OPT2020: 12th Annual Workshop on Optimization for Machine Learning

One-Point Gradient Estimators for Zeroth-Order Stochastic Gradient Langevin Dydamics

Lewis Liu ALLIS.ALGO2@GMAIL.COM
MILA and DIRO, Universit de Montral
Zhaoran Wang ZHAORANWANG@GMAIL.COM
Northwestern University, United States

Abstract

In this paper, we consider a zeroth-order stochastic gradient langevin dynamics (ZSGLD), which seeks approximate global minimizer in the derivative-free optimization scenario under adequately regular non-convex assumption. As a variant of the popular stochastic gradient langevin dynamics (SGLD), our recursion shares the additional properly scaled isotropic Gaussian noise but adopts a biased estimate of the gradient at each time-step. Our appropriately proposed one-point gradient estimator takes the advantage of both efficient zeroth-order approximation and the potential to escape local minima when embedded into SGLD recursion. We provide a nonasymptotic analysis to guarantee a finite-time convergence of ZSGLD in 2-Wasserstein distance with a general setting. The feasibility of the constraints in our theory is also validated given practical settings.

1. Introduction

Derivative-free optimization (DFO), also known as zeroth-order derivative optimization, has sparked new interest in recent years as a large number of practical machine learning applications require the optimization of black-box functions of which derivatives are inaccessible or computationally expensive. In such costly optimization scenarios, the evaluation of objective function usually requires a complex deterministic simulation based on solving the equations (for example, nonlinear eigenvalue problems, ordinary or partial differential equations) modeled from the original problems, and the sheer noise introduced into the simulation procedure which calls for legacy or proprietary codes makes the model neither pellucid nor reliable [16].

A number of strategies has been studied for DFO in diverse settings. In the literature of simulation optimization, there exist a huge bunch of such algorithms for solving both simulation and optimization tasks, see for example [5] for optimization and [1] for approximate sampling. The similarities between the task of optimization and that of averaging have been recently exploited in the papers (Dalalyan, 2014; Durmus and Moulines, 2016; Durmus et al., 2016) aiming to establish fast and accurate theoretical guarantees for sampling from and averaging with respect to the density using the Langevin Monte Carlo algorithm.

In the context of bandit optimization, a line of works [27] estimates the gradient with respect to policy parameters via finite-difference-like methods for black-box policy search, which is designed for model-free reinforcement learning settings. [28] considers zero-order stochastic convex optimization with two function evaluations per round. Moreover, to overcome the scalability of DFO to high dimensionality, [25] presents a sequential random embeddings (SRE) to reduce the embedding gap while running optimization algorithms in the low-dimensional spaces. On the other hand, the accuracy of posterior estimation plays an essential role in a bunch of black-box optimization problems. In some cases, the mismatch in posteriors may not hurt in terms of decision making, and we will still end up with good decisions. Unfortunately, in other cases, this mismatch together with its induced feedback loop will degenerate in a significant loss of performance.

In another line of works, stochastic gradient Langevin dynamics (SGLD) provides a fairy approximation to the global minima under non-convex learning frameworks with diverse guarantees. The approach seems to simply add properly scaled isotropic Gaussian noise to an unbiased estimate of the gradient at each iteration but turns out to be effective to escape local minima and suffices to guarantee asymptotic global convergence for appropriately regular non-convex objectives [17].

To leverage the approximate global convergence property of SGLD in derivative-free scenarios, we propose zeroth-order stochastic gradient langevin dydamics (ZSGLD) to fill this void with theoretical convergence guarantees.

2. Related Work

There has been a line of works developing effective derivative-free algorithms for optimization, both local and global. In traditional local optimization scenario, pattern search methods [9, 22] evaluate the cost function in a stencil-based fashion determined by a set of directions with intrinsic properties meant to be desirable from a geometric/algebraic point of view. [13] introduced a novel class of derivative-free optimization methods by applying certain non-commutative maps to approximate the gradient of the objective function with convergence guarantees. From the view of global search, the genetic algorithm (GA) [23], the particle swarm algorithm [2], and the differential evolution algorithm (DE) [15, 29] are the other popular derivative free optimization tools which are ubiquitously set in complicated industrial applications. However, these zeroth-order methods tend to be trapped in some spurious local minima.

Another line of work improving the performance of global optimization and posterior estimation has been based on Stochastic Gradient Langevin Dynamics (SGLD) [17]. Stochastic average gradient Langevin dynamics (SAGA-LD) and stochastic variance-reduced gradient Langevin dynamics (SVRG-LD) was proposed in [12] to boost the convergence rate of SGLD, followed by SVRG-LD$^+$ from [30] for a lower gradient complexity. In the light of theoretical analysis for diverse Langevin dynamics approaches, previous works [8, 20] provided guarantees for the distributions of iterates to converge rapidly to Gibbs measure and for the stationary distribution to concentrate around the global minimizers under large enough inverse temperature parameter [4]. [10, 11] demonstrated nonasymptotic rates of convergence for discretized Langevin iteration on convex functions and for sampling from log-concave densities. Another projected Langevin scheme while still in the convex literature was then analyzed by [6]. Recently, the convergence rates of SVRG-LD and SAGA-LD to the target distribution in 2-Wasserstein distance is analyzed in [7]. It turns out that SAGA-LD has a lower gradient complexity than SVRG-LD. [26] provided finite-time guarantees to reach approximate global minima with SGLD for both empirical and population loss by nonasymptotic analysis.

3. Notation

We use $\mathcal{N}(\mu, \sigma^2)$ to denote Gaussian distribution with mean μ and variance σ^2. We use $X \sim \mu_x$ to show that the X is subjected to distribution μ_x, and $X \sim Y$ to denote that two random variables X and Y have the same probability law. Let $\mathbb{E}_X(\cdot)$ be the expectation with respect to random variable X. If not specified, $\mathbb{E}(\cdot)$ takes expectation over all possible random variables. Other notations will be introduced when they are used.

4. Models and Algorithms

In this section, we present the details of construction of the gradient estimator for derivative-free optimization problems and our zeroth-order Langevin Monte Carlo Algorithm.

4.1. Gradient Estimator for Stochastic Derivative-free Optimization

We study stochastic derivative-free optimization in the context of limited computational budget and prohibitive access to the objective (loss) function. We define $L : \mathcal{X} \to \mathbb{R}$ to be a black-box function over domain $\mathcal{X} \subset \mathbb{R}^d$, and possibly non-differentiable or noisy. In machine learning scenario, L can always be formalized as

$$L(x) := \mathbb{E}_P[l(x, \widetilde{\xi})] = \int_\Omega l(x, \xi)\mathbb{P}(d\xi), \tag{4.1}$$

where $\widetilde{\xi}$ is a random variable on probability space Ω, from which i.i.d. training samples $(\widetilde{\xi}_1, \widetilde{\xi}_2, ..., \widetilde{\xi}_n)$ are drawn. In our case only the *stochastic zeroth-order oracle* is available:

$$y_m = L(x_m) + \epsilon, \qquad \epsilon \sim \mathcal{N}(0, \sigma^2) \tag{4.2}$$

such that evaluation of L is noisy although the expected value is correct and the gradient needs to be properly approximate. After designed m queries of the stochastic zeroth-order oracle in Eq. (4.2) by a sequence $\{X_0, X_1, ..., X_{m-1}\}$ of random variables , we focus on the unconstrained ($\mathcal{X} = \mathbb{R}^d$) derivative-free optimization problem

$$\min\{L(x) : x \in \mathcal{X}\} \tag{4.3}$$

through a minimization of the 2-Wasserstein distance between measures of X_m and the minimizer x^* of Eq. (4.3),

$$\mathcal{W}(\mathcal{L}(X_m), \mathcal{L}(x^*)) = (\inf \mathbb{E}\|X_m - x^*\|^2)^{1/2}, \tag{4.4}$$

which leads to a minimization of excess risk under mild assumptions [26]. Next, to give well-defined properties of our estimator and the convergence rate of our ZSGLD recursion, we impose assumptions usually used in Langevin dynamics literature as follows (the details will be discussed in Section 6.1).

Assumption 4.1 *The loss function L takes nonnegative real values, and the value at zero is bounded, that is, there exist constants $R, S \geq 0$, so that*

$$|L(0)| \leq R, \qquad \|\nabla L(0)\| \leq S. \tag{4.5}$$

Assumption 4.2 *The initial iterate X_0 is configured with a probability law with a strictly positive and bounded density function f_0 w.r.t. Lebesgue measure on $\mathbb{R}^d$, so that*

$$t_0 = \int_{\mathbb{R}^d} e^{\|x\|^2} f_0(x)dx \tag{4.6}$$

is finite.

