_s(\mathbf{x}^n) - D^k_s(\mathbf{x})| \to 0,
$$

which is slightly more (but it is needed for the induction to work).

The case $k=0$ is trivial. Otherwise, using Lipschitz-continuity of $H(\cdot,\cdot)$,

$$\begin{aligned}
&\mathbb{E}|D_s^{k+1}(\mathbf{x}^n)-D_s^{k+1}(\mathbf{x})|\\
\leq\ &\lambda\int_0^s \mathbb{E}|H(D_u^k(\mathbf{x}^n),\mathbf{x}^n_{\lfloor u\rfloor})-H(D_u^k(\mathbf{x}),\mathbf{x}_{\lfloor u\rfloor})|\,\mathrm{d}u\\
\leq\ &\lambda\int_0^s \mathbb{E}|H(D_u^k(\mathbf{x}^n),\mathbf{x}^n_{\lfloor u\rfloor})-H(D_u^k(\mathbf{x}),\mathbf{x}^n_{\lfloor u\rfloor})|+\mathbb{E}|H(D_u^k(\mathbf{x}),\mathbf{x}^n_{\lfloor u\rfloor})-H(D_u^k(\mathbf{x}),\mathbf{x}_{\lfloor u\rfloor})|\,\mathrm{d}u\\
\leq\ &\lambda K\int_0^t \left[\mathbb{E}|D_u^k(\mathbf{x}^n)-D_u^k(\mathbf{x})|+\max_{0\leq i\leq\lfloor t\rfloor}|\mathbf{x}_i^n-\mathbf{x}_i^n|\right]\mathrm{d}u.
\end{aligned}$$

It follows that

$$\sup_{0\leq s\leq t}\mathbb{E}|D_s^{k+1}(\mathbf{x}^n)-D_s^{k+1}(\mathbf{x})|\leq\quad \lambda Kt\left[\sup_{0\leq s\leq t}\mathbb{E}|D_s^k(\mathbf{x}^n)-D_s^k(\mathbf{x})|+\max_{0\leq i\leq\lfloor t\rfloor}|\mathbf{x}_i^n-\mathbf{x}_i^n|\right],$$

which tends to 0 as $n\to\infty$ by the induction hypothesis and the definition of the convergence in $\mathcal{R}$. We deduce that, for each k, the functional $\Phi_k:\mathcal{R}\to\mathcal{P}$ is continuous on $\mathcal{R}$.

Noting $\theta_0\in L^2$, it is well-known (see e.g. Theorem 6.2.2 of [1]) that $D_t^k(\mathbf{x})\to\tilde{Y}_t^\lambda(\mathbf{x})$, $k\to\infty$ in L^2. This implies $\Phi_k(\mathbf{x})\to\mathcal{L}(\tilde{Y}_t^\lambda(\mathbf{x}))$ in law, for each $\mathbf{x}\in\mathcal{R}$, which shows that the functional $\tilde{\mu}$ is measurable, being a pointwise limit of continuous functionals. The proof is complete. □

The next lemma shows that the existence of "good" couplings for a family of random variables implies the existence of good couplings for their mixtures, too. This is known, see Corollary 5.22 of [25], nevertheless we provide a complete proof.

Lemma 3.24. *Let $(\mathcal{R},\mathcal{X})$ be a measurable space and let the mappings $\mu:\mathcal{R}\to\mathcal{P}$, $\tilde{\mu}:\mathcal{R}\to\mathcal{P}$ be $\mathcal{X}/\mathcal{B}(\mathcal{P})$-measurable. Let ζ be a probability law on $\mathcal{X}$. If $W_1(\tilde{\mu}(u),\mu(u))\leq\kappa(u)$ holds for every $u\in\mathcal{R}$ where $\kappa:\mathcal{R}\to[0,1]$ is a measurable function then*

$$W_1\left(\int_{\mathcal{R}}\tilde{\mu}(u)\,\zeta(du),\int_{\mathcal{R}}\mu(u)\,\zeta(du)\right)\leq\int_{\mathcal{R}}\kappa(u)\,\zeta(du).$$

Proof. Let us consider $\mathcal{P}(\mathbb{R}^{2d})$, the set of probabilities on $\mathcal{B}(\mathbb{R}^{2d})$ equipped with some complete, separable metric inducing the topology of weak convergence. This is a Polish space. Consider

$$\mathbb{D}:=\{(\mathfrak{p},\mathfrak{p})\in\mathcal{P}(\mathbb{R}^d)\times\mathcal{P}(\mathbb{R}^d):\ \mathfrak{p}\in\mathcal{P}\}=\cap_{n\in\mathbb{N}}\{(\mathfrak{q}_1,\mathfrak{q}_2):\mathfrak{q}_1,\mathfrak{q}_2\in\mathcal{P}(\mathbb{R}^d),\ W_1(\mathfrak{q}_1,\mathfrak{q}_2)<1/n\},$$

which is a Borel subset of $\mathcal{P}(\mathbb{R}^d)\times\mathcal{P}(\mathbb{R}^d)$, being an intersection of open sets.

Define the mappings $F_i:\mathcal{P}(\mathbb{R}^{2d})\to\mathcal{P}(\mathbb{R}^d)$, $i=1,2$ by

$$F_1(\pi):=\pi(\cdot,\mathbb{R}^d),\quad F_2(\pi):=\pi(\mathbb{R}^d,\cdot),\quad \pi\in\mathcal{P}(\mathbb{R}^{2d}).$$

They are clearly continuous since weak convergence of a sequence of probabilities implies weak convergence of their marginals, too.

Finally, let $F_3:\mathcal{P}(\mathbb{R}^{2d})\to\mathbb{R}_+$ be defined by

$$F_3(\pi):=\int_{\mathbb{R}^{2d}}[|x-y|\wedge 1]\pi(dx,dy),\ \pi\in\mathcal{P}(\mathbb{R}^{2d}).$$

This is again continuous, by the definition of weak convergence.

