

关于相关数据流的随机梯度哈密尔顿蒙特卡罗算法的非渐近估计

杨金圆, 胡良剑*

东华大学理学院, 上海

收稿日期: 2024年1月26日; 录用日期: 2024年2月21日; 发布日期: 2024年2月28日

摘要

近年来, 基于朗之万蒙特卡罗方法的随机梯度下降算法得到了广泛应用。这些算法通过在梯度的估计中注入适当的高斯噪声以实现在非凸优化问题中的全局收敛。随机梯度哈密尔顿蒙特卡罗 (SGHMC) 是随机梯度下降带有动量的一种变体, 通常的研究以样本数据相互独立的假设为前提来分析 SGHMC 算法的收敛性, 然而实际中的样本数据往往存在相关性。本文在数据流具有相关性 (满足一定的条件混合特性) 的条件下, 给出了 SGHMC 算法的非渐近估计, 建立了全局 Lipschitz 条件下 SGHMC 算法的收敛性定理, 得到了迭代分布与目标分布之间 Wasserstein 距离的上界。

关键词

随机梯度哈密尔顿蒙特卡罗, 非渐近估计, 非凸优化, 相关数据流

Non-Asymptotic Estimates for Stochastic Gradient Hamiltonian Monte Carlo Algorithm with Dependent Data Streams

Jinyuan Yang, Liangjian Hu*

College of Science, Donghua University, Shanghai

Received: Jan. 26th, 2024; accepted: Feb. 21st, 2024; published: Feb. 28th, 2024

Abstract

In recent years, stochastic gradient descent algorithms based on the Langevin Monte Carlo method have been widely applied, which achieve global convergence in non-convex optimization problems

*通讯作者。

文章引用: 杨金圆, 胡良剑. 关于相关数据流的随机梯度哈密尔顿蒙特卡罗算法的非渐近估计[J]. 应用数学进展, 2024, 13(2): 569-583. DOI: 10.12677/aam.2024.132055

by injecting appropriate Gaussian noise into the gradient estimates. Stochastic gradient Hamiltonian Monte Carlo (SGHMC) is a variant of stochastic gradient descent with momentum. Usually, studies analyze the convergence of algorithm based on the assumption that the sample data are i.i.d. However, in practice, sample data may be dependent. This paper provides non-asymptotic estimates for SGHMC algorithm under the condition that the data streams are dependent (satisfying a certain conditional mixing property), establishes a convergence theorem for SGHMC algorithm under the global Lipschitz condition, and obtains an upper bound of the Wasserstein distance between the law of algorithm's iterates and the target distribution.

Keywords

Stochastic Gradient Hamiltonian Monte Carlo, Non-Asymptotic Estimates, Non-Convex Optimization, Dependent Data Streams

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随机梯度下降算法在解决非凸优化问题中有着广泛应用。考虑一个随机优化问题:

$$\text{最小化 } U(\theta) := \mathbb{E}[u(\theta, X)]$$

其中 U 可能是非凸损失函数, $\theta \in \mathbb{R}^d$, X 是随机数据。我们希望得到 θ 的估计量 $\hat{\theta}$ 使超额期望风险 $\mathbb{E}[U(\hat{\theta})] - \inf_{\theta \in \mathbb{R}^d} U(\theta)$ 达到最小。Raginsky 等[1]阐述了最小化超额期望风险与从一个集中在损失函数 U 的极小值附近的分布

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

中采样之间的密切联系, 其中 β 为逆温度系数。当 β 足够大时, 目标分布 π_β 集中在损失函数 U 的极小值附近[2], 且在适当的假设下, 目标分布 π_β 是过阻尼郎之万随机微分方程

$$d\theta_t = -h(\theta_t)dt + \sqrt{2\beta^{-1}}dB_t \tag{1}$$

的唯一平稳分布[3], 其中 $h := \nabla U$, $\{B_t\}_{t \geq 0}$ 为 d 维布朗运动。

在非凸情形下损失函数 U 可能具有多个极小值, 为算法收敛带来挑战。为了从目标分布 π_β 中采样, 基于朗之万蒙特卡罗(Langevin Monte Carlo, LMC)方法的随机梯度下降(Stochastic Gradient Descent, SGD)算法备受关注, 这些算法通过在梯度的估计中注入噪声保证全局收敛。考虑使用欧拉离散化方法近似模拟方程(1), 同时由于精确梯度 h 难以获得, 故使用其以观测数据驱动的无偏估计代替, 即得到随机梯度郎之万动力学(Stochastic Gradient Langevin Dynamics, SGLD)算法[4]

$$\theta_{n+1}^n := \theta_n^n - \eta H(\theta_n^n, X_{n+1}) + \sqrt{2\eta\beta^{-1}}\xi_{n+1}$$

其中 $\eta > 0$ 为步长, $H: \mathbb{R}^d \times \mathbb{R}^m \rightarrow \mathbb{R}^d$ 为满足 $\mathbb{E}H(\theta, X_n) = h(\theta)$ 的可测函数, $\{X_n\}_{n \in \mathbb{N}}$ 为适应于给定流 $\{\mathcal{G}_n\}_{n \in \mathbb{N}}$ 的 \mathbb{R}^m 值数据序列, $\{\xi_n\}_{n \in \mathbb{N}}$ 为独立的标准 d 维高斯噪声。

另一方面, 本文所考虑的随机梯度哈密尔顿蒙特卡罗(Stochastic Gradient Hamiltonian Monte Carlo, SGHMC)算法[5]基于欠阻尼郎之万随机微分方程

$$\begin{aligned}dV_t &= -\gamma V_t dt - h(\theta_t) dt + \sqrt{2\gamma\beta^{-1}} dB_t \\d\theta_t &= V_t dt\end{aligned}\quad (2)$$

其中 $\{\theta_t\}_{t \geq 0}$, $\{V_t\}_{t \geq 0}$ 分别为位置和动量过程, γ 为摩擦系数。在梯度的适当假设下, 马尔可夫过程 $(\theta_t, V_t)_{t \geq 0}$ 具有唯一平稳分布[6]

$$\bar{\pi}_\beta(d\theta, dv) \propto \exp\left(-\beta\left(\frac{1}{2}|v|^2 + U(\theta)\right)\right) d\theta dv$$

与 SGLD 算法类似, 精确梯度 h 难以获得, 但可以有效获得其无偏估计, 即得到 SGHMC 算法

$$\begin{aligned}V_{n+1}^n &= V_n^n - \eta \left[\gamma V_n^n + H(\theta_n^n, X_{n+1}) \right] + \sqrt{2\gamma\eta\beta^{-1}} \xi_{n+1} \\ \theta_{n+1}^n &= \theta_n^n + \eta V_n^n\end{aligned}\quad (3)$$

Eberle 等[7]证明了在某些假设条件下欠阻尼随机微分方程在 Wasserstein 距离下收敛到其平稳分布的速度比过阻尼情形快, 许多实验也印证了 SGHMC 算法表现出比 SGLD 算法更快的收敛速度[5]。