Assumption 4.3 *The loss function L satisfies (a, b)-dissipativeness [19], that is, there exist $a > 0$ and $b \geq 0$, such that for all $x \in \mathbb{R}^d$,*

$$\langle x, L(x)\rangle \geq a\|x\|^2 - b. \tag{4.7}$$

Assumption 4.4 *The gradient function of L, $\nabla L(\cdot)$, is T-Lipschitzian, namely, for a constant $T > 0$,*

$$\|\nabla L(x) - \nabla L(y)\| \leq T\|x - y\|, \qquad \forall x, y \in \mathbb{R}^d. \tag{4.8}$$

Similar to [14], below we introduce single point gradient estimates through sampling via a *probing distribution* μ. Let μ be a constant distribution supported on $\mathbb{R}^d$ and the random variable $u \sim \mu$, of which the probability-density function is denoted as p. Before giving the gradient estimator for ∇L, we first construct a *smoothed* version of function L,

$$\widetilde{L} := \mathbb{E}_{u\sim\mu}[L(x + u)], \tag{4.9}$$

which suppress the magnitude $|\widetilde{L}(x) - L(x)|$ and enables $\nabla\widetilde{L}(x)$ to be unbiasedly estimated. Then for any differentiable function $f : \mathbb{R}^d \to \mathbb{R}$, *Stein-Hudson identity* gives

$$\mathbb{E}[f(u)] = -\mathbb{E}[f(u)\nabla \log p(u)], \quad u \sim \mu. \tag{4.10}$$

By setting $f(u) = L(x + u)$ in Eq. (4.10) and combining Eq. (4.9), we can obtain

$$\nabla\widetilde{L}(x) = -\mathbb{E}_{u\sim\mu}[L(x + u)\nabla \log p(u)] \tag{4.11}$$

Hence we can intuitively pose an unbiased estimator $\widehat{h} : \mathbb{R} \times \Omega^d \to \mathbb{R}^d$ for $\nabla\widetilde{L}(x)$ with a control variate term reducing variance,

$$\widehat{h}(x, u) = (y - y')\nabla \log p(u), \tag{4.12}$$

where $y = L(x) + \epsilon$, $y' = L(x + u) + \epsilon'$, $u \sim \mu$. The control variate like y together with y' decorrelates dependency of the estimator on the magnitude of $|L(x)|$, which also helps give a finite-time convergence guarantee for ZSGLD in Section 4.2. The fitness of the estimator is concluded in the following result.

Proof From the independence of ϵ and u, $\mathbb{E}\widehat{h}(x, u) = \mathbb{E}[(L(x)+\epsilon)\nabla \log p(u)] - \mathbb{E}[(L(x+u)+\epsilon')\nabla \log p(u)] = L(x)\mathbb{E}[\nabla \log p(u)] - \mathbb{E}[L(x + u)\nabla \log p(u)]$. Under the regularity assumptions above, swapping of integration and differentiation is allowed. Thus we have $\mathbb{E}[\nabla \log p(u)] = \int p(u)\frac{\nabla p(u)}{p(u)}du = \int \nabla p(u)du = \nabla \int p(u)du = 0$. Combining with the previous equation, the proposition holds. ■

However, the estimator is biased to $\nabla L(x)$, while our ZSGLD recursion still gives reasonable finite-time convergence rate. The effect of the estimator on the recursion and related results are illustrated in the following sections.

4.2. Zeroth-order Stochastic Gradient Langevin Dynamics

We focus on the derivation and analysis of ZSGLD recursion in this section. In case of difficulties in accessing the gradient of objective function, we use the Stein-Hudson gradient estimator $\widehat{h}$ instead to obtain our main recursion:

$$X_{m+1} = X_m - \gamma\widehat{h}(X_m, u_m) + \sqrt{\frac{2\gamma}{\rho}}\zeta_m, \quad m \in \mathbb{N}. \tag{4.13}$$

where $\{\zeta_m\}_{m=0}^{\infty}$ are mutually independent standard Gaussian random vectors in $\mathbb{R}^d$, $\{u_m\}_{m=0}^{\infty}$ are mutually independent random vectors in $\mathbb{R}^d$ drawn from *probing distribution* μ, and for a fixed m, $X_m : \Xi \to \mathbb{R}^d$ is a random variable on space Ξ. We denote by $\rho > 0$ the inverse temperature parameter, and by $\gamma > 0$

the step length. We assume that X_0, u, $\{\zeta_m\}_{m=0}^{\infty}$ are mutually independent random vectors in $\mathbb{R}^d$. In this formulation, queries to the stochastic zeroth-order oracle can be obtained by a sequence of random iterates $\{X_0, X_1, ..., X_m, ...\}$, where X_m only depends on $\{X_0, ..., X_{m-1}\}$, $\{y_0, ..., y_{m-1}\}$ and randomness introduced by *probing distribution*. From a perspective of stochastic differential equation (SDE) [4], the above recursion can be regarded as a (zeroth-order) Euler-Maruyama [21] discretized version of Langevin diffusion with a gradient estimator,

$$dX(t) = -\nabla L(X(t))dt + \sqrt{\frac{2}{\rho}}dB(t), \quad t \geq 0, \tag{4.14}$$

where $X(t)$ is a stochastic process on $[0, \infty) \times \Xi$, $B(t)$ denotes the standard Brownian motion in $\mathbb{R}^d$. The $-\nabla L(x)$ term is usually referred as the *drift coefficient*. Given regularity assumptions in Section 4.1, according to detailed study in [8], (4.14) will converge rapidly to the unique invariant distribution $\omega \propto \exp(-\rho L)$. The upshot of the connection between (A.2) and (A.3) is that we can first prove the ZSGLD tracks the Langevin diffusion (4.14) with respect to distribution in 2-Wasserstein sense, then by using the fact that distributions of $X(t)$ converges to Gibbs measure and ω with large ρ concentrates around the global minimizer of L [20], along with the triangle inequality for 2-Wasserstein distance, we are able to give convergence guarantee for ZSGLD.

5. Main Theory

In this section, we will present extensive theoretical results to demonstrate the convergence rate of ZSGLD in 2-Wasserstein distance. Firstly, based on regularity assumptions and properties of the Stein-Hudson gradient estimator, we show bounds on the possible gradient norm of L, second-order moment error and first-order moment error brought by $\widehat{h}$. To quantify the closeness between X_m and $X(t)$, we then upsample the discrete-time stochastic process X_m with step size γ by interpolation. As a result, the second-order moment error from the value of a inner point to the end point is given, which also contains an upper bound on the second-order moment about the origin of X_m. Finally we present the overall bound on 2-Wasserstein distance between iterate distribution of (A.2) and Gibbs measure.

Let $\mathbb{P}_m := \mathcal{L}(X_k), \mathbb{Q}_t := \mathcal{L}(X(t))$. According to the bounded initial value and gradient of L, a simple but useful linear bound is given as follows.

Lemma 5.1 *Under Assumption A.1 and A.4, we have a linear bound on* $\|\nabla L\|$*,*

$$\|\nabla L(x)\| \leq T\|x\| + S, \qquad \forall x \in \mathbb{R}^d. \tag{5.1}$$

Remark 5.2 *Lemma B.1 is essentially a natural result from the T-Lipschitzian assumption, also a consequence of bounded second-order derivative if L is second-order differentiable. We pose no finite bound of either $|L(x)|$ or $\nabla L(x)$ while (B.4) controls the growth of the gradient norm within a linear speed.*

The following result gives a quadratic bound on the expected squared error of $\widehat{h}$.

Lemma 5.3 *For each $x \in \mathbb{R}^d$,*

$$\mathbb{E}\|\widehat{h}(x, u) - \nabla L(x)\|^2 \leq H_1\|x\|^2 + H_2\|x\| + H_3, \tag{5.2}$$

where

$$H_1 = 4T^2\mathbb{E}\|u\|^2\|\nabla \log p(u)\|^2 + 2T^2, \tag{5.3}$$

$$H_2 = 4TS(\mathbb{E}\|u\|^2\|\nabla \log p(u)\|^2 + 1), \tag{5.4}$$

$$H_3 = 2S^2\mathbb{E}\|u\|^2\|\nabla \log p(u)\|^2 + 4TS\mathbb{E}\|u\|^3\|\nabla \log p(u)\|^2 + 4T^2\mathbb{E}\|u\|^4\|\nabla \log p(u)\|^2 + 4\sigma^2\mathbb{E}\|\nabla \log p(u)\|^2 + 2S^2. \tag{5.5}$$

Next the bound on the first-order moment error is illustrated.

Lemma 5.4 *For each $x \in \mathbb{R}^d$,*

$$\mathbb{E}\|\widehat{h}(x, u) - \nabla L_(x)\| \le H_4\|x\| + H_5, \tag{5.6}$$

where

$$H_4 = T(\mathbb{E}\|u\|\|\nabla \log p(u)\| + 1), \tag{5.7}$$

$$H_5 = T\mathbb{E}\|u\|^2\|\nabla \log p(u)\| + S\mathbb{E}\|u\|\|\nabla \log p(u)\| + \frac{2\sigma}{\sqrt{\pi}}\mathbb{E}\|\nabla \log p(u)\| + S. \tag{5.8}$$

Then, we present a crucial lemma that describe the accumulative bias introduced by zeroth-order gradient estimator in ZSGLD recursion in the form of an interpolation error.

Lemma 5.5 (Second-order Moment Error by Interpolation) $\forall t \ge 0$, *let*

$$\widetilde{X}(t) = X_0 - \int_0^t \widehat{h}(\widetilde{X}(\lfloor q/\gamma \rfloor \gamma), \widetilde{u}(q))dq + \int_0^t dB(q), \tag{5.9}$$

be a continuous version of the discretized recursion using straightforward interpolation, where $\widetilde{u}(q) = u_m$ when $q \in [m\gamma, (m+1)\gamma)$. Then, for all $m \in \mathbb{N}, m \ge 0$ and some $q \in [m\gamma, (m+1)\gamma)$, when $\frac{a}{T} > \mathbb{E}\|u\|\|\nabla \log p(u)\| + 1$, there exists constants $0 < \delta < \frac{2(a - H_4)}{H_5}$ and $\alpha > 0$, so that for any $0 < \gamma < 1 \wedge \frac{2(a-H_4)-\delta H_5}{2T^2+H_1+2MH_4+(H_2/2+SH_4+TH_5)\delta}$, we have

$$\mathbb{E}\|\widetilde{X}(q) - \widetilde{X}(m\gamma)\|^2 \le 3\gamma^2\bigg(\big((H_1 + R^2) + H_{10}\alpha\big) \cdot (K_1(\delta) \vee K_2(\delta)) + \frac{H_{10}}{\alpha} + H_{11} \bigg) + \frac{6\gamma d}{\rho}, \tag{5.10}$$

where we define

$$
\begin{aligned}
H_6 &= H_1 + 2H_4T + 2T^2, \\
H_7 &= \frac{H_2}{2} + SH_4 + TH_5, \\
H_8 &= \frac{2(TH_5 + SH_4) + H_2}{2}, \\
H_9 &= H_3 + 2SH_5 + 2S^2, \\
H_{10} &= \frac{H_2 + 2RS}{2}, \quad H_{11} = H_3 + S^2, \\
K_1(\delta) &= \frac{2(TH_5 + H_5 + SH_4) + H_2}{2\delta} + H_3 \\
&\quad + 2SH_5 + 2S^2 + 2b + \frac{2d}{\rho}, \\
K_2(\delta) &= \tau_0 + \frac{\frac{H_8 + H_5}{\delta} + H_9 + 2b + \frac{2d}{\rho}}{2a - 2H_4 - H_6 - (H_5 + H_7)\delta}.
\end{aligned} \tag{5.11}
$$

The following main theorem shows that even if we pose a biased estimation, we are still able to reach reasonable rate of convergence.