Define

$$\begin{aligned}
&A:=\{(u,\pi)\in\mathcal{R}\times\mathcal{P}(\mathbb{R}^{2d}):\\
&\pi(\cdot\times\mathbb{R}^d)=\tilde{\mu}(u),\pi(\mathbb{R}^d\times\cdot)=\mu(u),\\
&\int_{\mathbb{R}^{2d}}[|x-y|\wedge 1]\pi(dx,dy)\leq\kappa(u)\}.
\end{aligned}$$

By hypothesis, for each $u\in\mathcal{R}$ there is $\pi\in\mathcal{P}(\mathbb{R}^{2d})$ such that $(u,\pi)\in A$ (note that the infimum in the definition of W_1 is always attained, see Theorem 4.1 of [25]).

We claim that $A\in\mathcal{X}\times\mathfrak{B}$. Indeed, this is clear from the identity

$$A=\{(u,\pi):(F_1(\pi),\tilde{\mu}(u))\in\mathbb{D}\}\cap\{(u,\pi):(F_2(\pi),\mu(u))\in\mathbb{D}\}\cap\{(u,\pi):F_3(\pi)\leq\kappa(u)\},$$

from Borel-measurability of $\mathbb{D}$ and from the continuity/measurability of the functionals involved.

Hence the measurable selection theorem (see III.44-45. of [11]) implies that there is an $\mathcal{X}/\mathcal{B}(\mathcal{P}(\mathbb{R}^{2d}))$-measurable $F:\mathcal{R}\to\mathcal{P}(\mathbb{R}^{2d})$ such that, for ζ-almost every $u\in\mathcal{R}$, $(u,F(u))\in A$. Now let υ be the unique probability on $\mathcal{B}(\mathbb{R}^{2d})$ that satisfies

$$\int_{\mathbb{R}^{2d}}\phi(z)\upsilon(dz)=\int_{\mathcal{R}}\left(\int_{\mathbb{R}^{2d}}\phi(z)\,F(u)(dz)\right)\zeta(du),$$

for each continuous and bounded $\phi:\mathbb{R}^{2d}\to\mathbb{R}$. Clearly, the respective marginals of υ are the mixtures

$$\int_{\mathcal{R}}\tilde{\mu}(u)\zeta(du),\quad \int_{\mathcal{R}}\mu(u)\zeta(du).$$

By construction,

$$\begin{aligned}
&W_1\left(\int_{\mathcal{R}}\tilde{\mu}(u)\,\zeta(du),\int_{\mathcal{R}}\mu(u)\,\zeta(du)\right)\\
\leq\;&\int_{\mathbb{R}^{2d}}[|x-y|\wedge 1]\upsilon(dx,dy)\\
=\;&\int_{\mathcal{R}}\int_{\mathbb{R}^{2d}}[|x-y|\wedge 1]F(u)(dx,dy)\zeta(du)\\
\leq\;&\int_{\mathcal{R}}\kappa(u)\zeta(du).
\end{aligned}$$

□

Corollary 3.25. *For each* $0<\lambda\leq\lambda_{\max}$ *and* $t\in\mathbb{R}_+$, *we get*

$$W_1(\mathcal{L}(\tilde{Y}_t^{\lambda}(\mathbf{X})),\mathcal{L}(Y_t^{\lambda}(\mathbf{X})))\leq C_{24}\lambda^{1/2}$$

where $C_{24}:=\sup_{n\in\mathbb{N}}E[C_{23}(\mathbf{X},n)]<\infty$.

Proof. Recall first that as X is conditionally L-mixing, $A:=\sup_{n\in\mathbb{N}}\mathbb{E}[|X_n|^4]<\infty$. Fix n such that $n\leq t<n+1$. Let us denote by $\mathcal{X}$ the Borel sigma-field of $\mathcal{R}:=(\mathbb{R}^m)^{\mathbb{N}}$ and let ζ be the law of $\mathbf{X}$. Define

$$\tilde{\mu}(\mathbf{x}):=\mathcal{L}(\tilde{Y}_t^{\lambda}(\mathbf{x})),\quad \mu(\mathbf{x}):=\mathcal{L}(Y_t^{\lambda}(\mathbf{x})).$$

Lemma 3.23 below implies the measurability of these functionals. Let $\kappa(\mathbf{x},t):=C_{23}(\mathbf{x},n)\sqrt{\lambda}$, for each $\mathbf{x}\in\mathcal{R}$, where $C_{23}(\mathbf{x},n)$ is given in (69). Now the statement follows by Lemma 3.24 provided that we show $\sup_{n\in\mathbb{N}}\mathbb{E}[C_{23}(\mathbf{X},n)]<\infty$. By the definition of $\hat{C}(\mathbf{x},n+1)$ and by the Cauchy inequality, this boils down to showing

$$\begin{aligned}
&\sup_k\left[\left(\lambda\sum_{j=(k-1)T}^{kT}\lambda\sum_{l=0}^{j}(1-a\lambda)^lE|X_{j-l}|^2\right)+\left(\lambda\sum_{l=(k-1)T}^{kT}E|X_l|^2\right)\right]\\
\leq\;&\sup_k\left[\sqrt{A}\lambda\left(\sum_{j=(k-1)T}^{kT}\lambda\frac{1}{a\lambda}\right)+\sqrt{A}\lambda(T+1)\right]\\
\leq\;&\sqrt{A}\lambda(T+1)\left(\frac{1}{a}+1\right)<\infty
\end{aligned}$$

and

$$\begin{aligned}
&\sup_k a\lambda\sum_{j=0}^{kT}(1-a\lambda)^j\,E|X_{(k-1)T-j}|^4\\
\leq\;&a\lambda\frac{1}{a\lambda}A<\infty.
\end{aligned}$$