近年来, 损失函数 U 为凸情形下的近似采样算法的收敛速度得到了深入研究, 放宽凸假设是一个更具挑战的问题。对于 U 非凸的情形, 通常考虑两种路径: (i) 考虑耗散条件[1] [8]; (ii) 假设损失函数 U 在无穷远处具有凸性[9]。此外, 大多数研究假设样本数据序列相互独立, 然而事实上数据可能相关而非独立。如金融市场价格分析中, 资产收益在适当的时间尺度上保持平稳, 但独立性失效[10]。在下达交易订单以最大化投资组合效益、最小化交易成本等金融问题中, 随机梯度下降算法往往在相关数据流的基础上进行[8]。Chau 等[8]在相关数据流(样本数据满足条件 L 混合特性)假设下研究了非凸情形下 SGLD 算法的收敛性。对于 SGHMC 算法, 在观测数据独立同分布的假设下, Chau 等[6]和 Gao 等[11]研究了 SGHMC 算法在全局 Lipschitz 条件下的收敛性。本文将考虑的数据样本 $\{X_n\}_{n \in \mathbb{N}}$ 是相关的, 在数据序列满足条件 L 混合特性的条件下, 得到 SGHMC 算法的非渐近估计, 并由此得到 SGHMC 算法收敛到其目标分布的速度。

2. 符号与假设

2.1. 符号

设 (Ω, \mathcal{F}, P) 为一个概率空间。记随机变量 X 的期望为 $\mathbb{E}X$, 用 L^p 表示 p 阶矩可积的实值随机变量的空间。给定整数 $d \geq 1$, 记 \mathbb{R}^d 上 Borel σ 代数 $\mathcal{B}(\mathbb{R}^d)$, 随机变量 X 在 $\mathcal{B}(\mathbb{R}^d)$ 上的分布用 $\mathcal{L}(X)$ 表示。记内积为 $\langle \cdot, \cdot \rangle$, 相应的范数用 $|\cdot|$ 表示。对于 $r \in \mathbb{R}_+$, 用 B_r 表示中心为 0 , 半径为 r 的闭球。可测空间 $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ 上的概率测度的集合记为 $\mathcal{P}(\mathbb{R}^d)$ 。对 $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, 用 $\mathcal{C}(\mu, \nu)$ 表示 $\mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d)$ 上的概率测度 Γ 的集合, 其边际分布分别为 μ, ν 。Wasserstein 距离定义为

$$W_2(\mu, \nu) := \inf_{\Gamma \in \mathcal{C}(\mu, \nu)} \left(\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |\theta - \theta'|^2 \Gamma(d\theta, d\theta') \right)^{1/2}.$$

2.2. 假设

假设 2.1 损失函数 U 取非负值, 即 $U(\theta) \geq 0$ 。

假设 2.2 $|\theta_0|, |v_0| \in \cap_{p \geq 1} L^p$ 。

假设 2.3 设 $\mathcal{G}_n, n \in \mathbb{N}$ 为过去信息的给定流, 且 $\mathcal{G}_0 = \{\emptyset, \Omega\}$, $\mathcal{G}_\infty := \sigma(\cup_{n \in \mathbb{N}} \mathcal{G}_n)$ 。设 $\mathcal{G}_n^*, n \in \mathbb{N}$ 为一个递减的 σ 代数族, 使得对所有 $n \in \mathbb{N}$, \mathcal{G}_n 和 \mathcal{G}_n^* 相互独立。过程 $\{X_n\}_{n \in \mathbb{N}}$ 关于 $\{\mathcal{G}_n, \mathcal{G}_n^*\}_{n \in \mathbb{N}}$ 是条件 L 混合的, 且满足对任意 $\theta \in \mathbb{R}^d$ 和 $n \geq 1$,

$$\mathbb{E}H(\theta, X_n) = h(\theta). \tag{4}$$

假设 2.4 存在正常数 L_1, L_2 , 使得对所有 $\theta, \theta' \in \mathbb{R}^d$ 和 $x, x' \in \mathbb{R}^m$, 有

$$|H(\theta, x) - H(\theta', x')| \leq L_1 |\theta - \theta'| + L_2 |x - x'|. \tag{5}$$

假设 2.5 存在正常数 a, b 使得对所有 $\theta \in \mathbb{R}^d$ 和 $x \in \mathbb{R}^m$,

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

注 2.6 由假设 2.4, 得到以下性质:

(i) 对所有 $\theta, x \in \mathbb{R}^d$,

$$|H(\theta, x)| \leq L_1 |\theta| + L_2 |x| + H_*. \tag{6}$$

其中 $H_* := |H(0, 0)|$.

(ii) 对所有 $\theta \in \mathbb{R}^d$,

$$|h(\theta)| \leq L_1 |\theta| + h_0 \tag{7}$$

其中 $h_0 := |h(0)|$.

3. 主要结果

3.1. 条件 L 混合

Gerencsér [12]介绍了 L 混合过程, L 混合表示随机序列中相邻元素间相关性快速减小。Chau 等[13]在随机梯度方法的背景下引入了条件 L 混合的概念, 条件 L 混合表示在给定一些信息或条件下序列的混合性质仍然成立。考虑概率空间 $(\Omega, \mathcal{F}, \mathbb{P})$, 具有离散时间流 $\{\mathcal{G}_n\}_{n \in \mathbb{N}}$ 和一个递减的 σ 域序列 $\{\mathcal{G}_n^*\}_{n \in \mathbb{N}}$, 使得对所有 $n \in \mathbb{N}$, \mathcal{G}_n 和 \mathcal{G}_n^* 相互独立。当 $r \geq 1$ 时, 若 $\sup_{n \in \mathbb{N}} \mathbb{E}^{1/r} [|X_n|^r] < \infty$, 随机过程 $\{X_n\}_{n \in \mathbb{N}}$ 称为 L' 有界。

对每个 $n, \tau \in \mathbb{N}$ 和 $i = 1, \dots, d$, 定义

$$\begin{aligned} \tilde{M}_r^n(X, i) &:= \sup_{m \in \mathbb{N}} \mathbb{E}^{1/r} [|X_{n+m}^i|^r | \mathcal{G}_n] \\ \tilde{\gamma}_r^n(\tau, X, i) &:= \sup_{m \geq \tau} \mathbb{E}^{1/r} [|X_{n+m}^i - \mathbb{E}[X_{n+m}^i | \mathcal{G}_{n+m-\tau}^* \vee \mathcal{G}_n]|^r | \mathcal{G}_n] \end{aligned}$$

其中 X_{n+m}^i 为 X_{n+m} 的第 i 个分量, 且设

$$\tilde{\Gamma}_r^n(X, i) := \sum_{\tau=0}^{\infty} \tilde{\gamma}_r^n(\tau, X, i), M_r^n(X) := \sum_{i=1}^d \tilde{M}_r^n(X, i), \Gamma_r^n(X) := \sum_{i=1}^d \tilde{\Gamma}_r^n(X, i). \tag{8}$$

定义 3.1 (条件 L 混合) 称随机过程 $\{X_n\}_{n \in \mathbb{N}}$ 关于 $\{\mathcal{G}_n, \mathcal{G}_n^*\}_{n \in \mathbb{N}}$ 为条件 L 混合, 如果满足以下条件:

- (i) 对所有 $\theta \in \Theta$, $\{X_n\}_{n \in \mathbb{N}}$ 均为 $\{\mathcal{G}_n\}_{n \in \mathbb{N}}$ 适应;
- (ii) 对所有 $r \geq 1L'$, 随机过程 $\{X_n\}_{n \in \mathbb{N}}$ 为 L' 有界;
- (iii) 对所有 $r \geq 1$, 序列 $\{M_r^n(X)\}_{n \in \mathbb{N}}$, $\{\Gamma_r^n(X)\}_{n \in \mathbb{N}}$ 为 L' 有界。

由上述定义, 若 $\{X_n\}_{n \in \mathbb{N}}$ 为独立序列且满足(i)和(ii), 则(iii)也成立, 可视为条件 L 混合过程的特殊情况。设 $\{X_n\}_{n \in \mathbb{N}}$ 为条件 L 混合过程, 对 $r \geq 1$ 引入量 $\mathcal{M}_r(X) := \sup_{n \in \mathbb{N}} \mathbb{E} |X_n|^r$ 。下面的结论分别来自参考文献[14]

的定理 B.5、引理 6.4 和例 2.4、引理 A.3。

定理 3.2 对 $r > 2$, 设 $\{W_t\}_{t \in \mathbb{R}_+}$ 为 L^r 有界, 且设 $M_r(W) + \Gamma_r(W) < \infty$ a.s.。假设对 $t \in \mathbb{R}_+$, $\mathbb{E}[W_t | \mathcal{G}_0] = 0$ a.s.。设 $f: [0, T] \rightarrow \mathbb{R}$ 为 $\mathcal{B}([0, T])$ 可测, 且有 $\int_0^T f_t^2 dt < \infty$, 则存在一个常数 $C'(r)$ 使得

$$\mathbb{E}^{1/r} \left[\sup_{s \in [0, T]} \left| \int_0^s f_t W_t dt \right|^r \middle| \mathcal{G}_0 \right] \leq C'(r) \left(\int_0^T f_t^2 dt \right)^{1/2} [M_r(W) + \Gamma_r(W)], \text{ a.s.}$$

这里可以取 $C'(r) = \sqrt{r-1} (2^{1/2} - 2^{1/r})^{-1}$ 。

引理 3.3 若假设 2.4 成立, 则对每个 $j \in \mathbb{N}$, $H(\theta, W_t), t \in \mathbb{R}_+, \theta \in B_j$ 满足

$$M_r(H(\theta, W)) \leq L_1 j + L_2 M_r(W) + H_*, \Gamma_r(H(\theta, W)) \leq 2L_2 \Gamma_r(W).$$

引理 3.4 若假设 2.4 成立, 且设 $j \in \mathbb{N}$ 。令 $(Z_s)_{s \geq 0}$ 为 B_j 值随机变量族, 满足 $Z: \mathbb{R}_+ \times \Omega \rightarrow \mathbb{R}^d$ 是 $\mathcal{B}(\mathbb{R}_+) \otimes \mathcal{G}_0$ 可测的。对 $t \in \mathbb{R}_+$, 定义过程 $Y_t = H(Z_t, W_t)$, 则

$$M_r(Y) \leq L_1 j + L_2 M_r(W) + H_*, \Gamma_r(Y) \leq 2L_2 \Gamma_r(W).$$

3.2. 收敛定理

首先定义

$$\eta_{\max} = \min \left\{ 1, \frac{K_3}{K_2}, \frac{\gamma\lambda}{2K_1}, \frac{2}{\gamma\lambda}, \frac{\gamma\lambda}{2\tilde{K}_1}, \frac{\lambda\gamma}{4K_4} \right\}$$

其中常数 K_1, K_2, K_3 将在引理 4.2 的证明中给出, 常数 \tilde{K}_1 和 K_4 将在引理 4.3 的证明中给出, 参数 λ 在(9)中引入。下面给出 SGHMC 算法到目标分布的收敛性的主要结果。

定理 3.5 若假设 2.1~2.5 成立且 $0 < \eta \leq \eta_{\max}$, 则存在常数 $C_1, C_2, C_3, C_4 > 0$, 使得

$$W_2(\mathcal{L}(\theta_n^\eta, V_n^\eta), \bar{\pi}_\beta) \leq C_1 \eta^{1/2} + C_2 \eta^{1/4} + C_3 e^{-C_4 \eta^n}$$

其中 C_1, C_2 依赖于 $\gamma, \beta, d, L_1, L_2$ 和数据流 $(X_n)_{n \in \mathbb{N}}$, 具体形式分别在定理 4.6 和定理 4.8 的证明中给出, C_3, C_4 依赖于 γ, β, d, L_1 , 具体形式在定理 3.5 的证明中给出。

本文在数据流存在相关性的假设下给出了 SGHMC 算法的非渐近估计, 阐明了 SGHMC 算法在有限迭代次数 n 下的性能。定理 3.5 表明算法的误差在迭代次数 n 上一致有界且选择足够小的步长 $\eta > 0$ 可以使误差达到任意小, 从而保证了算法的收敛性在相关数据流情形下依然成立。与独立数据流的情形相比, 本文在定理证明过程中采用与文献[6]不同的证明路线。本文将 SGHMC 算法与依赖于条件 L 混合的辅助过程进行比较, 得到相关数据流情形下二者之间的 Wasserstein 距离。此外, 本文在算法的误差估计中得到不同的系数表达式, 条件 L 混合特性为这些系数的有界性提供了保证。

4. 证明

4.1. 辅助过程

证明主要结果的中心思想是引入合适的 Lyapunov 函数及 SGHMC 算法的连续时间插值过程等辅助过程, 该连续时间插值过程在离散时间下的分布与 SGHMC 算法相同。不同于 SGLD 算法中经典的 Lyapunov 函数 $\mathcal{V}(\theta) = (1 + |\theta|^2)^{p/2}$, 本文使用 Eberle [7] 定义的 Lyapunov 函数:

$$\mathcal{V}(\theta, v) = \beta U(\theta) + \frac{\beta}{4} \gamma^2 \left(|\theta + \gamma^{-1} v|^2 + |\gamma^{-1} v|^2 - \lambda |\theta|^2 \right) \tag{9}$$

其中 $\lambda \in (0, 1/4]$, 该函数依赖于目标函数 $U(\theta)$ 。

下面定义 (θ_η, V_η) 的缩放过程 $(\zeta_t^\eta, Z_t^\eta) := (\theta_\eta, V_\eta)$,

$$\begin{aligned} dZ_t^\eta &= -\eta(\gamma Z_t^\eta + h(\zeta_t^\eta))dt + \sqrt{2\gamma\eta\beta^{-1}}dB_t^\eta \\ d\zeta_t^\eta &= \eta Z_t^\eta dt \end{aligned} \tag{10}$$

其中 $B_t^\eta := B_{\eta t} / \sqrt{\eta}, t \geq 0$ 。记 $\{B_t\}_{t \geq 0}$ 自然流为 $\{\mathcal{F}_t\}_{t \geq 0}$, 且 $\{\mathcal{F}_t\}_{t \geq 0}$ 与 $\mathcal{G}_\infty \vee \sigma(\theta_0, v_0)$ 相互独立, 则 $\{B_t^\eta\}_{t \geq 0}$ 自然流记为 $\{\mathcal{F}_t^\eta\}_{t \geq 0}$ 且 $\mathcal{F}_t^\eta := \mathcal{F}_{\eta t}$, $\{\mathcal{F}_t^\eta\}_{t \geq 0}$ 也与 $\mathcal{G}_\infty \vee \sigma(\theta_0, v_0)$ 相互独立。此外, 记 $\mathcal{F}_\infty := \sigma(\cup_{t \geq 0} \mathcal{F}_t)$ 。