Theorem 5.6 *When it holds that*

$$
\frac{a}{T} > \mathbb{E}\|u\|\|\nabla \log p(u)\| + 1, \tag{5.12}
$$

there exists constants $\delta < \frac{2(a-H_4)}{H_5}, \alpha > 0, \beta > 0$, *so that for* $m \in \mathbb{N}$ *and any* $0 < \gamma < \frac{2(a-H_4)-\delta H_5}{2T^2+H_1+2MH_4+(H_2/2+SH_4+TH_5)\delta}$, *we have*

$$
\begin{aligned}
\mathcal{W}(\mathbb{P}_m, \omega) &\le (\sqrt{\widetilde{C}_1(\alpha,\delta)}\gamma^{\frac{1}{4}} + \sqrt{\widetilde{C}_2(\beta,\delta)})m\gamma \\
&\quad + C_3 e^{-\frac{m\gamma}{\rho C_{LS}}},
\end{aligned} \tag{5.13}
$$

where

$$
\begin{aligned}
C_1(\alpha,\delta) &= \frac{3\rho T^2}{2}\Big(\big((H_1 + R^2) + H_{10}\alpha\big)\left(K_1(\delta) \vee K_2(\delta)\right) \\
&\quad + \frac{H_{10}}{\alpha} + H_{11} + \frac{2d}{\rho}\Big), \\
C_2(\beta,\delta) &= \frac{\rho}{2}\left(\left(H_1 + \frac{H_2\beta}{2}\right)(K_1(\delta) \vee K_2(\delta) + \frac{H_2}{2\beta} + H_3\right), \\
\widetilde{C}_1(\alpha,\delta) &= \left(12 + 8\left(\tau_0 + 2b + \frac{2d}{\rho}\right)m\gamma\right)\cdot(\sqrt{C_1(\alpha,\delta)} + C_1(\alpha,\delta)), \\
\widetilde{C}_2(\beta,\delta) &= \left(12 + 8\left(\tau_0 + 2b + \frac{2d}{\rho}\right)m\gamma\right)\cdot(\sqrt{C_2(\beta,\delta)} + C_1(\beta,\delta)), \\
C_3 &= \sqrt{2C_{LS}\left(\rho C_{30} + \log\|f_0\|_\infty + \frac{d}{2}\log\frac{3\pi}{a\rho}\right)}, \\
C_{30} &= \frac{T\tau_0}{3} + S\sqrt{\tau_0} + R + \frac{b}{2}\log 3.
\end{aligned} \tag{5.14}
$$

See Section B for a detailed proof of Lemma B.1, B.2, B.3, B.4 and Theorem 5.6.

6. Discussions and Examples

In this section, we will first give extensive illustration to the feasibility of the regularity assumptions proposed in Section 4.1. To show how our theory works more intuitively, we take a specific *probing distribution* μ as an example that enable our algorithm to work while satisfying all the aforementioned conditions.

6.1. Feasibility of Assumptions

For general non-convex optimization study, assumption A.1 and A.4 are trivial to guarantee a further analysis. In the following we focus on the exponential integrability assumption (A.2) and dissipative assumption (A.3). On one hand, when X_0 is randomly initialized with the natural Gaussian distribution $X_0 \sim \mathcal{N}(0, \sigma^2 I_d)$ of which $\sigma^2 < 1/2$, Eq. (A.5) gives $t_0 = \frac{1}{\sqrt{2\pi}\sigma}\int_{\mathbb{R}^d}\exp\{\frac{(2\sigma^2-1)\|x\|^2}{2\sigma^2}\}dx < \frac{1}{\sqrt{2\pi}\sigma}\int_{\mathbb{R}^d}\exp\{\frac{-\epsilon_\sigma\|x\|^2}{2\sigma^2}\}dx < \infty$ for some constant $\epsilon_\sigma > 0$. Hence the former assumption is easy to be satisfied. On the other hand, when we construct the function L with 2-norm regularization as $L(x) = L_0(x) + \dfrac{\theta}{2}\|x\|^2$, where L_0 is T_0-Lipschizian. As a result, (A.6) holds for L with $a = \dfrac{\theta}{2}$ and $b = \dfrac{T_0^2}{2\theta}$. Thus, our analysis applies to a wide range of loss functions with 2-norm regularization of weights. From a higher point of view, dissipativity implies that critical points (where $\langle x, L(x)\rangle = 0$) of a dynamical system described by $\dot{x}(t) = -\nabla L(x)$ ($x(0) = x_0$) will fall into an *absorbing set* which turns out to be a ball centered at the origin with radius $\sqrt{\dfrac{b}{a}}$. However, their distribution can be arbitrary. From all above, the assumption A.3 only poses little restriction on the objective function.

Another important difference between our assumptions and those of non-convex global optimization is the bounded variance of inexact gradient. In our derivative-free scenario, the gradient in SGLD is replaced with the proposed gradient estimator of which the variance bound and bias bound are proved as our theoretical results.

6.2. Examples of Probing Distribution

Below we will take Gaussian distribution as an example, instead of the abstract distribution μ, for the *probing distribution* which plays a crucial role in our one-point gradient estimation, to show how this usual distribution meet our requirements and enable the theory to work.

For $u \in \mathbb{R}^d$, we consider the Gaussian distribution

$$p(u) = p(u^{(1)}, ..., u^{(d)}) = \prod_{i=1}^{d} p_i(u^{(i)}),$$
$$p_i(u^{(i)}) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{{u^{(i)}}^2}{2\sigma^2}\right\}, \tag{6.1}$$

where $\sigma > 0$ denotes the variance of $u^{(i)}$ (i=1,2,...,d). In this configuration, we note that $\|\nabla \log p(u)\| = \| - u/\sigma^2\| = \|u\|/\sigma^2$. Our gradient estimator gives $\widehat{h}(x, u) = (y' - y)u/\sigma^2$. Next we check that the expectation terms with respect to u in H_i ($i = 1, 2, ..., 10$) are bounded. Firstly, from standard Gaussian tail inequalities $\mathbb{P}\{|u| \geq s\} \leq 2\exp\{-\frac{s^2}{2\sigma^2}\}, \forall i \in \{1,2,...,d\}, \forall t \geq 0$, and furthermore a union bound of the inequalities over $\{1,2,...,d\}$, *i.e.*, $\mathbb{P}\{\max_i |u^{(i)}| \geq s\} \leq 2d\exp\{-\frac{s^2}{2\sigma^2}\}$, we have for all even $n \in \mathbb{N}$,

$$\begin{aligned}\mathbb{E}\|u\|^n &\leq \sigma^n(2d\log d)^{n/2}\mathbb{P}\{\max_i |u^{(i)}| \leq \sigma\sqrt{2\log d}\} \\ &\quad + \int_{\sigma\sqrt{2\log d}}^{\infty} s^n d^{n/2}\mathbb{P}\{\max_i |u^{(i)}| \geq s\}ds \\ &\leq \sigma^n(2d\log d)^{n/2} + 2d^{\frac{n}{2}+1}\int_{\sigma\sqrt{2\log d}}^{\infty} s^n e^{-s^2/2\sigma^2}ds \\ &\leq \sigma^n(2d\log d)^{n/2} + 2d^{\frac{n}{2}}\sigma^{n+1}n(2\log d)^{n/2} \\ &\lesssim (1+\sigma)\sigma^n(d\log d)^{n/2}. \end{aligned} \tag{6.2}$$

Hence the terms $\mathbb{E}\|u\|\|\nabla\log p(u)\|$, $\mathbb{E}\|\nabla\log p(u)\|^2$, $\mathbb{E}\|u\|^2\|\nabla\log p(u)\|^2$, and $\mathbb{E}\|u\|^4\|\nabla\log p(u)\|^2$ are proved to be bounded. Furthermore, by Jensen's inequality for $f(x) = -\sqrt{x}$, we have for all odd $n' \in \mathbb{N}$,

$$\mathbb{E}\|u\|^{n'} \leq \sqrt{\mathbb{E}\|u\|^{2n'}}, \tag{6.3}$$

where the right-hand side have been upper bounded by Eq. (6.2). Therefore, terms like $\mathbb{E}\|\nabla\log p(u)\|$, $\mathbb{E}\|u\|^2\|\nabla\log p(u)\|$, and $\mathbb{E}\|u\|^3\|\nabla\log p(u)\|^2$ are all bounded. To guarantee a finite-time convergence of the example, we need one more step to check condition (5.12). It follows that

$$\mathbb{E}\|u\|\|\nabla\log p(u)\| = \frac{1}{\sigma^2}\mathbb{E}\|u\|^2 = d. \tag{6.4}$$

Hence it requires $a > T(d+1)$ to hold for our algorithm, it is reasonable since a and T are decorrelated with the dimension d. Thus the example of Gaussian probing distribution is validated.