□

Lemma 3.26. *The contraction constant* C_8 *in Proposition 3.12 is given by the* $\min\{\phi,C_6(p),4C_7(p)\epsilon C_6(p)\}/2$, *where the explicit expressions for* $C_6(p)$ *and* $C_7(p)$ *can be found in Lemma 3.6. Furthermore,* ϵ *satisfies the following inequality*

$$\epsilon\leq 1\wedge\left(8C_7(2)\sqrt{\frac{\pi}{K_1}}\int_0^{\tilde{b}}\exp\left(\left(\frac{\sqrt{K_1}}{2}s+\frac{2}{\sqrt{K_1}}\right)^2\right)ds\right)^{-1},$$

and ϕ is given by

$$\phi = \left(\sqrt{\frac{4\pi}{K_1}} \bar{b} \exp \left(\left(\frac{\sqrt{K_1}}{2} \bar{b} + \frac{2}{\sqrt{K_1}} \right)^2 \right) \right)^{-1},$$

where $\tilde{b} = \sqrt{2C_7(p)/C_6(p) - 1}$, $\bar{b} = \sqrt{4C_7(p)(1 + C_6(p))/C_6(p) - 1}$ *and* $\Phi(\cdot)$ *is the c.d.f. of a standard normal random variable. The constant* C_9 *is given as the ratio of* C_{11}/C_{10}*, where* C_{11}*,* C_{10} *are given explicitly in the proof.*

Proof. Consider the Lyapunov function $V_p(\theta) = (|\theta|^2 + 1)^{p/2}$, for any $\theta \in \mathbb{R}^d$ and $p \geq 2$. Notice that $\nabla V_p(\theta) = p\theta(|\theta|^2 + 1)^{p/2-1}$. For any $\theta \in \mathbb{R}^d$, according to Lemma 3.6 (where $C_6(p)$ and $C_7(p)$ are given explicitly)

$$\mathcal{L}V_p(\theta) \leq C_7(p) - C_6(p)V_p(\theta).$$

As in [15], define a bounded non-decreasing function: $Q(\epsilon) : (0, \infty) \to [0, \infty)$ by

$$Q(\epsilon) = \sup \frac{|\nabla V_p|}{\max\{V_p, 1/\epsilon\}}.$$

In order to express $Q(\epsilon)$ in a more clear form using ϵ, we consider the following three cases:

1. Consider $\epsilon \in (0, 2^{-p/2})$. For $|\theta| < \sqrt{(1/\epsilon)^{2/p} - 1}$, we have $V_p(\theta) < 1/\epsilon$, and

   $$Q(\epsilon) = \sup_{|\theta| < \sqrt{(1/\epsilon)^{2/p} - 1}} \epsilon p|\theta|(|\theta|^2 + 1)^{p/2-1} = \epsilon^{2/p} p\sqrt{(1/\epsilon)^{2/p} - 1}.$$

   On the other hand, for $|\theta| \geq \sqrt{(1/\epsilon)^{2/p} - 1}$, $V_p(\theta) \geq 1/\epsilon$, and

   $$Q(\epsilon) = \sup \frac{p|\theta|}{|\theta|^2 + 1} = \epsilon^{2/p} p\sqrt{(1/\epsilon)^{2/p} - 1},$$

   since for $\epsilon \in (0, 2^{-p/2})$, $|\theta| > 1$. Therefore, $Q(\epsilon) = \epsilon^{2/p} p\sqrt{(1/\epsilon)^{2/p} - 1} < p/2$ for all $\epsilon \in (0, 2^{-p/2})$.

2. For the second case, consider $\epsilon \in (2^{-p/2}, 1)$. Then, by using the same arguments as above, one obtains for $|\theta| < \sqrt{(1/\epsilon)^{2/p} - 1}$, $Q(\epsilon) = \epsilon^{2/p} p\sqrt{(1/\epsilon)^{2/p} - 1}$, while for $|\theta| \geq \sqrt{(1/\epsilon)^{2/p} - 1}$, $Q(\epsilon) = p/2$. Thus, one obtains $Q(\epsilon) = p/2$ for all $\epsilon \in (2^{-p/2}, 1)$.

3. Finally, for $\epsilon \in (1, \infty)$, we have $Q(\epsilon) = p/2$, since $V_p(\theta) \geq 1$ for all $\theta \in \mathbb{R}^d$.

In the first two cases above, we use the fact that $p/2 \geq \epsilon^{2/p} p\sqrt{(1/\epsilon)^{2/p} - 1}$ for all $\epsilon \in (0, 1)$. Indeed, squaring both sides, we have

$$1 \geq 4\epsilon^{4/p}((1/\epsilon)^{2/p} - 1) \iff 4\epsilon^{4/p} - 4\epsilon^{2/p} + 1 \geq 0 \iff (2\epsilon^{2/p} - 1)^2 \geq 0.$$

Combining all the three cases, one obtains $Q(\epsilon) \leq p/2$ for all $\epsilon > 0$. To calculate R_1 and R_2, notice that

$$R_1 \leq 2 \sup\{|\theta| : V_p(\theta) \leq 2C_7(p)/C_6(p)\} \implies R_1 \leq \sqrt{(2C_7(p)/C_6(p))^{2/p} - 1};$$
$$R_2 \leq 2 \sup\{|\theta| : V_p(\theta) \leq 4C_7(p)(1 + C_6(p))/C_6(p)\} \implies R_2 \leq \sqrt{(4C_7(p)(1 + C_6(p))/C_6(p))^{2/p} - 1}.$$

According to Theorem 2.2 in [15], we require that

$$(4C_7(p)\epsilon)^{-1} \geq \int_0^{R_1} \int_0^s \exp\left(\frac{1}{2}\int_r^s u\kappa(u)\, du + 2Q(\epsilon)(s - r)\right) dr\, ds,$$

where in our case $\kappa(u) = K_1$ and $Q(\epsilon) = p/2$. Then, one calculates

$$\begin{aligned}
(4C_7(p)\epsilon)^{-1} &\geq \int_0^{R_1} \int_0^s \exp\left(\frac{1}{2}\int_r^s K_1 u\, du + p(s - r)\right) dr\, ds \\
&= \int_0^{R_1} \int_0^s \exp\left(\frac{K_1}{4}(s^2 - r^2) + p(s - r)\right) dr\, ds \\
&= \int_0^{R_1} \exp\left(\left(\frac{\sqrt{K_1}}{2}s + \frac{p}{\sqrt{K_1}}\right)^2\right) \int_0^s \exp\left(-\left(\frac{\sqrt{K_1}}{2}r + \frac{p}{\sqrt{K_1}}\right)^2\right) dr\, ds,
\end{aligned}$$

which implies by setting $v/\sqrt{2} = \sqrt{K_1}r/2 + p/\sqrt{K_1}$, $(dv = \sqrt{K_1/2}dr)$