然后定义 SGHMC 算法(3)的连续时间插值过程:

$$\begin{aligned} d\bar{V}_t^\eta &= -\eta(\gamma \bar{V}_{\lfloor t \rfloor}^\eta + H(\bar{\theta}_{\lfloor t \rfloor}^\eta, X_{\lfloor t \rfloor+1}))dt + \sqrt{2\gamma\eta\beta^{-1}}dB_t^\eta \\ d\bar{\theta}_t^\eta &= \eta \bar{V}_{\lfloor t \rfloor}^\eta dt \end{aligned} \tag{11}$$

注意到, 插值过程(11)与 SGHMC 算法在格点上具有相同的分布, 即 $\mathcal{L}(\theta_n^\eta, V_n^\eta) = \mathcal{L}(\bar{\theta}_n^\eta, \bar{V}_n^\eta)$ 。此外, 考虑一个初值条件为 $\hat{\theta}_s^{s,u,v,\eta} = u$ 和 $\hat{V}_s^{s,u,v,\eta} = v$ 连续时间过程 $(\hat{\zeta}_t^{s,u,v,\eta}, \hat{Z}_t^{s,u,v,\eta})$:

$$\begin{aligned} d\hat{Z}_t^{s,u,v,\eta} &= -\eta(\gamma \hat{Z}_t^{s,u,v,\eta} + h(\hat{\zeta}_t^{s,u,v,\eta}))dt + \sqrt{2\gamma\eta\beta^{-1}}dB_t^\eta \\ d\hat{\zeta}_t^{s,u,v,\eta} &= \eta \hat{Z}_t^{s,u,v,\eta} dt \end{aligned} \tag{12}$$

定义 4.1 给定 $n \in \mathbb{N}$ 且对 $T := \lfloor 1/\eta \rfloor$, 定义过程

$$\bar{\zeta}_t^{\eta,n} = \hat{\zeta}_t^{nT, \bar{\theta}_{nT}^\eta, \bar{V}_{nT}^\eta, \eta}, \bar{Z}_t^{\eta,n} = \hat{Z}_t^{nT, \bar{\theta}_{nT}^\eta, \bar{V}_{nT}^\eta, \eta} \tag{13}$$

直观看来, 过程 $(\bar{\zeta}_t^{\eta,n}, \bar{Z}_t^{\eta,n}), t \geq nT$ 是一个开始于时间 nT 且初值为 $(\bar{\theta}_{nT}^\eta, \bar{V}_{nT}^\eta)$ 的欠阻尼 Langevin 过程。最后, 对 $t \in \mathbb{R}_+$ 定义一个含样本数据 $\{X_n\}_{n \in \mathbb{N}}$ 信息流和布朗运动 $\{B_t\}_{t \geq 0}$ 随机性的连续时间流 $\{\mathcal{H}_t\}_{t \geq 0}$:

$$\mathcal{H}_t := \mathcal{F}_\infty \vee \mathcal{G}_{\lfloor t \rfloor}, \mathcal{H}_t^* := \mathcal{G}_{\lfloor t \rfloor}^*$$

4.2. 引理准备

引理 4.2 若假设 2.1~2.5 成立。则对 $0 < \eta \leq \eta_{max}$,

$$\begin{aligned} \sup_{k \geq 0} \mathbb{E} |\theta_k^\eta|^2 &\leq C_\theta := 8(1-2\lambda)^{-1} \beta^{-1} \gamma^{-2} C_{\mathcal{V}} \\ \sup_{k \geq 0} \mathbb{E} |V_k^\eta|^2 &\leq C_v := 4(1-2\lambda)^{-1} \beta^{-1} C_{\mathcal{V}} \end{aligned}$$

其中 $C_{\mathcal{V}} := \int_{\mathbb{R}^{2d}} \mathcal{V}(\theta, v) \mu_0(d\theta, dv) + 4\lambda^{-1}(A_c + d)$ 。

证明 设对 $k = 0, 1, \dots$,

$$L(k) = \frac{\mathbb{E} \mathcal{V}(\theta_k^\eta, V_k^\eta)}{\beta} = \mathbb{E} \left[U(\theta_k^\eta) + \frac{\gamma^2}{4} (|\theta_k^\eta + \gamma^{-1} V_k^\eta|^2 + |\gamma^{-1} V_k^\eta|^2 - \lambda |\theta_k^\eta|^2) \right] \tag{14}$$

类似文献[15]中引理 3.2 的证明, 由(3)和注 2.6 的(6)得到

$$L(k+1) - L(k) \leq -\frac{\eta\gamma}{2} \mathbb{E} \langle \theta_k^\eta, h(\theta_k^\eta) \rangle - \frac{\eta\gamma}{2} \mathbb{E} |V_k^\eta|^2 - \frac{\eta\gamma^2\lambda}{2} \mathbb{E} \langle \theta_k^\eta, V_k^\eta \rangle + \eta\gamma\beta^{-1}d + M_k\eta^2 \tag{15}$$

其中

$$M_k = \left(\frac{L_1}{2} + \frac{\gamma^2}{4} + \frac{\gamma}{4} - \frac{\gamma^2\lambda}{4} \right) \mathbb{E} |V_k^\eta|^2 + \left(\frac{\tilde{L}_1}{2} + \frac{\gamma L_1^2}{2} \right) \mathbb{E} |\theta_k^\eta|^2 + \frac{\tilde{C}_1}{2} + \frac{\gamma h_0^2}{2} \tag{16}$$

且 $\tilde{L}_1 = 2L_1^2$, $\tilde{C}_1 = 4L_2^2\mathcal{M}_2(X) + 4H_*^2$ 。注意到对 $0 < \lambda \leq 1/4$, 由假设 2.1,

$$\mathcal{V}(\theta, v) \geq \max \left\{ \frac{\beta}{8}(1-2\lambda)\gamma^2|\theta|^2, \frac{\beta}{4}(1-2\lambda)|v|^2 \right\} \tag{17}$$

由(14)中 $L(k)$ 的定义及不等式 $\max\{a, b\} \geq (a+b)/2$, 得到

$$L(k) \geq \frac{1}{16}(1-2\lambda)\gamma^2\mathbb{E}|\theta_k^\eta|^2 + \frac{1}{8}(1-2\lambda)\mathbb{E}|V_k^\eta|^2 \tag{18}$$

因此, 由和, 得到

$$M_k \leq K_1L(k) + K_2$$

其中

$$K_1 = \max \left\{ \frac{\frac{\tilde{L}_1}{2} + \frac{\gamma L_1^2}{2}}{\frac{1}{16}(1-2\lambda)\gamma^2}, \frac{L_1 + \frac{\gamma^2}{4} + \frac{\gamma}{4} - \frac{\gamma^2\lambda}{4}}{\frac{1}{8}(1-2\lambda)} \right\}, K_2 = \frac{\tilde{C}_1}{2} + \frac{\gamma h_0^2}{2}$$