7. Conclusion

In this paper, we propose a simple yet provably efficient one-point gradient estimator which takes the advantage of both zeroth-order approximation and the potential to escape local minima of SGLD. For such a zeroth-order SGLD, we provide a nonasymptotic analysis to guarantee a finite-time convergence in 2-Wasserstein distance with a general setting. In addition, our work provides more insight into the combination of derivative-free optimization and global optimization in simulation-based real-world systems.

References

[1] Atchadé, Y., Fort, G., Moulines, E., and Priouret, P. Adaptive markov chain monte carlo: theory and methods. *Preprint*, 2009.

[2] Biswal, B., Dash, P. K., and Panigrahi, B. K. Power quality disturbance classification using fuzzy c-means algorithm and adaptive particle swarm optimization. *IEEE Transactions on Industrial Electronics*, 56(1):212–220, 2009.

[3] Bolley, F. and Villani, C. Weighted csiszár-kullback-pinsker inequalities and applications to transportation inequalities. In *ANNALES-FACULTE DES SCIENCES TOULOUSE MATHEMATIQUES*, volume 14, pp. 331. Université Paul Sabatier, 2005.

[4] Borkar, V. S. and Mitter, S. K. A strong approximation theorem for stochastic recursive algorithms. *Journal of optimization theory and applications*, 100(3):499–513, 1999.

[5] Boyd, S. and Vandenberghe, L. *Convex optimization*. Cambridge university press, 2004.

[6] Bubeck, S., Eldan, R., and Lehec, J. Sampling from a log-concave distribution with projected langevin monte carlo. *Discrete & Computational Geometry*, 59(4):757–783, 2018.

[7] Chatterji, N. S., Flammarion, N., Ma, Y.-A., Bartlett, P. L., and Jordan, M. I. On the theory of variance reduction for stochastic gradient monte carlo. *arXiv preprint arXiv:1802.05431*, 2018.

[8] Chiang, T.-S., Hwang, C.-R., and Sheu, S. J. Diffusion for global optimization in r^n. *SIAM Journal on Control and Optimization*, 25(3):737–753, 1987.

[9] Conn, A. R., Scheinberg, K., and Vicente, L. N. *Introduction to derivative-free optimization*, volume 8. Siam, 2009.

[10] Dalalyan, A. S. Theoretical guarantees for approximate sampling from smooth and log-concave densities. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3):651–676, 2017.

[11] Dalalyan, A. S. and Tsybakov, A. B. Sparse regression learning by aggregation and langevin monte-carlo. *Journal of Computer and System Sciences*, 78(5):1423–1443, 2012.

[12] Dubey, K. A., Reddi, S. J., Williamson, S. A., Poczos, B., Smola, A. J., and Xing, E. P. Variance reduction in stochastic gradient langevin dynamics. In *Advances in neural information processing systems*, pp. 1154–1162, 2016.

[13] Feiling, J., Zeller, A., and Ebenbauer, C. Derivative-free optimization algorithms based on non-commutative maps. *arXiv preprint arXiv:1805.07248*, 2018.

[14] Flaxman, A. D., Kalai, A. T., and McMahan, H. B. Online convex optimization in the bandit setting: gradient descent without a gradient. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pp. 385–394. Society for Industrial and Applied Mathematics, 2005.

[15] Fleetwood, K. An introduction to differential evolution. In *Proceedings of Mathematics and Statistics of Complex Systems (MASCOS) One Day Symposium, 26th November, Brisbane, Australia*, pp. 785–791, 2004.

[16] Fowler, K. R., Reese, J. P., Kees, C. E., Dennis Jr, J., Kelley, C. T., Miller, C. T., Audet, C., Booker, A. J., Couture, G., Darwin, R. W., et al. Comparison of derivative-free optimization methods for groundwater supply and hydraulic capture community problems. *Advances in Water Resources*, 31(5):743–757, 2008.

[17] Gelfand, S. B. and Mitter, S. K. Recursive stochastic algorithms for global optimization in rˆd. *SIAM Journal on Control and Optimization*, 29(5):999–1018, 1991.

[18] Gyöngy, I. Mimicking the one-dimensional marginal distributions of processes having an itô differential. *Probability theory and related fields*, 71(4):501–516, 1986.

[19] Hale, J. K. *Asymptotic behavior of dissipative systems*. Number 25. American Mathematical Soc., 2010.

[20] Hwang, C.-R. Laplace's method revisited: weak convergence of probability measures. *The Annals of Probability*, pp. 1177–1182, 1980.

[21] Kloeden, P. E. and Platen, E. Higher-order implicit strong numerical schemes for stochastic differential equations. *Journal of statistical physics*, 66(1-2):283–314, 1992.

[22] Kolda, T. G., Lewis, R. M., and Torczon, V. Optimization by direct search: New perspectives on some classical and modern methods. *SIAM review*, 45(3):385–482, 2003.

[23] Lenwari, W., Sumner, M., and Zanchetta, P. The use of genetic algorithms for the design of resonant compensators for active filters. *IEEE Transactions on Industrial Electronics*, 56(8):2852–2861, 2009.

[24] Norros, I., Valkeila, E., Virtamo, J., et al. An elementary approach to a girsanov formula and other analytical results on fractional brownian motions. *Bernoulli*, 5(4):571–587, 1999.

[25] Qian, H., Hu, Y.-Q., and Yu, Y. Derivative-free optimization of high-dimensional non-convex functions by sequential random embeddings. In *IJCAI*, pp. 1946–1952, 2016.

[26] Raginsky, M., Rakhlin, A., and Telgarsky, M. Non-convex learning via stochastic gradient langevin dynamics: a nonasymptotic analysis. *arXiv preprint arXiv:1702.03849*, 2017.

[27] Sehnke, F., Osendorfer, C., Rückstieß, T., Graves, A., Peters, J., and Schmidhuber, J. Parameter-exploring policy gradients. *Neural Networks*, 23(4):551–559, 2010.

[28] Shamir, O. An optimal algorithm for bandit and zero-order convex optimization with two-point feedback. *Journal of Machine Learning Research*, 18(52):1–11, 2017.

[29] Storn, R. and Price, K. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11(4):341–359, 1997.

[30] Zou, D., Xu, P., and Gu, Q. Subsampled stochastic variance-reduced gradient langevin dynamics. In *International Conference on Uncertainty in Artificial Intelligence*, 2018.

Appendix A. Proof Sketch

We lay out a proof sketch of the main theory. Our proof utilizes several technical tools including the theory of Markov diffusion operators and associated functional inequalities, properties of stochastic differential equations, and transportation cost properties. Given an estimator $\widehat{h} : \mathbb{R} \times \Omega^d \to \mathbb{R}^d$ of the $\nabla L(x)$:

$$\widehat{h}(w, u) = (y - y')\nabla \log p(u), \tag{A.1}$$

where $y = L_\xi(w) + \epsilon$, $y' = L_\xi(w + u) + \epsilon'$, $u \sim \mu$. The zeroth-order SGLD recursion can be described as follows:

$$X_{m+1} = X_m - \gamma \widehat{h}(X_m, u_m) + \sqrt{\frac{2\gamma}{\rho}}\zeta_m, \qquad m = 0, 1, 2, ... \tag{A.2}$$

where $\{\zeta_m\}_{m=0}^{\infty}$ are mutually independent standard Gaussian random vectors in $\mathbb{R}^d$ and $\{u_m\}_{m=0}^{\infty}$ are mutually independent random vectors in $\mathbb{R}^d$ drawn from distribution μ. We denote by $\rho > 0$ the inverse temperature parameter. We assume that X_0, u, $\{\zeta_m\}_{m=0}^{\infty}$ are mutually independent standard Gaussian random vectors in $\mathbb{R}^d$. From a perspective of stochastic differential equation [4], the above recursion can be regarded as a (zeroth-order) discretized version of Langevin diffusion,

$$dX(t) = -\nabla L(X(t))dt + \sqrt{\frac{2}{\rho}}dB(t). \tag{A.3}$$

$X(t)$ is a stochastic process on $[0, \infty) \times \Xi$.

To guarantee the convergence of our discretized recursion, we impose assumptions usually used in Langevin dynamics literature as follows.

Assumption A.1 *The loss function L takes nonnegative real values, and the value at zero is bounded, that is, there exist constants $R, S \geq 0$, so that*

$$|L(0)| \leq R, \qquad \|\nabla L(0)\| \leq S. \tag{A.4}$$

Assumption A.2 *The initial iterate X_0 is configured with a probability law with a strictly positive and bounded density function f_0 w.r.t. Lebesgue measure on $\mathbb{R}^d$, so that*

$$\tau_0 = \int_{\mathbb{R}^d} e^{\|x\|^2} f_0(x)dx \tag{A.5}$$

is finite.