$$\begin{aligned}(4C_7(p)\epsilon)^{-1} &\geq \sqrt{\frac{2}{K_1}} \int_0^{R_1} \exp\left(\left(\frac{\sqrt{K_1}}{2}s + \frac{p}{\sqrt{K_1}}\right)^2\right) \int_{p\sqrt{2/K_1}}^{\sqrt{K_1/2}s+p\sqrt{2/K_1}} \exp\left(-\frac{v^2}{2}\right) dv\, ds \\ &= \sqrt{\frac{4\pi}{K_1}} \int_0^{\tilde{b}} \exp\left(\left(\frac{\sqrt{K_1}}{2}s + \frac{p}{\sqrt{K_1}}\right)^2\right) \left(\Phi\left(\sqrt{K_1/2}s + p\sqrt{2/K_1}\right) - \Phi\left(p\sqrt{2/K_1}\right)\right) ds,\end{aligned}$$

where $\tilde{b} = \sqrt{(2C_7(p)/C_6(p))^{2/p} - 1} > 0$ and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. By using the property of the cumulative distribution function and to ease the calculations of C_{10} and C_{11} below, it is enough for ϵ to satisfy the following inequality:

$$\epsilon \leq 1 \wedge \left(8C_7(p)\sqrt{\frac{\pi}{K_1}} \int_0^{\tilde{b}} \exp\left(\left(\frac{\sqrt{K_1}}{2}s + \frac{p}{\sqrt{K_1}}\right)^2\right) ds\right)^{-1}.$$

To obtain ϕ, we set $Q(\epsilon) = p/2$ and calculate

$$\begin{aligned}\phi^{-1} &= \int_0^{R_2} \int_0^s \exp\left(\frac{1}{2}\int_r^s K_1 u\, du + p(s-r)\right) dr\, ds \\ &= \sqrt{\frac{4\pi}{K_1}} \int_0^{\bar{b}} \exp\left(\left(\frac{\sqrt{K_1}}{2}s + \frac{p}{\sqrt{K_1}}\right)^2\right) \left(\Phi\left(\sqrt{K_1/2}s + p\sqrt{2/K_1}\right) - \Phi\left(p\sqrt{2/K_1}\right)\right) ds,\end{aligned}$$

where $\bar{b} = \sqrt{(4C_7(p)(1 + C_6(p))/C_6(p))^{2/p} - 1} > 0$. One notices that $\phi \geq \bar{\phi} := \left(\sqrt{\frac{4\pi}{K_1}}\bar{b} \exp\left(\left(\frac{\sqrt{K_1}}{2}\bar{b} + \frac{2}{\sqrt{K_1}}\right)^2\right)\right)^{-1}$, and

$$C_8 = \min\{\phi, C_6(p), 4C_7(p)\epsilon C_6(p)\}/2 \geq \bar{C}_8 := \min\{\bar{\phi}, C_6(p), 4C_7(p)\epsilon C_6(p)\}/2,$$

which implies $e^{-C_8} \leq e^{-\bar{C}_8}$. Thus, we set $\phi = \bar{\phi} := \left(\sqrt{\frac{4\pi}{K_1}}\bar{b} \exp\left(\left(\frac{\sqrt{K_1}}{2}\bar{b} + \frac{2}{\sqrt{K_1}}\right)^2\right)\right)^{-1}$ and Proposition 3.12 still holds.

As for C_{10} and C_{11}, one notes that $\rho_2 = f(|\theta-\theta'|)(1+\epsilon V_2(\theta)+\epsilon V_2(\theta'))$, where $\epsilon \in (0, 1)$, $\frac{1}{2}F(r) \leq f(r) \leq F(r)$ for $r \leq R_2$ and $f(r) = f(R_2)$ for $r \geq R_2$, moreover, $r \exp(-K_1 R_2^2/4 - pR_2) \leq F(r) \leq r$ for all $r \leq R_2$ and $f(r) \leq R_2$ for all $r > 0$. To calculate C_{10}, one calculates for $r \leq R_2$

$$\begin{aligned}[1 \wedge |\theta - \theta'|](1 + V_2(\theta) + V_2(\theta')) &\leq \epsilon^{-1}|\theta - \theta'|(\epsilon + \epsilon V_2(\theta) + \epsilon V_2(\theta')) \\ &\leq 2\epsilon^{-1} \exp(K_1 R_2^2/4 + pR_2)\left(\frac{1}{2}F(|\theta - \theta'|)\right)(1 + \epsilon V_2(\theta) + \epsilon V_2(\theta')) \\ &\leq \bar{C}_{10}^{-1} f(|\theta - \theta'|)(1 + \epsilon V_2(\theta) + \epsilon V_2(\theta')),\end{aligned}$$

where $\bar{C}_{10} = \epsilon/2 \exp(-K_1 R_2^2/4 - pR_2)$. For $r > R_2$

$$\begin{aligned}f(|\theta - \theta'|)(1 + \epsilon V_2(\theta) + \epsilon V_2(\theta')) &= f(R_2)(1 + \epsilon V_2(\theta) + \epsilon V_2(\theta')) \\ &\geq \tilde{C}_{10}[1 \wedge |\theta - \theta'|](1 + V_2(\theta) + V_2(\theta')),\end{aligned}$$

where $\tilde{C}_{10} = \epsilon/2R_2 \exp(-K_1 R_2^2/4 - pR_2)$, and $C_{10} = \min\{\bar{C}_{10}, \tilde{C}_{10}\}$. To calculate C_{11}, one considers, for $r \leq R_2$

$$\begin{aligned}f(|\theta - \theta'|)(1 + \epsilon V_2(\theta) + \epsilon V_2(\theta')) &\leq |\theta - \theta'|(1 + \epsilon V_2(\theta) + \epsilon V_2(\theta')) \\ &\leq C_{11}[1 \wedge |\theta - \theta'|](1 + V_2(\theta) + V_2(\theta')),\end{aligned}$$

where $C_{11} = 1 + R_2$. Then, for $r > R_2$

$$\begin{aligned}f(|\theta - \theta'|)(1 + \epsilon V_2(\theta) + \epsilon V_2(\theta')) &= f(R_2)(1 + \epsilon V_2(\theta) + \epsilon V_2(\theta')) \\ &\leq C_{11}[1 \wedge |\theta - \theta'|](1 + V_2(\theta) + V_2(\theta')).\end{aligned}$$