由参考文献[15]中的引理 B.5, 选择 $\lambda = \min\{1/4, a/(L_1 + 2L_1h_0 + \gamma^2/2)\}$, 可得到对所有 $x \in \mathbb{R}^d$, 有

$$\langle \theta, h(\theta) \rangle \geq 2\lambda \left(U(\theta) + \frac{\gamma^2}{4}|\theta|^2 \right) - \frac{2A_c}{\beta} \tag{19}$$

其中 $A_c = \beta(b + 2\lambda u_0 + 2\lambda L_1 h_0)/2$ 。故将(19)代入(15), 得到

$$L(k+1) - L(k) \leq -\eta\gamma\lambda\mathbb{E}U(\theta_k^\eta) - \frac{\eta\gamma^3\lambda}{4}\mathbb{E}|\theta_k^\eta|^2 + \eta A_c\gamma\beta^{-1} - \frac{\eta\gamma}{2}\mathbb{E}|V_k^\eta|^2 - \frac{\eta\gamma^2\lambda}{2}\mathbb{E}\langle \theta_k^\eta, V_k^\eta \rangle + \eta\gamma\beta^{-1}d + (K_1L(k) + K_2)\eta^2 \tag{20}$$

又 $0 < \lambda \leq 1/4$ 时, 由(14)得到

$$L(k) \leq \mathbb{E}U(\theta_k^\eta) + \frac{\gamma^2}{4}\mathbb{E}|\theta_k^\eta|^2 + \frac{1}{2\lambda}\mathbb{E}|V_k^\eta|^2 + \frac{\gamma}{2}\mathbb{E}\langle \theta_k^\eta, V_k^\eta \rangle \tag{21}$$

则对 $0 < \eta \leq \min\left\{\frac{K_3}{K_2}, \frac{\gamma\lambda}{2K_1}, \frac{2}{\gamma\lambda}\right\}$, 由(20)和(21), 得到

$$L(k+1) \leq \left(1 - \frac{\eta\gamma\lambda}{2}\right)L(k) + 2K_3\eta \leq L(0) + \frac{4}{\gamma\lambda}K_3 \tag{22}$$

其中 $K_3 := (A_c + d)\gamma\beta^{-1}$ 。结合(17), 最终得证。

引理 4.3 设 $0 < \eta \leq \eta_{\max}$, 且假设 2.1~2.5 成立, 则有

$$\sup_{k \in \mathbb{N}} \mathbb{E}[\mathcal{V}^2(\theta_k^\eta, V_k^\eta)] \leq \mathbb{E}[\mathcal{V}^2(\theta_0, v_0)] + \frac{4D}{\gamma\lambda}$$

其中 $D = \mathcal{O}(d^2)$ 为与 η 无关的常数。

证明 首先, 定义 $\Delta_k^1 = V_k^\lambda - \eta[\gamma V_k^\lambda + H(\theta_k^\eta, X_{k+1})]$, $\Delta_k^2 = \theta_k^\eta + \gamma^{-1}V_k^\eta - \eta\gamma^{-1}H(\theta_k^\eta, X_{k+1})$ 。类似参考文献[15]引理 3.4 和文献[6]引理 5.5 的证明, 得到

$$\begin{aligned} \frac{\mathcal{V}_{k+1}}{\beta} &\leq (1 - \eta\gamma\lambda)\frac{\mathcal{V}_k}{\beta} + \eta\langle V_k^\eta, h(\theta_k^\eta) - H(\theta_k^\eta, X_{k+1}) \rangle + \eta\gamma A_c\beta^{-1} \\ &\quad + \frac{1}{2}\eta\gamma\langle \theta_k^\eta, h(\theta_k^\eta) - H(\theta_k^\eta, X_{k+1}) \rangle + \eta\gamma\beta^{-1}|\xi_{k+1}|^2 + N_k\eta^2 + \Sigma_k \end{aligned} \tag{23}$$

其中

$$N_k = \left(\frac{L_1}{2} + \frac{\gamma^2}{4} - \frac{\gamma^2 \lambda}{4} + \frac{\gamma}{4} \right) |V_k^\eta|^2 + \left(1 + \frac{\gamma}{2} \right) L_1^2 |\theta_k^\eta|^2 + (2 + \gamma) L_2^2 |X_{k+1}|^2 + (2 + \gamma) H_*^2 \tag{24}$$

$$\Sigma_k = \frac{1}{2} \gamma^2 \sqrt{2\eta\gamma^{-1}\beta^{-1}} \langle \Delta_k^2, \xi_{k+1} \rangle + \frac{1}{2} \sqrt{2\eta\gamma\beta^{-1}} \langle \Delta_k^1, \xi_{k+1} \rangle$$

下面求(24)中第一式的上界, 由(17)和不等式 $\max\{a, b\} \geq (a+b)/2$, 有

$$N_k \leq \tilde{K}_1 \frac{\mathcal{V}_k}{\beta} + \tilde{K}_2 \tag{25}$$

其中

$$\tilde{K}_1 = \max \left\{ \frac{\left(1 + \frac{\gamma}{2} \right) L_1^2}{\frac{1}{16} (1 - 2\lambda) \gamma^2}, \frac{\frac{L_1}{2} + \frac{\gamma^2}{4} - \frac{\gamma^2 \lambda}{4} + \frac{\gamma}{4}}{\frac{1}{8} (1 - 2\lambda)} \right\}, \tilde{K}_2 = (2 + \gamma) L_2^2 |X_{k+1}|^2 + (2 + \gamma) H_*^2$$

将(25)代入(23), 对 $0 < \eta \leq \frac{\gamma\lambda}{2\tilde{K}_1}$, 整理得到

$$\mathcal{V}_{k+1} \leq \phi \mathcal{V}_k + \bar{K}_{k+1} \tag{26}$$

其中,

$$\phi := 1 - \frac{1}{2} \eta \gamma \lambda$$

$$\bar{K}_{k+1} := \eta \gamma A_c + \eta \beta \langle V_k^\eta, h(\theta_k^\eta) - H(\theta_k^\eta, X_{k+1}) \rangle + \frac{1}{2} \eta \gamma \beta \langle \theta_k^\eta, h(\theta_k^\eta) - H(\theta_k^\eta, X_{k+1}) \rangle$$