Assumption A.3 *The loss function L satisfies (a, b)-dissipativeness [19], that is, there exist $a > 0$ and $b \geq 0$, such that for all $x \in \mathbb{R}^d$,*

$$\langle x, L(x)\rangle \geq a\|x\|^2 - b. \tag{A.6}$$

Assumption A.4 *The gradient function of L, $\nabla L(\cdot)$, is T-Lipschitzian, namely, for a constant T¿0,*

$$\|\nabla L(x) - \nabla L(y)\| \leq T\|x - y\|, \qquad \forall x, y \in \mathbb{R}^d. \tag{A.7}$$

Appendix B. Proof of Main Results

B.1. Notation

Consider the stochastic optimization problem

$$\text{minimize} \qquad L(x) := \mathbb{E}_P[l(x,\xi)] = \int_\Omega l(x,\xi)\mathbb{P}(d\xi). \tag{B.1}$$

As it is prohibitive to get access to $\mathbb{P}$, we choose to minimize

$$L_\xi(x) := \frac{1}{n}\sum_{i=1}^{n} l(x,\xi_i), \tag{B.2}$$

the empirical risk of a hypothesis $w \in \mathbb{R}^d$ on a dataset $\xi = (\xi_1, ..., \xi_n) \in \Omega^n$ In our case only the *stochastic zeroth-order oracle* is available:

$$y = L(x) + \epsilon, \qquad \epsilon \sim \mathcal{N}(0,\sigma^2). \tag{B.3}$$

We use $X \sim Y$ to denote that two random variables X and Y have the same probability law. Let $\mathbb{E}_X(\cdot)$ be the expectation with respect to random variable X. If not specified, $\mathbb{E}(\cdot)$ takes expectation over all possible random variables. Let $\mathbb{P}_m := \mathcal{L}(X_k), \mathbb{Q}_t := \mathcal{L}(X(t))$, and ω be the Gibbs distribution measure, that is, the unique invariant distribution to which distributions of $X(t)$ converge to.

The following results shows that even if we pose a biased estimation $\widehat{g}(w,u) = (y - y')\nabla \log p(u)$ (where $y = L_\xi(w) + \epsilon$, $y' = L_\xi(w+u) + \epsilon'$, $u \sim \mu$) of $\nabla L_\xi(w)$, we are still able to reach reasonable rate of convergence.

Lemma B.1 *Under Assumption A.1 and A.4, we have a linear bound on $\|\nabla L\|$,*

$$\|\nabla L(x)\| \le T\|x\| + S, \qquad \forall x \in \mathbb{R}^d. \tag{B.4}$$

Proof For any $x \in \mathbb{R}^d$, from Assumption A.4, we have

$$\|\nabla L(x) - \nabla L(0)\| \le T\|x\|. \tag{B.5}$$

Notice that $\|\nabla L(x)\| - \|\nabla L(0)\| \le \|\nabla L(x) - \nabla L(0)\|$. Therefore, $\|L(x)\| \le T\|x\| + \|\nabla L(0)\| \le T\|x\| + S$. ■

Lemma B.2 *For each $x \in \mathbb{R}^d$,*

$$\mathbb{E}\|\widehat{h}(x,u) - \nabla L(x)\|^2 \le H_1\|x\|^2 + H_2\|x\| + H_3, \tag{B.6}$$

where

$$H_1 = 4T^2\mathbb{E}\|u\|^2\|\nabla \log p(u)\|^2 + 2T^2, \tag{B.7}$$

$$H_2 = 4TS(\mathbb{E}\|u\|^2\|\nabla \log p(u)\|^2 + 1), \tag{B.8}$$

$$\begin{aligned} H_3 = {} & 2S^2\mathbb{E}\|u\|^2\|\nabla \log p(u)\|^2 + 4TS\mathbb{E}\|u\|^3\|\nabla \log p(u)\|^2 \\ & + 4T^2\mathbb{E}\|u\|^4\|\nabla \log p(u)\|^2 + 4\sigma^2\mathbb{E}\|\nabla \log p(u)\|^2 + 2S^2, \end{aligned} \tag{B.9}$$

Proof

$$\begin{aligned}
\mathbb{E}\|\widehat{h}(x,u) - \nabla L(x)\|^2 &= \mathbb{E}\|(L(x) + \epsilon - L_(x+u) - \epsilon')\nabla \log p(u) - \nabla L(x)\|^2 \\
&\leq 2\mathbb{E}\|(L(x) - L(x+u) + \epsilon - \epsilon')\nabla \log p(u)\|^2 + 2\mathbb{E}\|\nabla L(x)\|^2 \\
&= 2\mathbb{E}(|L(x) - L(x+u)|^2 + 2\sigma^2)\|\nabla \log p(u)\|^2 + 2\mathbb{E}\|\nabla L(x)\|^2 \\
&\leq 2\mathbb{E}(\|\nabla L(x+\delta(u)u)\|^2\|u\|^2 + 2\sigma^2)\|\nabla \log p(u)\|^2 + 2(T\|x\| + S)^2 \\
&\leq 2\mathbb{E}((T\|x+\delta(u)u\| + S)^2\|u\|^2 + 2\sigma^2)\|\nabla \log p(u)\|^2 + 2(T\|x\| + S)^2 \\
&\leq 2\mathbb{E}((2T^2(\|x\|^2 + \|\delta(u)u\|^2) + 2TS(\|x\| + \|\delta(u)u\|) + S^2)\|u\|^2 + 2\sigma^2)\|\nabla \log p(u)\|^2 \\
&+ 2(T^2\|x\|^2 + 2TS\|x\| + S^2) \\
&= H_1\|x\|^2 + H_2\|x\| + H_3. \qquad \text{(B.10)}
\end{aligned}$$

■

Lemma B.3 *For each $x \in \mathbb{R}^d$,*

$$\mathbb{E}\|\widehat{h}(x,u) - \nabla L_(x)\| \leq H_4\|x\| + H_5, \tag{B.11}$$

where

$$H_4 = T(\mathbb{E}\|u\|\|\nabla \log p(u)\| + 1), \tag{B.12}$$

$$H_5 = T\mathbb{E}\|u\|^2\|\nabla \log p(u)\| + S\mathbb{E}\|u\|\|\nabla \log p(u)\| + \frac{2\sigma}{\sqrt{\pi}}\mathbb{E}\|\nabla \log p(u)\| + S \tag{B.13}$$

Proof Following the triangle inequality of the Euclidean norm, we have

$$\begin{aligned}
\mathbb{E}\|\widehat{h}(x,u) - \nabla L_(x)\| &= \mathbb{E}\|(L(x) - L(x+u) + \varepsilon - \varepsilon')\nabla \log p(u) - \nabla L(x)\| \\
&\leq \mathbb{E}|L(x) - L(x+u) + \varepsilon - \varepsilon'|\|\nabla \log p(u)\| + \mathbb{E}\|\nabla L(x)\|, \qquad \text{(B.14)}
\end{aligned}$$

the second term in B.14 can be upper-bounded by the linear bound of the norm of the gradient,

$$\mathbb{E}\|\nabla L(x)\| \leq T\|x\| + S \tag{B.15}$$

According to the triangle inequality and the mean-value theorem, for any $u \in \mathbb{R}^d$, there exists $\delta(u) \in (0,1)$ so that

$$\mathbb{E}|L(x) - L(x+u) + \varepsilon - \varepsilon'|\|\nabla \log p(u)\| \leq \mathbb{E}\|\nabla L(x+\delta(u)u)u\|\|\nabla \log p(u)\| + \frac{2\sigma}{\sqrt{\pi}}\mathbb{E}\|\nabla \log p(u)\| \tag{B.16}$$

$$\begin{aligned}
&\leq \mathbb{E}(T\|x+\delta(u)u\| + S)\|u\|\|\nabla \log p(u)\| + \frac{2\sigma}{\sqrt{\pi}}\mathbb{E}\|\nabla \log p(u)\| \\
&\leq T\mathbb{E}\|u\|\|\nabla \log p(u)\| \cdot \|x\| + T\mathbb{E}\|u\|^2\|p\nabla \log p(u)\| \\
&+ T\mathbb{E}\|u\|\|\nabla \log p(u)\| + \frac{2\sigma}{\sqrt{\pi}}\mathbb{E}\|\nabla \log p(u)\| \qquad \text{(B.17)}
\end{aligned}$$

The second term in B.16 comes from the expectation of $|\varepsilon - \varepsilon'|$ where $\varepsilon - \varepsilon' \sim \mathcal{N}(0, 2\sigma^2)$. Combining (B.15), (B.17), and (B.14), we can obtain the desired results. ■

Lemma B.4 (Second-order Error by Interpolation) *Let*

$$\widetilde{X}(t) = X_0 - \int_0^t \widehat{h}(\widetilde{X}(\lfloor q/\gamma \rfloor \gamma), \widetilde{u}(q))dq + \int_0^t dB(q), \qquad \forall t \geq 0 \tag{B.18}$$

be a continuous version of the discretized recursion using straightforward interpolation, where $\widetilde{u}(q) = u_m$ *when* $q \in [m\gamma, (m+1)\gamma)$*. Then, for all* $m \in \mathbb{N}, m \geq 0$ *and some* $q \in [m\gamma, (m+1)\gamma)$*, when* $\frac{a}{T} > \mathbb{E}\|u\|\|\nabla \log p(u)\| + 1$*, there exists constants* $0 < \delta < \frac{2(a - H_4)}{H_5}$ *and* $\alpha > 0$*, so that for any* $0 < \gamma < 1 \wedge \frac{2(a-H_4)-\delta H_5}{2T^2+H_1+2MH_4+(H_2/2+SH_4+TH_5)\delta}$*, we have*

$$\mathbb{E}\|\widetilde{X}(q) - \widetilde{X}(m\gamma)\|^2 \leq 3\gamma^2 \left(\left((H_1 + R^2) + H_{10}\alpha\right)(K_1(\delta) \vee K_2(\delta)) + \frac{H_{10}}{\alpha} + H_{11} \right) + \frac{6\gamma d}{\rho}, \tag{B.19}$$

where we define

$$\begin{aligned}
&H_6 = H_1 + 2H_4T + 2T^2, \qquad H_7 = \frac{H_2}{2} + SH_4 + TH_5, \\
&H_8 = \frac{2(TH_5 + SH_4) + H_2}{2}, \qquad H_9 = H_3 + 2SH_5 + 2S^2, \\
&H_{10} = \frac{H_2 + 2RS}{2}, \qquad H_{11} = H_3 + S^2, \\
&K_1(\delta) = \frac{2(TH_5 + H_5 + SH_4) + H_2}{2\delta} + H_3 + 2SH_5 + 2S^2 + 2b + \frac{2d}{\rho}, \\
&K_2(\delta) = \tau_0 + \frac{\frac{H_8 + H_5}{\delta} + H_9 + 2b + \frac{2d}{\rho}}{2a - 2H_4 - H_6 - (H_5 + H_7)\delta}.
\end{aligned} \tag{B.20}$$