Thus $C_{11} = 1 + R_2$. □

3.5 Proof of Main Results

Proof of Theorem 2.7. Under our assumptions, π_β is a stationary law for L_t^λ, $t \in \mathbb{R}_+$. As

$$\begin{aligned} & W_1(\mathcal{L}(Y_t^\lambda(\mathbf{X}), \pi_\beta) \\ \leq\ & W_1(\mathcal{L}(Y_t^\lambda(\mathbf{X}), \mathcal{L}(\tilde{Y}_t^\lambda(\mathbf{X})) + W_1(\mathcal{L}(\tilde{Y}_t^\lambda(\mathbf{X})), \mathcal{L}(L_t^\lambda)) + W_1(\mathcal{L}(L_t^\lambda), \pi_\beta) \\ \leq\ & [C_{19} + C_{24}]\lambda^{1/2} + C_9 \mathrm{e}^{-C_8 \lambda t} w_{1,2}(\theta_0, \pi_\beta), \end{aligned}$$

by Corollary 3.25, Lemma 3.19, Proposition 3.12 and by (22). This implies the statement. $\square$

Proof of Theorem 2.9. Using Corollary 2 in [14] instead of Corollary 2.3 in [15] we can establish Proposition 3.12 with $\tilde{W}_1$ in lieu of $w_{1,2}$. Now the arguments for the proof of Theorem 2.7 can be repeated verbatim up to (61). We then get

$$\begin{aligned} & \tilde{W}_1(\mathcal{L}(\overline{Z}_t^{\lambda,n}), \mathcal{L}(L_t^\lambda)) \\ \leq\ & \sum_{k=1}^n \tilde{W}_1(\mathcal{L}(\overline{Z}_t^{\lambda,k}), \mathcal{L}(\overline{Z}_t^{\lambda,k-1})) \\ =\ & \sum_{k=1}^n \tilde{W}_1(\mathcal{L}(Z^\lambda(t, kT, \tilde{Y}_{kT}^\lambda(\mathbf{X}))), \mathcal{L}(Z^\lambda(t, (k-1)T, \tilde{Y}_{(k-1)T}^\lambda(\mathbf{X}))) \\ \leq\ & C_9 \sum_{k=1}^n \exp(-C_8(n-k))\, \tilde{W}_1(\mathcal{L}(\tilde{Y}_{kT}^\lambda(\mathbf{X})), \mathcal{L}(Z^\lambda(kT, (k-1)T, \tilde{Y}_{(k-1)T}^\lambda(\mathbf{X})))) \\ =\ & C_9 \sum_{k=1}^n \exp(-C_8(n-k))\, \tilde{W}_1(\mathcal{L}(\tilde{Y}_{kT}^\lambda(\mathbf{X})), \mathcal{L}(\overline{Z}_{kT}^{\lambda,k-1})) \\ \leq\ & \frac{C_9}{1 - \exp(-C_8)} \max_{1 \leq k \leq n} \tilde{W}_1(\mathcal{L}(\tilde{Y}_{kT}^\lambda(\mathbf{X})), \mathcal{L}(\overline{Z}_{kT}^{\lambda,k-1})) \\ \leq\ & \frac{C_9}{1 - \exp(-C_8)} \max_{1 \leq k \leq n} \tilde{W}_2(\mathcal{L}(\tilde{Y}_{kT}^\lambda(\mathbf{X})), \mathcal{L}(\overline{Z}_{kT}^{\lambda,k-1})) \\ \leq\ & C_{22}\sqrt{\lambda}, \end{aligned}$$

hence Lemma 3.19 holds with $\tilde{W}_1$ instead of W_1 and the result follows just like Theorem 2.7 above. $\square$

Proof of Theorem 2.11. Using Proposition 1 of [13] instead of Corollary 2.3 in [15] we can establish Proposition 3.12 with $\tilde{W}_2$ in lieu of $w_{1,2}$. The arguments for the proof of Theorem 2.7 can be repeated verbatim up to (61), which need to be replaced by

$$\begin{aligned} & \tilde{W}_2(\mathcal{L}(\overline{Z}_t^{\lambda,n}), \mathcal{L}(L_t^\lambda)) \\ \leq\ & \sum_{k=1}^n \tilde{W}_2(\mathcal{L}(\overline{Z}_t^{\lambda,k}), \mathcal{L}(\overline{Z}_t^{\lambda,k-1})) \\ =\ & \sum_{k=1}^n \tilde{W}_2(\mathcal{L}(Z^\lambda(t, kT, \tilde{Y}_{kT}^\lambda(\mathbf{X}))), \mathcal{L}(Z^\lambda(t, (k-1)T, \tilde{Y}_{(k-1)T}^\lambda(\mathbf{X}))) \\ \leq\ & C_9 \sum_{k=1}^n \exp(-C_8(n-k))\, \tilde{W}_2(\mathcal{L}(\tilde{Y}_{kT}^\lambda(\mathbf{X})), \mathcal{L}(Z^\lambda(kT, (k-1)T, \tilde{Y}_{(k-1)T}^\lambda(\mathbf{X})))) \\ =\ & C_9 \sum_{k=1}^n \exp(-C_8(n-k))\, \tilde{W}_2(\mathcal{L}(\tilde{Y}_{kT}^\lambda(\mathbf{X})), \mathcal{L}(\overline{Z}_{kT}^{\lambda,k-1})) \\ \leq\ & \frac{C_9}{1 - \exp(-C_8)} \max_{1 \leq k \leq n} \tilde{W}_2(\mathcal{L}(\tilde{Y}_{kT}^\lambda(\mathbf{X})), \mathcal{L}(\overline{Z}_{kT}^{\lambda,k-1})) \\ \leq\ & \frac{C_9}{1 - \exp(-C_8)} \max_{1 \leq k \leq n} \tilde{W}_2(\mathcal{L}(\tilde{Y}_{kT}^\lambda(\mathbf{X})), \mathcal{L}(\overline{Z}_{kT}^{\lambda,k-1})) \\ \leq\ & C_{22}\sqrt{\lambda}, \end{aligned}$$

hence Lemma 3.19 holds with $\tilde{W}_2$ instead of W_1 and the result follows just like Theorem 2.7 above. $\square$