$$+ \eta \gamma |\xi_{k+1}|^2 + \eta^2 \beta \tilde{K}_2 + \beta \Sigma_k$$

(26)两边平方取条件期望, 注意到 $\mathbb{E}\Sigma_k = 0$, 由假设 2.3 的(4)和注 2.6, 得到

$$\begin{aligned} & \mathbb{E}[\mathcal{V}_{k+1}^2 | \mathcal{G}_{k+1}] \\ & \leq \phi^2 \mathcal{V}_k^2 + 2\phi \mathcal{V}_k (\eta \gamma A_c + \eta \gamma d + \eta^2 \beta \hat{K}_2) + \mathbb{E}[|\bar{K}_{k+1}|^2 | \mathcal{G}_{k+1}] \\ & \leq \phi^2 \mathcal{V}_k^2 + 2\phi \mathcal{V}_k (\eta \gamma A_c + \eta \gamma d + \eta^2 \beta \hat{K}_2) + 6\eta^2 \gamma^2 A_c^2 + \frac{3}{2} \eta^2 \beta^2 |V_k^\eta|^4 + 6\eta^2 \beta^2 \left(\frac{1}{4} \gamma + (2\gamma + 4) L_1^2 \right)^2 |\theta_k^\eta|^4 \\ & \quad + 6\eta^2 \gamma^2 d(d+2) + 6\eta^4 \beta^2 \hat{K}_2^2 + 6\eta^2 \beta^2 \left(\frac{1}{2} \gamma + 1 \right)^2 (4L_2^2 \mathcal{M}_2(X) + 2h_0^2 + 4H_*^2)^2 + \beta^2 \mathbb{E}[|\Sigma_k|^2 | \mathcal{G}_{k+1}] \\ & \leq \phi^2 \mathcal{V}_k^2 + 2\phi \eta \gamma \tilde{c}_1 + 2\phi \eta^2 \beta \mathcal{V}_k \hat{K}_2 + \eta^2 \tilde{c}_2 + \eta^2 \beta^2 \tilde{c}_3 |\theta_k^\eta|^4 + \frac{3}{2} \eta^2 \beta^2 |V_k^\eta|^4 \\ & \quad + \eta^2 \beta^2 \tilde{c}_4 + \eta^4 \beta^2 \hat{c}_4 + \beta^2 \mathbb{E}[|\Sigma_k|^2 | \mathcal{G}_{k+1}] \end{aligned} \tag{27}$$

其中 $\hat{K}_2 = (2 + \gamma) L_2^2 \mathcal{M}_2(X) + (2 + \gamma) H_*^2$ 且

$$\tilde{c}_1 = (A_c + d)\gamma, \quad \tilde{c}_2 = 6\gamma^2 (A_c^2 + d(d+2)), \quad \tilde{c}_3 = \frac{3}{4} \gamma^2 + 48(\gamma + 2)^2 L_1^4$$

$$\tilde{c}_4 = (\gamma + 2)^2 (72L_2^4 \mathbb{E}|X_0|^4 + 18h_0^4 + 72H_*^4), \quad \hat{c}_4 = 6\hat{K}_2^2$$

对于 $\mathbb{E}[|\Sigma_k|^2 | \mathcal{G}_{k+1}]$, 由注 2.6 的(6)和参考文献[15]引理 3.4 的证明, 可得到

$$\begin{aligned} \mathbb{E}\left[|\Sigma_k|^2 \mid \mathcal{G}_{k+1}\right] &\leq 2\eta\gamma\beta^{-1}d\left(5|V_k^\eta|^2 + (3\gamma^2 + 15L_1^2)|\theta_k^\eta|^2 + 15L_2^2\mathcal{M}_2(X) + 15H_*^2\right) \\ &\leq \eta\tilde{c}_5\frac{\mathcal{V}_k}{\beta^2} + \eta\frac{\tilde{c}_6}{\beta} \end{aligned} \tag{28}$$

其中

$$\tilde{c}_5 = \max\left\{\frac{10\gamma d}{\frac{1}{8}(1-2\lambda)}, \frac{2\gamma d(3\gamma^2 + 15L_1^2)}{\frac{1}{16}(1-2\lambda)\gamma^2}\right\}, \quad \tilde{c}_6 = 30\gamma d(L_2^2\mathcal{M}_2(X) + H_*^2)$$

将(28)代入(27), 利用 $\eta \leq 1$, 结合(17), 得到

$$\begin{aligned} \mathbb{E}\left[\mathcal{V}_{k+1}^2 \mid \mathcal{G}_{k+1}\right] &\leq \phi^2\mathcal{V}_k^2 + \eta(2\phi\tilde{c}_1 + \tilde{c}_5)\mathcal{V}_k + 2\phi\eta^2\beta\mathcal{V}_k\hat{K}_2 + \eta^2\tilde{c}_2 + \eta^2\beta^2\tilde{c}_3|\theta_k^\eta|^4 \\ &\quad + \frac{3}{2}\eta^2\beta^2|V_k^\eta|^4 + \eta^2\beta^2(\tilde{c}_4 + \hat{c}_4) + \eta\beta\tilde{c}_6 \\ &\leq (\phi^2 + \eta^2K_4)\mathcal{V}_k^2 + \eta(2\phi\tilde{c}_1 + \tilde{c}_5)\mathcal{V}_k + 2\phi\eta^2\beta\mathcal{V}_k\hat{K}_2 \\ &\quad + \eta^2\tilde{c}_2 + \eta^2\beta^2(\tilde{c}_4 + \hat{c}_4) + \eta\beta\tilde{c}_6 \end{aligned} \tag{29}$$

其中

$$K_4 = \max\left\{\frac{\tilde{c}_3}{\frac{1}{128}(1-2\lambda)^2\gamma^4}, \frac{\frac{3}{2}}{\frac{1}{32}(1-2\lambda)^2}\right\}$$

对(29)两边取期望, 由于 $0 < \phi \leq 1$, 有

$$\mathbb{E}\mathcal{V}_{k+1}^2 \leq \left(1 - \frac{\eta\lambda\gamma}{2} + \eta^2K_4\right)\mathbb{E}\mathcal{V}_k^2 + \eta\tilde{c}_7\mathbb{E}\mathcal{V}_k + 2\eta^2\beta\hat{K}_2\mathbb{E}\mathcal{V}_k + \eta^2\beta\tilde{c}_2 + \eta^2\beta^2(\tilde{c}_4 + \hat{c}_4) + \eta\beta\tilde{c}_6$$

其中 $\tilde{c}_7 = 2\tilde{c}_1 + \tilde{c}_5$ 。由引理 4.2 的证明过程中的(22)和 $\eta \leq 1$, 当 $\eta \leq \lambda\gamma/(4K_4)$ 时, 递归计算得到

$$\begin{aligned} \mathbb{E}\mathcal{V}_{k+1}^2 &\leq \left(1 - \frac{\eta\lambda\gamma}{2} + \eta^2K_4\right)\mathbb{E}\mathcal{V}_k^2 + D\eta \\ &\leq \mathbb{E}\mathcal{V}_0^2 + \frac{4D}{\lambda\gamma} \end{aligned}$$

其中 $D = \tilde{c}_7\tilde{c}_8 + (2\hat{K}_2\tilde{c}_8 + \tilde{c}_2 + \tilde{c}_6)\beta + (\hat{c}_4 + \tilde{c}_4)\beta^2$, $\tilde{c}_8 := \mathbb{E}\mathcal{V}_0 + 4(A_c + d)/\lambda$ 。

下面给出参考文献[8]的引理 3.15 和引理 3.16。

引理 4.4 对每个 $n \in \mathbb{N}$, 存在一个可测函数 $h: \Omega \times [nT, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ 使得对任意 $t \geq nT$ 和 $\theta \in \mathbb{R}^d$, $h_{nT,t}(\theta)(\omega)$ 是 $\mathbb{E}\left[H(\theta, X_{\lfloor t \rfloor+1}) \mid \mathcal{H}_{nT}\right]$ 的一个版本, 且对几乎每个 $\omega \in \Omega$, $\theta \rightarrow h_{nT,t}(\theta)(\omega)$ 是连续的。

引理 4.5 若假设 2.2 和假设 2.3 成立, 且设 $p \geq 1$, 则

$$\sup_{n \in \mathbb{N}} \mathbb{E}^{1/p} \left[\left(\sum_{k=nT}^{\infty} \sup_{\theta \in \mathbb{R}^d} \|h_{k,nT}(\theta) - h(\theta)\| \right)^p \right] \leq 2L_2\Gamma_p^0(X).$$