Proof According to (B.18), It is shown that $\widetilde{X}(0) = X_0$, $\int_0^{m\gamma} dB(q) = B(m\gamma) \sim \sum_{i=0}^{m-1} \sqrt{\gamma}\zeta_i$, and $\int_0^{m\gamma} \widehat{h}(\widetilde{X}(\lfloor q/\gamma \rfloor \gamma), \widetilde{u}(q))dq = \gamma \sum_{i=0}^{m-1} \widehat{h}(X_i, u_i)$, hence $\widetilde{X}(m\gamma)$ and X_m are subjected to the same probability law. Therefore, $\widehat{h}(\widetilde{X}(\lfloor q/\gamma \rfloor \gamma), \widetilde{u}(q) \equiv \widehat{h}(X_m, u_m)$ when $q \in [m\gamma, (m+1)\gamma)$, which gives

$$\begin{aligned}
\mathbb{E}\|\widetilde{X}(q) - \widetilde{X}(m\gamma)\|^2 &= \mathbb{E}\|(m\gamma - q)\widehat{h}(X_m, u_m) + \sqrt{\frac{2}{\rho}}(B(q) - B(m\gamma))\|^2 \\
&= \mathbb{E}\|(q - m\gamma)(\nabla L(X_m) - \widehat{h}(X_m, u_m)) - (q - m\gamma)\nabla L(X_m) + \sqrt{\frac{2}{\rho}}(B(q) - B(m\gamma))\|^2 \\
&\leq 3\gamma^2 \mathbb{E}\|\nabla L(X_m) - \widehat{h}(X_m, u_m)\|^2 + 3\gamma^2 \mathbb{E}\|\nabla L(X_m)\|^2 + \frac{6}{\rho}\mathbb{E}\|B(q) - B(m\gamma)\|^2 && \text{(B.21)} \\
&= 3\gamma^2 \mathbb{E}\|\nabla L(X_m) - \widehat{h}(X_m, u_m)\|^2 + 3\gamma^2 \mathbb{E}\|\nabla L(X_m)\|^2 + \frac{6d}{\rho}(q - m\gamma) && \text{(B.22)} \\
&\leq 3\gamma^2 (H_1 \mathbb{E}\|X_m\|^2 + H_2 \mathbb{E}\|X_m\| + H_3) + 3\gamma^2 (R^2 \mathbb{E}\|X_m\|^2 + 2RS\mathbb{E}\|X_m\| + S^2) \\
&\quad + \frac{6d\gamma}{\rho}, && \text{(B.23)}
\end{aligned}$$

where $q - m\gamma \leq \gamma$ is used, (B.22) comes from Jensen's inequality for $\|\cdot\|^2$, and (B.21) is due to Lemma B.1 and Lemma B.2. From the fact

$$x \leq \frac{\alpha}{2}x^2 + \frac{1}{2\alpha}, \qquad \forall x \geq 0, \forall \alpha > 0, \tag{B.24}$$

we have

$$\mathbb{E}\|\widetilde{X}(q) - \widetilde{X}(m\gamma)\|^2 \le \left[3\gamma^2(H_1 + R^2) + \frac{3\gamma^2\alpha}{2}(H_2 + 2RS)\right]\mathbb{E}\|X_m\|^2 + 3\gamma^2\left(\frac{H_2 + 2RS}{2\alpha} + H_3 + S^2\right) + \frac{6\gamma d}{\rho}. \tag{B.25}$$

The term of second-order moment about the origin $\mathbb{E}\|X_m\|^2$ can be bounded through a contracted recursion. To show this, we rewrite X_{m+1} by X_m, u_m using (A.2),

$$\begin{aligned}
\mathbb{E}\|X_{m+1}\|^2 &= \mathbb{E}\|X_m - \gamma\widehat{h}(X_m, u_m)\|^2 + 2\sqrt{\frac{2\gamma}{\rho}}\mathbb{E}\langle X_m - \gamma\widehat{h}(X_m, u_m), \zeta_m\rangle + \frac{2\gamma}{\rho}\mathbb{E}\|\zeta_m\|^2 \\
&= \mathbb{E}\|X_m - \gamma\widehat{h}(X_m, u_m)\|^2 + \frac{2\gamma d}{\rho} \\
&= \mathbb{E}\|X_m - \gamma\nabla L(X_m)\|^2 + 2\gamma\mathbb{E}\langle X_m - \gamma\nabla L(X_m), \nabla L(X_m) - \widehat{h}(X_m, u_m)\rangle \\
&+ \gamma^2\mathbb{E}\|\nabla L(X_m) - \widehat{h}(X_m, u_m)\|^2 + \frac{2\gamma d}{\rho} \\
&\le \mathbb{E}\|X_m - \gamma\nabla L(X_m)\|^2 + 2\gamma\mathbb{E}\langle X_m - \gamma\nabla L(X_m), \nabla L(X_m) - \widehat{h}(X_m, u_m)\rangle \\
&+ \gamma^2(H_1\mathbb{E}\|X_m\|^2 + H_2\mathbb{E}\|X_m\| + H_3) + \frac{2\gamma d}{\rho},
\end{aligned} \tag{B.26}$$

where the second equality is due to the independence of ζ_m and $X_m - \gamma\widehat{h}(X_m, u_m)$ as well as $\mathbb{E}\zeta_m = 0$. We used Lemma B.2 again in the last inequality, of which the second term can be bounded by Cauchy-Schwarz inequality, triangle inequality for $\|\cdot\|$ and Lemma B.3,

$$\begin{aligned}
2\gamma\mathbb{E}\langle X_m - \gamma\nabla L(X_m), \nabla L(X_m) - \widehat{h}(X_m, u_m)\rangle &\le 2\gamma\mathbb{E}\|X_m - \gamma\nabla L(X_m)\|\|\nabla L(X_m) - \widehat{h}(X_m, u_m)\| \\
&= 2\gamma\mathbb{E}_{X_m}\left(\|X_m - \gamma\nabla L(X_m)\|\mathbb{E}_{u_m}\|\nabla L(X_m) - \widehat{h}(X_m, u_m)\|\right) \\
&\le 2\gamma\mathbb{E}_{X_m}(\|X_m\| + \gamma\|\nabla L(X_m)\|)(H_4\|X_m\| + H_5) \\
&\le 2\gamma\Big((1 + \gamma T)H_4\mathbb{E}\|X_m\|^2 \\
&+ ((1 + \gamma T)H_5 + \gamma S H_4)\mathbb{E}\|X_m\| + \gamma S H_5\Big).
\end{aligned} \tag{B.27}$$

The first term of the RHS of the last inequality in (B.26) is estimated by dissipativity assumption (A.3) and Lemma B.1 as follows,

$$\begin{aligned}
\mathbb{E}\|X_m - \gamma\nabla L(X_m)\|^2 &= \mathbb{E}\|X_m\|^2 - 2\gamma\mathbb{E}\langle X_m, \nabla L(X_m)\rangle + \gamma^2\mathbb{E}\|\nabla L(X_m)\|^2 \\
&\le \mathbb{E}\|X_m\|^2 - 2\gamma(a\mathbb{E}\|X_m\|^2 - b) + 2\gamma^2(T^2\mathbb{E}\|X_m\|^2 + S^2) \\
&= (1 + 2\gamma^2T^2 - 2\gamma a)\mathbb{E}\|X_m\|^2 + 2\gamma b + 2\gamma^2 S^2.
\end{aligned} \tag{B.28}$$

Plugging (B.27) and (B.28) into (B.26), we have

$$
\begin{aligned}
\mathbb{E}\|X_{m+1}\|^2 &\le (1+2\gamma^2T^2-2\gamma a+2\gamma(1+\gamma T)H_4+\gamma^2H_1)\mathbb{E}\|X_m\|^2 \\
&\quad + (2\gamma((1+\gamma T)H_5+\gamma SH_4)+\gamma^2H_2)\mathbb{E}\|X_m\| \\
&\quad + (H_3+2SH_5+2S^2)\gamma^2+2b\gamma+\frac{2\gamma d}{\rho} \\
&\le \left(\left(H_1+2H_4T+2T^2+\left(\frac{H_2}{2}+SH_4+TH_5\right)\delta\right)\gamma^2+(2H_4+\delta H_5-2a)\gamma+1\right)\mathbb{E}\|X_m\|^2 \\
&\quad + \left(\frac{2(TH_5+SH_4)+H_2}{2\delta}+H_3+2SH_5+2S^2\right)\gamma^2+\left(\frac{H_5}{\delta}+2b\right)\gamma+\frac{2\gamma d}{\rho}, \qquad \text{(B.29)}
\end{aligned}
$$

where $\delta > 0$, we have used the fact that $\mathbb{E}\|X_m\| \le \frac{\delta}{2}\mathbb{E}\|X_m\|^2 + \frac{1}{2\delta}$ in the last inequality. Then for $0<\gamma<1\wedge\frac{2(a-H_4)-\delta H_5}{2T^2+H_1+2MH_4+(H_2/2+SH_4+TH_5)\delta}$, we will have several conditions to consider as follows.