Remark 3.27. A careful examination of our estimates reveals how the constant in Theorem 2.7 depends on β and on d. It's easy to see that $C_8 \leq C_6(2) = a/2$ and $C_{24} = \mathcal{O}(\beta^{1/2})$ which do not depend on d. From (31), and noting that $p_1 = 5$, we have $\overline{M}(10) = \mathcal{O}\left(\sqrt{\frac{d}{\beta}}\right)$ and thus $v_{10}(\overline{M}(10)) = \mathcal{O}\left(\frac{d^5}{\beta^5}\right)$. It follows that $C_{20} = \mathcal{O}\left(\frac{d^{5/2}}{\beta^{5/2}}\right)$, $C_{21} = \mathcal{O}\left(\frac{d^{5/4}}{\beta^{5/4}}\right)$, $C_{22} = \mathcal{O}\left(\frac{d^{9/4}}{\beta^{9/4}}\right)$ and therefore $C_{19} = C_{21} + C_{22} = \mathcal{O}\left(\frac{d^{9/4}}{\beta^{9/4}}\right)$. Observing from Lemma 3.6 that the constant $C_7(p) = \mathcal{O}\left(\frac{d^{p/2}}{\beta^{p/2}}\right)$, one obtains that $R_2 = \mathcal{O}\left(\frac{d^{1/2}}{\beta^{1/2}}\right)$ and $C_{11} = \mathcal{O}\left(\frac{d^{1/2}}{\beta^{1/2}}\right)$. On the other hand, we observe that

$$\epsilon = \mathcal{O}\left(\frac{\beta}{d\int_0^{\sqrt{d/\beta}} e^{s^2} ds}\right)$$

and

$$\bar{C}_{10} = \mathcal{O}\left(\frac{\beta}{d\int_0^{\sqrt{d/\beta}} e^{s^2} ds} e^{-d/\beta}\right), \qquad \tilde{C}_{10} = \mathcal{O}\left(\frac{\beta}{d\int_0^{\sqrt{d/\beta}} e^{s^2} ds} \frac{d^{1/2}}{\beta^{1/2}} e^{-d/\beta}\right).$$

The constant $C_{10} = \min\{\bar{C}_{10}, \tilde{C}_{10}\}$ and so

$$C_9 = C_{11}/C_{10} = \mathcal{O}\left(\frac{d^{3/2}}{\beta^{3/2}} e^{d/\beta} \int_0^{\sqrt{d/\beta}} e^{s^2} ds\right) \leq \mathcal{O}\left(\frac{d^{3/2}}{\beta^{3/2}} e^{2d/\beta}\right).$$

One further observes at this point that C_9 is a consequence of Corollary 2.3 in [15]. Thus, any further improvement with the coupling arguments which relate to the dependency on the dimension will provide a significant improvement here.

Lemma 3.28. *Let $\mu, \nu \in \mathcal{P}_{V_4}$. Then*

$$w_{1,2}(\mu,\nu) \leq \tilde{W}_2(\mu,\nu)[1 + \mu^{1/2}(V_4) + \nu^{1/2}(V_4)].$$

Proof. Indeed, by the Cauchy and Minkowski inequalities,

$$\begin{aligned} w_{1,2}(\mu,\nu) \leq & \inf_{\zeta \in \mathcal{C}(\mu,\nu)} \left(\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |\theta - \theta'|^2 \zeta(\mathrm{d}\theta \mathrm{d}\theta')\right)^{1/2} \left(\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} [1 + V_2(\theta) + V_2(\theta')]^2 \zeta(\mathrm{d}\theta \mathrm{d}\theta')\right)^{1/2} \\ \leq & \tilde{W}_2(\mu,\nu)[1 + \mu^{1/2}(V_2^2) + \nu^{1/2}(V_2^2)] = \tilde{W}_2(\mu,\nu)[1 + \mu^{1/2}(V_4) + \nu^{1/2}(V_4)]. \end{aligned}$$

using also the definition of V_p for $p = 2, 4$. The statement follows. ☐

4 Appendix

Lemma 4.1. *Let f be a piecewise continuous càdlàg function on $[a, b]$. It holds that*

$$\sup_{u \in [a,b]} \left|\int_a^u f(s) ds\right| = \lim_{N \to \infty} \frac{1}{N} \max_{1 \leq k \leq \lfloor (b-a)N \rfloor} \left|\sum_{j=0}^{k} f(a + j/N)\right|.$$

Proof. Without loss of generality, we may assume that $[a, b] = [0, 1]$. As càdlàg functions defined on a compact interval are bounded, the sequence of simple functions

$$f_N(s) = \sum_{j=0}^{N-1} f(j/N) 1_{s \in [j/N, (j+1)/N)}, N \in \mathbb{N}$$

converges a.e. to f on $[0, 1]$. Therefore,

$$\left|\sup_{u \in [0,1]} \left|\int_0^u f_N(s) ds\right| - \sup_{u \in [0,1]} \left|\int_0^u f(s) ds\right|\right| \leq \sup_{u \in [0,1]} \left|\int_0^u (f_N(s) - f(s)) ds\right| \leq \int_0^1 |f_N(s) - f(s)| ds \to 0,$$

by the dominated convergence theorem. ☐

We present a simpler version of [19, Theorem 7.19], which is suitable for the purposes of this article.