4.3. 主要结果的证明

为证明主要结果, 考虑将证明目标分解成如下形式:

$$W_2(\mathcal{L}(\theta_t^\eta, V_t^\eta), \bar{\pi}_\beta) \leq W_2(\mathcal{L}(\bar{\theta}_t^\eta, \bar{V}_t^\eta), \mathcal{L}(\bar{\zeta}_t^{\eta,n}, \bar{Z}_t^{\eta,n})) + W_2(\mathcal{L}(\bar{\zeta}_t^{\eta,n}, \bar{Z}_t^{\eta,n}), \mathcal{L}(\zeta_t^\eta, Z_t^\eta)) + W_2(\mathcal{L}(\zeta_t^\eta, Z_t^\eta), \bar{\pi}_\beta) \tag{30}$$

将证明目标转化为控制(30)右边三项。对给定步长 $\eta > 0$ ，将时间轴划分为长度为 $T = \lfloor 1/\eta \rfloor$ 的区间。对(30)右边第一项，将比较连续插值过程的分布 $\mathcal{L}(\bar{\theta}_t^\eta, \bar{V}_t^\eta)$ 与区间上由(13)定义过程的分布之间的距离，二者在区间起点上具有相同初始分布。对(30)右式第二、三项的控制主要依赖于[7]中收缩估计的结果。

定理 4.6 若假设 2.1~2.5 成立，且 $0 < \eta \leq \eta_{\max}$ ，则，

$$W_2(\mathcal{L}(\bar{\theta}_t^\eta, \bar{V}_t^\eta), \mathcal{L}(\bar{\zeta}_t^{\eta,n}, \bar{Z}_t^{\eta,n})) \leq C_1 \eta^{1/2}$$

其中 $C_1 = \mathcal{O}(d^{1/2})$ 是有限常数，具体形式在证明中给出。

证明 注意到由 W_2 的定义，

$$W_2(\mathcal{L}(\bar{\theta}_t^\eta, \bar{V}_t^\eta), \mathcal{L}(\bar{\zeta}_t^{\eta,n}, \bar{Z}_t^{\eta,n})) \leq \mathbb{E}^{1/2} |\bar{\theta}_t^\eta - \bar{\zeta}_t^{\eta,n}|^2 + \mathbb{E}^{1/2} |\bar{V}_t^\eta - \bar{Z}_t^{\eta,n}|^2 \tag{31}$$

首先求(31)右式第一项的上界。由过程 $\bar{\theta}_t^\eta$ 和 $\bar{\zeta}_t^{\eta,n}$ 的定义，有

$$|\bar{\theta}_t^\eta - \bar{\zeta}_t^{\eta,n}| \leq \eta \int_{nT}^t |\bar{V}_{\lfloor s \rfloor}^\eta - \bar{Z}_s^{\eta,n}| ds$$

两边平方取期望，由于被积元素非负，因此

$$\sup_{nT \leq u \leq t} \mathbb{E} |\bar{\theta}_u^\eta - \bar{\zeta}_u^{\eta,n}|^2 \leq \eta \int_{nT}^t \mathbb{E} |\bar{V}_{\lfloor s \rfloor}^\eta - \bar{Z}_s^{\eta,n}|^2 ds \tag{32}$$

下面求 $\mathbb{E} |\bar{V}_{\lfloor t \rfloor}^\eta - \bar{Z}_t^{\eta,n}|^2$ 的上界。对 $t \in [nT, (n+1)T)$ ，由(11)~(13)及假设 2.4，得到

$$\begin{aligned} |\bar{V}_{\lfloor t \rfloor}^\eta - \bar{Z}_t^{\eta,n}| &\leq |\bar{V}_{\lfloor t \rfloor}^\eta - \bar{V}_t^\eta| + |\bar{V}_t^\eta - \bar{Z}_t^{\eta,n}| \\ &\leq |\bar{V}_{\lfloor t \rfloor}^\eta - \bar{V}_t^\eta| + \eta \gamma \int_{nT}^t |\bar{V}_{\lfloor s \rfloor}^\eta - \bar{Z}_s^{\eta,n}| ds + \eta \int_{nT}^t |H(\bar{\theta}_{\lfloor s \rfloor}^\eta, X_{\lfloor s \rfloor+1}) - H(\bar{\zeta}_s^{\eta,n}, X_{\lfloor s \rfloor+1})| ds \\ &\quad + \eta \left| \int_{nT}^t H(\bar{\zeta}_s^{\eta,n}, X_{\lfloor s \rfloor+1}) - h_{k,nT}(\bar{\zeta}_s^{\eta,n}) ds \right| + \eta \left| \int_{nT}^t |h_{k,nT}(\bar{\zeta}_s^{\eta,n}) - h(\bar{\zeta}_s^{\eta,n})| ds \right| \\ &\leq |\bar{V}_{\lfloor t \rfloor}^\eta - \bar{V}_t^\eta| + \eta \gamma \int_{nT}^t |\bar{V}_{\lfloor s \rfloor}^\eta - \bar{Z}_s^{\eta,n}| ds + \eta L_1 \int_{nT}^t |\bar{\theta}_{\lfloor s \rfloor}^\eta - \bar{\zeta}_s^{\eta,n}| ds \\ &\quad + \eta \left| \int_{nT}^t H(\bar{\zeta}_s^{\eta,n}, X_{\lfloor s \rfloor+1}) - h_{k,nT}(\bar{\zeta}_s^{\eta,n}) ds \right| + \eta \left| \int_{nT}^t |h_{k,nT}(\bar{\zeta}_s^{\eta,n}) - h(\bar{\zeta}_s^{\eta,n})| ds \right| \end{aligned}$$

由文献[15]引理 B.4，得到 $\mathbb{E} |\bar{V}_{\lfloor t \rfloor}^\eta - \bar{V}_t^\eta|^2 \leq \sigma_V \eta$ ，其中 $\sigma_V = 4\eta\gamma^2 C_v + 4\eta(\tilde{L}_1 C_\theta + \tilde{C}_1) + 4\gamma\beta^{-1}d$ 。故利用 $\eta T \leq 1$ ，得到

$$\begin{aligned} \mathbb{E} |\bar{V}_{\lfloor t \rfloor}^\eta - \bar{Z}_t^{\eta,n}|^2 &\leq 4\sigma_V \eta + 4\eta\gamma^2 \int_{nT}^t \mathbb{E} |\bar{V}_{\lfloor s \rfloor}^\eta - \bar{Z}_s^{\eta,n}|^2 ds + 6\eta L_1^2 \int_{nT}^t \mathbb{E} |\bar{\theta}_{\lfloor s \rfloor}^\eta - \bar{\zeta}_s^{\eta,n}|^2 ds \\ &\quad + 6\eta^2 \mathbb{E} \left| \int_{nT}^t H(\bar{\zeta}_s^{\eta,n}, X_{\lfloor s \rfloor+1}) - h_{k,nT}(\bar{\zeta}_s^{\eta,n}) ds \right|^2 + 6\eta \int_{nT}^t \mathbb{E} |h_{k,nT}(\bar{\zeta}_s^{\eta,n}) - h(\bar{\zeta}_s^{\eta,n})|^2 ds \end{aligned}$$