- When $(H_6+H_7\delta)\gamma^2+(2H_4+\delta H_5-2a)\gamma+1\le 0$, which requires $(H_2+2SH_4)(a-H_4)+2TH_5(a-2H_4-T)-H_1H_5>0$ to ensure the existence of such a δ that $4(H_6+H_7\delta)\le(2H_4+\delta H_5-2a)^2$, it follows from (B.29) that

$$
\begin{aligned}
\mathbb{E}\|X_m\|^2 &\le \left(\frac{H_8}{\delta}+H_9\right)\gamma^2+\left(\frac{H_5}{\delta}+2b+\frac{2d}{\rho}\right)\gamma \\
&\le \frac{2(TH_5+H_5+SH_4)+H_2}{2\delta}+H_3+2SH_5+2S^2+2b+\frac{2d}{\rho} \\
&= K_1(\delta). \qquad \text{(B.30)}
\end{aligned}
$$

- When $0<(H_6+H_7\delta)\gamma^2+(2H_4+\delta H_5-2a)\gamma+1<1$, which can be always satisfied due to the property of quadratic functions. Then we have the following summation according to recursion (B.29),

$$
\begin{aligned}
\mathbb{E}\|X_m\|^2 &= \left((H_6+H_7\delta)\gamma^2+(2H_4+\delta H_5-2a)\gamma+1\right)^m\mathbb{E}\|X_0\|^2 \\
&\quad + \left(\left(\frac{H_8}{\delta}+H_9\right)\gamma^2+\left(\frac{H_5}{\delta}+2b+\frac{2d}{\rho}\right)\gamma\right)\sum_{i=0}^{m-1}\left((H_6+H_7\delta)\gamma^2+(2H_4+\delta H_5-2a)\gamma+1\right)^i \\
&\le \mathbb{E}\|X_0\|^2+\frac{\frac{H_8+H_5}{\delta}+H_9+2b+\frac{2d}{\rho}}{2a-2H_4-H_6-(H_5+H_7)\delta} \\
&\le \log\mathbb{E}e^{\|X_0\|^2}+\frac{\frac{H_8+H_5}{\delta}+H_9+2b+\frac{2d}{\rho}}{2a-2H_4-H_6-(H_5+H_7)\delta} \\
&= \tau_0+\frac{\frac{H_8+H_5}{\delta}+H_9+2b+\frac{2d}{\rho}}{2a-2H_4-H_6-(H_5+H_7)\delta} \\
&= K_2(\delta). \qquad \text{(B.31)}
\end{aligned}
$$

From all above, we conclude that for all $m\in\mathbb{N}$,

$$
\mathbb{E}\|X_m\|^2 \le K_1(\delta)\vee K_2(\delta). \qquad \text{(B.32)}
$$

Combining (B.32) with (B.25) we obtain

$$\mathbb{E}\|\widetilde{X}(q) - \widetilde{X}(m\gamma)\|^2 \leq 3\gamma^2 \left(\left((H_1 + R^2) + H_{10}\alpha\right) \left(K_1(\delta) \vee K_2(\delta)\right) + \frac{H_{10}}{\alpha} + H_{11} \right) + \frac{6\gamma d}{\rho}. \tag{B.33}$$

Thus we proved the desired result.

∎

Lemma B.5 *When* $\frac{a}{T} > \mathbb{E}\|u\|\|\nabla \log p(u)\| + 1$*, there exists constants* $\delta < \frac{2(a-H_4)}{H_5}, \alpha > 0, \beta > 0$*, so that for* $m \in \mathbb{N}$ *and any* $0 < \gamma < \frac{2(a - H_4) - \delta H_5}{2T^2 + H_1 + 2MH_4 + (H_2/2 + SH_4 + TH_5)\delta}$*, we have*

$$\mathcal{W}(\mathbb{P}_m, \omega) \leq (\sqrt{\widetilde{C}_1(\alpha, \delta)}\gamma^{\frac{1}{4}} + \sqrt{\widetilde{C}_2(\beta, \delta)})m\gamma + C_3 e^{-\frac{m\gamma}{\rho C_{LS}}}, \tag{B.34}$$

where

$$\begin{aligned}
\widetilde{C}_1(\alpha, \delta) &= \left(12 + 8\left(\tau_0 + 2b + \frac{2d}{\rho}\right) m\gamma\right) (\sqrt{C_1(\alpha, \delta)} + C_1(\alpha, \delta)), \\
\widetilde{C}_2(\beta, \delta) &= \left(12 + 8\left(\tau_0 + 2b + \frac{2d}{\rho}\right) m\gamma\right) \cdot (\sqrt{C_2(\beta, \delta)} + C_1(\beta, \delta)), \\
C_3 &= \sqrt{2C_{LS}\left(\rho\left(\frac{T\tau_0}{3} + S\sqrt{\tau_0} + R + \frac{b}{2}\log 3\right) + \log\|f_0\|_\infty + \frac{d}{2}\log\frac{3\pi}{a\rho}\right)}.
\end{aligned} \tag{B.35}$$

Proof Let

$$\begin{aligned}
\sigma_k &:= \mathcal{L}(X_m), \qquad m = 1, 2, 3, ... \\
\lambda_t &:= \mathcal{L}(X(t)), \qquad t \geq 0
\end{aligned} \tag{B.36}$$

be the probability laws of discretized zeroth-order Langevin recursion (A.2) and Langevin diffusion (A.3) respectively. From (A.2), $\{X_m\}_{m=0}^{\infty}$ is a time-homogeneous Markov process. Based on (B.18) in Lemma B.4, another It$\widehat{o}$ process can be constructed using [18] as follows,

$$Y(t) = X_0 - \int_0^t h_q(Y(q))dq + \sqrt{\frac{2}{\rho}}\int_0^t dB(q), \tag{B.37}$$

where $h_q(Y(q)) = \mathbb{E}[\widehat{h}(\widetilde{X}(\lfloor q/\gamma \rfloor \gamma), \widetilde{u}(q))|\widetilde{X}(q) = Y(q)]$. Let

$$\begin{aligned}
\sigma_t &= \mathcal{L}(Y(q), 0 \leq q \leq t) \\
\lambda_t &= \mathcal{L}(X(q), 0 \leq q \leq t)
\end{aligned} \tag{B.38}$$

be the probability laws be two collections of probability laws of (B.5) and (A.3) respectively. By Girsanov formula [24], we can derive the Radon -Nikodym derivative for a change of measure as follows,

$$\frac{d\lambda_t}{d\sigma_t}(Y) = \exp\left\{\frac{\rho}{2}\int_0^t (\nabla L(Y(q)) - h_q(Y(q)))^* dB(q) - \frac{\rho}{4}\int_0^t \|\nabla L(Y(q)) - h_q(Y(q))\|^2 dq\right\}, \tag{B.39}$$

from which we can obtain the Kullback-Leibler divergence from λ_t to σ_t

$$
\begin{aligned}
\mathcal{D}_{KL}(\sigma_t \| \lambda_t) &= -\int_{\Xi} d\sigma_t \log \frac{d\lambda_t}{d\sigma_t} \\
&= -\frac{\rho}{2}\int_{\Xi} d\sigma_t \int_0^t (\nabla L(Y(q)) - h_q(Y(q)))^* dB(q) + \frac{\rho}{4}\int_{\Xi} d\sigma_t \int_0^t \|\nabla L(Y(q)) - h_q(Y(q))\|^2 dq \\
&= \frac{\rho}{4}\int_0^t \mathbb{E}\|\nabla L(Y(q)) - h_q(Y(q))\|^2 dq, \qquad \text{(B.40)}
\end{aligned}
$$

where the last equation is due to the property of stochastic integral, that is, for $\nabla L(Y(q)) - h_q(Y(q)) \in L^2([0,\infty) \times \Xi)$,

$$
\mathbb{E}\int_0^t (\nabla L(Y(q)) - h_q(Y(q)))^* dB(q) = 0. \qquad \text{(B.41)}
$$

Next we discretize t by $t = m\gamma$ for all $m \in \mathbb{N}$ and proceed to expand (B.40),

$$
\begin{aligned}
\mathcal{D}_{KL}(\sigma_{m\gamma} \| \lambda_{m\gamma}) &= \frac{\rho}{4}\sum_{i=0}^{m-1}\int_{i\gamma}^{(i+1)\gamma} \mathbb{E}\|\nabla L(Y(q)) - h_q(Y(q))\|^2 dq \\
&= \frac{\rho}{4}\sum_{i=0}^{m-1}\int_{i\gamma}^{(i+1)\gamma} \mathbb{E}_{\widetilde{X}(q)}\|\nabla L(\widetilde{X}(q)) - h_q(\widetilde{X}(q))\|^2 dq \\
&= \frac{\rho}{4}\sum_{i=0}^{m-1}\int_{i\gamma}^{(i+1)\gamma} \mathbb{E}_{\widetilde{X}(q)}\|\nabla L(\widetilde{X}(q)) - \mathbb{E}_{\widetilde{X}(\lfloor q/\gamma\rfloor\gamma),\widetilde{u}(q)}\widehat{h}(\widetilde{X}(\lfloor q/\gamma\rfloor\gamma), \widetilde{u}(q))\|^2\|^2 dq \\
&= \frac{\rho}{4}\sum_{i=0}^{m-1}\int_{i\gamma}^{(i+1)\gamma} \mathbb{E}_{\widetilde{X}(q)}\|\mathbb{E}_{\widetilde{X}(\lfloor q/\gamma\rfloor\gamma),\widetilde{u}(q)}[\nabla L(\widetilde{X}(q)) - \widehat{h}(\widetilde{X}(\lfloor q/\gamma\rfloor\gamma), \widetilde{u}(q))]\|^2 dq \\
&\le \frac{\rho}{4}\sum_{i=0}^{m-1}\int_{i\gamma}^{(i+1)\gamma} \mathbb{E}_{\widetilde{X}(q),\widetilde{X}(\lfloor q/\gamma\rfloor\gamma),\widetilde{u}(q)}\|\nabla L(\widetilde{X}(q)) - \widehat{h}(\widetilde{X}(\lfloor q/\gamma\rfloor\gamma), \widetilde{u}(q))\|^2 dq, \quad \text{(B.42)}
\end{aligned}
$$