Lemma 4.2. *Let $(\xi_t)_{t \geq 0}$ and $(\eta_t)_{t \geq 0}$ be two diffusion type processes with*

$$\mathrm{d}\xi_t = \alpha_t(\xi)\mathrm{d}t + \sigma \mathrm{d}B_t, \qquad \textit{for } t > 0, \tag{71}$$

and

$$\mathrm{d}\eta_t = b_t(\eta)\mathrm{d}t + \sigma \mathrm{d}B_t \qquad \textit{for } t > 0, \tag{72}$$

where $\xi_0 = \eta_0$ is an $\mathcal{F}_0$ measurable random variable and c is a positive constant. Suppose also that the nonanticipative functionals $\alpha_t(x)$ and $b_t(x)$ are such that a unique (continuous) strong solution exist for (71) and (72) respectively. If, for any fixed $T > 0$,

$$\int_0^T [|\alpha_s(\xi)|^2 + |b_s(\xi)|^2]\mathrm{d}s < \infty \text{ (a.s.) and } \int_0^T [|\alpha_s(\eta)|^2 + |b_s(\eta)|^2]\mathrm{d}s < \infty \text{ (a.s.),}$$

then $\mu_\xi^T = \mathcal{L}(\xi_{[0,T]}) \backsim \mu_\eta^T = \mathcal{L}(\eta_{[0,T]})$ and the densities are given by

$$\frac{\mathrm{d}\mu_\eta^T}{\mathrm{d}\mu_\xi^T}(\xi) = \exp\left(-\sigma^{-2}\int_0^T \langle \alpha_s(\xi) - b_s(\xi), \mathrm{d}\xi_s\rangle + \frac{1}{2\sigma^2}\int_0^T [|\alpha_s(\xi)|^2 - |b_s(\xi)|^2]\mathrm{d}s\right) \tag{73}$$

and

$$\frac{\mathrm{d}\mu_\xi^T}{\mathrm{d}\mu_\eta^T}(\eta) = \exp\left(\sigma^{-2}\int_0^T \langle \alpha_s(\eta) - b_s(\eta), \mathrm{d}\eta_s\rangle - \frac{1}{2\sigma^2}\int_0^T [|\alpha_s(\eta)|^2 - |b_s(\eta)|^2]\mathrm{d}s\right). \tag{74}$$

Finally, the Kullback-Leibler divergence is given by

$$\mathrm{KL}(\mu_\xi^T, \mu_\eta^T) = \frac{1}{2}\mathbb{E}\left[\int_0^T |\alpha_s(\xi) - b_s(\xi)|^2\mathrm{d}s\right]. \tag{75}$$

Proof. The proof follows from a straightforward extension to the vector-case of [19, Theorem 7.19]. The computation of the Kullback-Leibler distance is a direct application of the definition. □

Let $V : \mathbb{R}^d \to [1, \infty)$ be a measurable function. For a measurable function $f : \mathbb{R}^d \to \mathbb{R}$, the V-norm of f is given by $\|f\|_V = \sup_{x\in\mathbb{R}^d} |f(x)|/V(x)$. For ξ and ξ' two probability measures on $\mathbb{R}^d$, the V-total variation distance of ξ and ξ' is given by

$$\|\xi - \xi'\|_V = \sup_{\|f\|_V\leq 1}\int_{\mathbb{R}^d} f(\theta)\mathrm{d}\{\xi - \xi'\}(\theta).$$

If $V \equiv 1$, then $\|\cdot\|_V$ is the total variation distance. The V-total variation distance is also characterized in terms of coupling (see [20, Theorem 19.1.7]):

$$\|\xi - \xi'\|_V = \inf_{\zeta\in\mathcal{C}(\xi,\xi')} \iint_{\mathbb{R}^d\times\mathbb{R}^d} \{V(\theta) + V(\theta')\}\mathbb{1}_{\{\theta\neq\theta'\}}\zeta(\mathrm{d}\theta, \mathrm{d}\theta')$$

where $\mathcal{C}(\xi, \xi')$ is the set of coupling of ξ and ξ'. An optimal coupling is given by (see [20, Theorem 19.1.6])

$$\gamma^*(B) = \{1 - \xi \wedge \xi'(\mathbb{R}^d)\}\beta(B) + \int_B \xi \wedge \xi'(\mathrm{d}\theta)\delta_\theta(\mathrm{d}\theta')$$

where $\xi \wedge \xi'$ is the infimum of probability measures ξ and ξ' and β is any coupling of η and η' where

$$\eta = \frac{\xi - \xi \wedge \xi'}{1 - \xi \wedge \xi'(\mathbb{R}^d)} \quad \text{and} \quad \eta' = \frac{\xi - \xi \wedge \xi'}{1 - \xi \wedge \xi'(\mathbb{R}^d)}$$

Lemma 4.3. *For any probability measures ξ and ξ' on $\mathbb{R}^d$, and $p \geq 1$, we get*

$$w_{1,p}(\xi, \xi') \leq \sqrt{2}\left\{1 + [\xi(V_{2p})]^{1/2} + [\xi'(V_{2p})]^{1/2}\right\}\{\mathrm{KL}(\xi, \xi')\}^{1/2}.$$

Proof.

$$\begin{aligned}
w_{1,p}(\xi, \xi') &= \inf_{\zeta\in\mathcal{C}(\xi,\xi')} \iint_{\mathbb{R}^{2d}} (1 \wedge |\theta - \theta'|)\{1 + V_p(\theta) + V_p(\theta')\}\zeta(\mathrm{d}\theta\mathrm{d}\theta') \\
&\leq \iint_{\mathbb{R}^{2d}} (1 \wedge |\theta - \theta'|)\{1 + V_p(\theta) + V_p(\theta')\}\gamma^*(\mathrm{d}\theta\mathrm{d}\theta') \\
&\leq \{1 - \xi \wedge \xi'(\mathbb{R}^d)\} \iint_{\mathbb{R}^{2d}} \{1 + V_p(\theta) + V_p(\theta')\}\beta(\mathrm{d}\theta\mathrm{d}\theta') \\
&= \{1 - \xi \wedge \xi'(\mathbb{R}^d)\} + (\xi - \xi \wedge \xi')(V_p) + (\xi' - \xi \wedge \xi')(V_p) \\
&= \|\xi - \xi'\|_{\mathrm{TV}} + \|\xi - \xi'\|_{V_p}.
\end{aligned}$$

The proof then follows from the weighted Pinsker's inequality; see [12, Lemma 24]. □

Lemma 4.4. *Let x, $y \in \mathbb{R}^d$, then*

$$\sum_{\substack{i+j+k=p\\ \{i\neq p-1\}\cap\{j\neq 1\}}} \frac{p!}{i!j!k!}\|x\|^{2i}\big(2\langle x,y\rangle\big)^j\|y\|^{2k} \leq \sum_{\substack{k=0\\ k\neq 1}}^{2p} \binom{2p}{k}\|x\|^{2p-k}\|y\|^k$$

Proof. Note that

$$\sum_{\substack{i+j+k=p\\ \{i\neq p-1\}\cap\{j\neq 1\}}} \frac{p!}{i!j!k!}\|x\|^{2i}\big(2\langle x,y\rangle\big)^j\|y\|^{2k} \qquad \leq \qquad \sum_{\substack{i+j+k=p\\ \{i\neq p-1\}\cap\{j\neq 1\}}} \frac{p!}{i!j!k!}\|x\|^{2i}\big(2\|x\|\|y\|\big)^j\|y\|^{2k}. \tag{76}$$

Moreover,

$$\begin{aligned}\sum_{k=0}^{2p} \binom{2p}{k}\|x\|^{2p-k}\|y\|^k =& (\|x\|+\|y\|)^{2p} = (\|x\|^2+2\|x\|\|y\|+\|y\|^2)^p\\ =& \sum_{i+j+k=p} \frac{p!}{i!j!k!}\|x\|^{2i}\big(2\|x\|\|y\|\big)^j\|y\|^{2k}.\end{aligned}$$

Consequently,

$$\sum_{\substack{k=0\\ k\neq 1}}^{2p} \binom{2p}{k}\|x\|^{2p-k}\|y\|^k = \sum_{\substack{i+j+k=p\\ \{i\neq p-1\}\cap\{j\neq 1\}}} \frac{p!}{i!j!k!}\|x\|^{2i}\big(2\|x\|\|y\|\big)^j\|y\|^{2k}. \tag{77}$$

Thus, in view of (76) and (77), the desired result is obtained. □

References

[1] L. Arnold. *Stochastic differential equations: theory and applications.* Wiley & Sons, New York, 1974.

[2] M. Barkhagen, N. H. Chau, É. Moulines, M. Rásonyi, S. Sabanis and Y. Zhang. On stochastic gradient Langevin dynamics with stationary data streams in the logconcave case. *Preprint*, 2018. arXiv:1812.02709

[3] Bogachev. *Measure theory.* vol. 2. Springer, 2007.

[4] N. H. Chau, Ch. Kumar, M. Rásonyi and S. Sabanis. On fixed gain recursive estimators with discontinuity in the parameters. *ESAIM Probability and Statistics*, 23:217–244, 2019.

[5] X. Cheng and P. Bartlett. Convergence of Langevin MCMC in KL-divergence. *Preprint arXiv:1705.09048*, 2017.

[6] X. Cheng, N. S. Chatterji, Y. Abbasi-Yadkori, P. L. Bartlett and M. I. Jordan. Sharp convergence rated for Langevin dynamics in the nonconvex setting. *Preprint*, 2018. arXiv:1805.01648v1

[7] A. S. Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79:651–676, 2017.

[8] A. S. Dalalyan. Further and stronger analogy between sampling and optimization: Langevin Monte Carlo and gradient descent. *In Conference on Learning Theory*, pages 678–689, 2017.

[9] A. S. Dalalyan and A. Karagulyan. User-friendly guarantees for the Langevin Monte Carlo with inaccurate gradient. *To appear in Stochastic Processes and their Applications*, 2019.

[10] A. S. Dalalyan and A. B. Tsybakov. Sparse regression learning by aggregation and Langevin Monte Carlo. *Journal of Computer and Systems Science*, 78:1423–1443, 2012.

[11] C. Dellacherie and P.-A. Meyer. *Probabilités et potentiel. Chapitres I á IV.* Hermann, Paris, 1975.

[12] A. Durmus and É. Moulines. Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *Ann. Appl. Probab.*, 27:1551–1587, 2017.

[13] A. Durmus and É. Moulines. High-dimensional Bayesian inference via the unadjusted Langevin algorithm. *Preprint*, 2018. arXiv:1605.01559v3

[14] A. Eberle. Reflection couplings and contraction rates for diffusions. *Probab. Theory Related Fields*, 166:851–886, 2016.

[15] A. Eberle, A. Guillin and R. Zimmer. Quantitative Harris-type theorems for diffusions and McKean-Vlasov processes. *In Press, Transactions of the American Mathematical Society*, 2018. https://doi.org/10.1090/tran/7576

[16] L. Gerencsér. On a class of mixing processes. *Stochastics*, 26:165–191, 1989.

[17] N. V. Krylov. *Controlled diffusion processes.* Springer, 2008.

[18] E. Lenglart. Relation de domination entre deux processus. *Ann. Inst. H. Poincaré Prob.Statist.* 13:171–179, 1977.

[19] R. Liptser and A. N. Shiryaev. *Statistics of random processes.* Springer, 2013

[20] R. Douc, E. Moulines, P. Priouret and P. Soulier. *Markov chains.* Springer, 2018

[21] M. B. Majka, A. Mijatović and Ł. Szpruch. Non-asymptotic bounds for sampling algorithms without log-concavity. *Preprint*, 2018. arXiv:1808.07105v1

[22] X. Mao. *Stochastic Differential Equations and Their Applications.* Horwood Pub., 2007.

[23] M. Raginsky, A. Rakhlin, and M. Telgarsky. Non-convex learning via Stochastic Gradient Langevin Dynamics: a nonasymptotic analysis. *Proceedings of Machine Learning Research*, (65)1674–1703, 2017.

[24] D. Revuz and M. Yor. *Continuous martingales and Brownian motion.* Springer, 2005.

[25] C. Villani. *Optimal transport. Old an new.* Springer, 2009.

[26] M. Welling, and Y. W. Teh. Bayesian learning via stochastic gradient Langevin dynamics. *Proceedings of the 28th international conference on machine learning* (ICML-11). 2011.

[27] P. Xu, J. Chen, D. Zhou and Q. Gu. Global convergence of Langevin dynamics based algorithms for nonconvex optimization. *Advances in Neural Information Processing Systems*, 2018.