where the second equality is due to the construction of (B.37) where $\mathcal{L}(\widetilde{X}(q)) = \mathcal{L}(Y(q))$, the last line follows from Jensen's inequality. Then, using Lipschitzian condition (A.7), the defintion of $\widetilde{u}(q)$ and Jensen's inequality again, it follows that

$$
\begin{aligned}
\mathcal{D}_{KL}(\sigma_{m\gamma} \| \lambda_{m\gamma}) &\le \frac{\rho}{2}\Bigg(\sum_{i=0}^{m-1}\int_{i\gamma}^{(i+1)\gamma} \mathbb{E}_{\widetilde{X}(q),\widetilde{X}(\lfloor q/\gamma\rfloor\gamma),\widetilde{u}(q)}\|\nabla L(\widetilde{X}(q)) - \nabla L(\widetilde{X}(\lfloor q/\gamma\rfloor\gamma))\|^2 dq \\
&\quad + \sum_{i=0}^{m-1}\int_{i\gamma}^{(i+1)\gamma} \mathbb{E}_{\widetilde{X}(q),\widetilde{X}(\lfloor q/\gamma\rfloor\gamma),\widetilde{u}(q)}\|\nabla L(\widetilde{X}(\lfloor q/\gamma\rfloor\gamma)) - \widehat{h}(\widetilde{X}(\lfloor q/\gamma\rfloor\gamma), \widetilde{u}(q))\|^2 dq\Bigg), \\
&\le \frac{\rho T^2}{2}\sum_{i=0}^{m-1}\int_{i\gamma}^{(i+1)\gamma} \mathbb{E}\|\widetilde{X}(q) - \widetilde{X}(\lfloor q/\gamma\rfloor\gamma)\|^2 dq + \frac{\rho\gamma}{2}\sum_{i=0}^{m-1}\mathbb{E}\|\nabla L(X_i) - \widehat{h}(X_i, u_i)\|^2. \qquad \text{(B.43)}
\end{aligned}
$$

Using Lemma B.2 and Lemma B.4 for each corresponding term in (B.42) while noting $0 < \gamma < 1$, we can write

$$\begin{aligned}
\mathcal{D}_{KL}(\sigma_{m\gamma}\|\lambda_{m\gamma}) &\leq \frac{\rho T^2}{2}\left(3m\gamma^3\left(\left((H_1+R^2)+H_{10}\alpha\right)(K_1(\delta)\vee K_2(\delta))+\frac{H_{10}}{\alpha}+H_{11}\right)+\frac{6m\gamma^2 d}{\rho}\right)\\
&\quad+\frac{\rho\gamma}{2}\sum_{i=0}^{m-1}(H_1\mathbb{E}\|X_i\|^2+H_2\|X_i\|+H_3)\\
&\leq C_1(\alpha,\delta)m\gamma^2+\frac{\rho m\gamma}{2}\left(\left(H_1+\frac{H_2\beta}{2}\right)(K_1(\delta)\vee K_2(\delta)+\frac{H_2}{2\beta}+H_3\right)\\
&= C_1(\alpha,\delta)m\gamma^2+C_2(\beta,\delta)m\gamma, \qquad \text{(B.44)}
\end{aligned}$$

where $C_1(\alpha,\delta)=\dfrac{3\rho T^2}{2}\left(\left((H_1+R^2)+H_{10}\alpha\right)(K_1(\delta)\vee K_2(\delta))+\dfrac{H_{10}}{\alpha}+H_{11}+\dfrac{2d}{\rho}\right), C_2(\beta,\delta)=\dfrac{\rho}{2}\left(\left(H_1+\dfrac{H_2\beta}{2}\right)(\right.$ with a constant $\beta > 0$. Next, from Data-processing inequality we have

$$\mathcal{D}_{KL}(\mathbb{P}_m\|\mathbb{Q}_{m\gamma}) \leq \mathcal{D}_{KL}(\sigma_{m\gamma}\|\lambda_{m\gamma}) \leq C_1(\alpha,\delta)m\gamma^2+C_2(\beta,\delta)m\gamma. \qquad \text{(B.45)}$$

Then, letting $\tau_0 = \log t_0$, we proceed to use a weighted transportation cost inequality from [3] and exponential integrability of Langevin diffusion in [26] to obtain the Euclidean Wasserstein distance of $\mathbb{P}_m$ and $\mathbb{Q}_{m\gamma}$,

$$\begin{aligned}
\mathcal{W}^2(\mathbb{P}_m,\mathbb{Q}_{m\gamma}) &\leq \left(12+8\left(\tau_0+2b+\frac{2d}{\rho}\right)m\gamma\right)\left(\sqrt{\mathcal{D}_{KL}(\mathbb{P}_m\|\mathbb{Q}_{m\gamma})}+\mathcal{D}_{KL}(\mathbb{P}_m\|\mathbb{Q}_{m\gamma})\right)\\
&\leq \left(12+8\left(\tau_0+2b+\frac{2d}{\rho}\right)m\gamma\right)\left((C_1(\alpha,\delta)\gamma+C_2(\beta,\delta))m\gamma+(\sqrt{C_1(\alpha,\delta)\gamma}+\sqrt{C_2(\beta,\delta)})m\gamma\right)\\
&\leq \left(12+8\left(\tau_0+2b+\frac{2d}{\rho}\right)m\gamma\right)\left((\sqrt{C_2(\beta,\delta)}+C_2(\beta,\delta))m\gamma+(\sqrt{C_1(\alpha,\delta)}+C_1(\alpha,\delta))m\gamma^{\frac{3}{2}}\right)\\
&= (\widetilde{C}_1(\alpha,\delta)\gamma^{\frac{1}{2}}+\widetilde{C}_2(\beta,\delta))m^2\gamma^2, \qquad \text{(B.46)}
\end{aligned}$$

where $0<\gamma<1$ and $m\gamma \geq 1$ are used in the second and the third inequalities. We also have let $\widetilde{C}_1(\alpha,\delta) := \left(12+8\left(\tau_0+2b+\dfrac{2d}{\rho}\right)m\gamma\right)(\sqrt{C_1(\alpha,\delta)}+C_1(\alpha,\delta))$ and $\widetilde{C}_2(\beta,\delta) := \left(12+8\left(\tau_0+2b+\dfrac{2d}{\rho}\right)m\gamma\right)$ $\cdot(\sqrt{C_2(\beta,\delta)}+C_1(\beta,\delta))$. From logarithmic Sobolev inequality [26] for ω

$$\mathcal{W}(\mathbb{Q}_t,\omega) \leq \sqrt{2C_{LS}\mathcal{D}_{KL}(\mathbb{Q}_t\|\omega)} \qquad \text{(B.47)}$$

with the constant $C_{LS} \leq \dfrac{8T^2+2a^2}{\rho a^2 T}+\dfrac{1}{\Delta_\lambda}\left(\dfrac{6(\rho+d)T}{a}+2\right)$. Then, by the theorem on exponential decay of entropy and Otto-Villani theorem [26], we have

$$\mathcal{D}_{KL}(\mathbb{Q}_t\|\omega) \leq e^{-\frac{2t}{\rho C_{LS}}}\mathcal{D}_{KL}(\mathbb{P}_0\|\omega). \qquad \text{(B.48)}$$

According to the expression of τ_0 (Eq. (A.2)), we can obtain a relative entropy bound of $\mathbb{P}_0$ and ω,

$$\mathcal{D}_{KL}(\mathbb{P}_0\|\omega) \leq \rho\left(\frac{T\tau_0}{3}+S\sqrt{\tau_0}+R+\frac{b}{2}\log 3\right)+\log\|f_0\|_\infty+\frac{d}{2}\log\frac{3\pi}{a\rho}. \qquad \text{(B.49)}$$

By combining Eqs. (B.45)-(B.49) and triangle inequality for $\mathcal{W}(\cdot,\cdot)$, we can derive that for all $m \in \mathbb{N}$,

$$\begin{aligned}
\mathcal{W}(\mathbb{P}_m, \omega) &\leq \mathcal{W}(\mathbb{P}_m, \mathbb{Q}_{m\gamma}) + \mathcal{W}(\mathbb{Q}_{m\gamma}, \omega) \\
&\leq (\sqrt{\widetilde{C}_1(\alpha,\delta)}\gamma^{\frac{1}{4}} + \sqrt{\widetilde{C}_2(\beta,\delta)})m\gamma \\
&+ e^{-\frac{m\gamma}{\rho C_{LS}}}\sqrt{2C_{LS}\left(\rho\left(\frac{T\tau_0}{3} + S\sqrt{\tau_0} + R + \frac{b}{2}\log 3\right) + \log\|f_0\|_\infty + \frac{d}{2}\log\frac{3\pi}{a\rho}\right)} \\
&= (\sqrt{\widetilde{C}_1(\alpha,\delta)}\gamma^{\frac{1}{4}} + \sqrt{\widetilde{C}_2(\beta,\delta)})m\gamma + C_3 e^{-\frac{m\gamma}{\rho C_{LS}}}. \qquad \text{(B.50)}
\end{aligned}$$

■