应用 Grönwall 不等式，注意 $\eta T \leq 1$ ，对 $t \in [nT, (n+1)T)$ ，得到

$$\begin{aligned} \mathbb{E} |\bar{V}_{\lfloor t \rfloor}^\eta - \bar{Z}_t^{\eta,n}|^2 &\leq 4c_1 \sigma_V \eta + 6c_1 \eta L_1^2 \int_{nT}^t \mathbb{E} |\bar{\theta}_{\lfloor s \rfloor}^\eta - \bar{\zeta}_s^{\eta,n}|^2 ds + 6c_1 \eta^2 \mathbb{E} \left| \int_{nT}^t H(\bar{\zeta}_s^{\eta,n}, X_{\lfloor s \rfloor+1}) - h_{k,nT}(\bar{\zeta}_s^{\eta,n}) ds \right|^2 \\ &\quad + 6c_1 \eta \int_{nT}^t \mathbb{E} |h_{k,nT}(\bar{\zeta}_s^{\eta,n}) - h(\bar{\zeta}_s^{\eta,n})|^2 ds \end{aligned} \tag{33}$$

其中 $c_1 = \exp(4\gamma^2)$ 。结合(32)和(33)，且由 $\eta T \leq 1$ ，得到

$$\begin{aligned}
 & \sup_{nT \leq u \leq t} \mathbb{E} \left| \bar{\theta}_u^\eta - \bar{\zeta}_u^{\eta,n} \right|^2 \\
 & \leq 6c_1 \eta^2 L_1^2 \int_{nT}^t \int_{nT}^s \mathbb{E} \left| \bar{\theta}_{[s']}^\eta - \bar{\zeta}_{s'}^{\eta,n} \right|^2 ds' ds + 6c_1 \eta^3 \int_{nT}^t \mathbb{E} \left| \int_{nT}^s H \left(\bar{\zeta}_{s'}^{\eta,n}, X_{[s'+1]} \right) - h_{k,nT} \left(\bar{\zeta}_{s'}^{\eta,n} \right) ds' \right|^2 ds \\
 & \quad + 6c_1 \eta^2 \int_{nT}^t \int_{nT}^s \mathbb{E} \left| h_{k,nT} \left(\bar{\zeta}_{s'}^{\eta,n} \right) - h \left(\bar{\zeta}_{s'}^{\eta,n} \right) \right|^2 ds' ds + 4c_1 \sigma_V \eta \\
 & \leq 6c_1 \eta L_1^2 \sup_{nT \leq s \leq t} \int_{nT}^s \mathbb{E} \left| \bar{\theta}_{[s']}^\eta - \bar{\zeta}_{s'}^{\eta,n} \right|^2 ds' + 6c_1 \eta^2 \mathbb{E} \left[\sup_{nT \leq s \leq (n+1)T} \left| \int_{nT}^s H \left(\bar{\zeta}_{s'}^{\eta,n}, X_{[s'+1]} \right) - h_{k,nT} \left(\bar{\zeta}_{s'}^{\eta,n} \right) ds' \right|^2 \right] \\
 & \quad + 6c_1 \eta \int_{nT}^\infty \mathbb{E} \left[\sup_{\theta \in \mathbb{R}^d} \left| h_{k,nT}(\theta) - h(\theta) \right|^2 \right] ds' + 4c_1 \sigma_V \eta
 \end{aligned} \tag{34}$$

首先, 求(34)右式第一项的上界。由于被积项非负, 故

$$\sup_{nT \leq s \leq t} \int_{nT}^s \mathbb{E} \left| \bar{\theta}_{[s']}^\eta - \bar{\zeta}_{s'}^{\eta,n} \right|^2 ds' \leq \int_{nT}^t \sup_{nT \leq u \leq s'} \mathbb{E} \left| \bar{\theta}_u^\eta - \bar{\zeta}_u^{\eta,n} \right|^2 ds' \tag{35}$$

然后, 求(34)中第二项的上界。对任意 $j \in \mathbb{N}$, 引入 \mathcal{H}_{nT} 可测事件:

$$Z_j^{nT} := \left\{ j \leq \sup_{nT \leq s < (n+1)T} \left| \bar{\zeta}_s^{\eta,n} \right| < j+1 \right\} \tag{36}$$

对 $s \in [nT, (n+1)T)$, 定义两个辅助随机过程:

$$\tilde{W}_s^\theta := H \left(\theta, X_{[s+1]} \right) 1_{Z_j^{nT}}, \quad \hat{W}_s := H \left(\bar{\zeta}_s^{\eta,n}, X_{[s+1]} \right) 1_{Z_j^{nT}}.$$

由引理 3.3, 对任意固定的 θ , 随机过程 \tilde{W}_s^θ 满足

$$M_p^{nT} \left(\tilde{W}^\theta \right) \leq L_1 j + L_2 M_p^{nT} (X) + H_*, \quad \Gamma_p^{nT} \left(\tilde{W}^\theta \right) \leq 2L_2 \Gamma_p^{nT} (X)$$

因此, 由引理 3.4, 我们可以将 \mathcal{H}_{nT} 可测随机过程 $\bar{\zeta}_s^{\eta,n}$ 插入随机过程 \tilde{W}_s^θ 中, 即得到随机过程 \hat{W}_s 满足

$$M_p^{nT} \left(\hat{W} \right) \leq L_1 j + L_2 M_p^{nT} (X) + H_*, \quad \Gamma_p^{nT} \left(\hat{W} \right) \leq 2L_2 \Gamma_p^{nT} (X)$$

对任意 $j \in \mathbb{N}$, $s \geq nT$, 定义随机过程:

$$W_s^{nT,j} := \left(H \left(\bar{\zeta}_s^{\eta,n}, X_{[s+1]} \right) - h_{nT,s} \left(\bar{\zeta}_s^{\eta,n} \right) \right) 1_{Z_j^{nT}}$$

记 $W_s = W_s^{nT,j}$, 显然对 $s \geq nT$ 有 $\mathbb{E} \left[W_s^{nT,j} | \mathcal{H}_{nT} \right] = 0$ 。下面估计定理 3.2 中的量 $M_p(W)$, $\Gamma_p(W)$ 。注意到,

$$W_s^{nT,j} = \hat{W}_s - \mathbb{E} \left[\hat{W}_s | \mathcal{H}_{nT} \right] \tag{37}$$

故由 Minkowski 不等式和(8), 得到

$$\begin{aligned}
 M_p^{nT} (W) & \leq 2 \sum_{i=1}^d \sup_{nT \leq s < (n+1)T} \mathbb{E}^{1/p} \left[\left| \hat{W}_s^i \right|^p \mathcal{G}_{nT} \right] \\
 & \leq 2 \left(L_1 j + L_2 M_p^{nT} (X) + H_* \right)
 \end{aligned} \tag{38}$$

且将(37)代入(8), 易得

$$\Gamma_p^{nT} (W) = \Gamma_p^{nT} \left(\hat{W} \right) \leq 2K_2 \Gamma_p^{nT} (X) \tag{39}$$

应用定理 3.2, 取 $r = 3$, 则 $C'(3) \leq 10$, 且由(36)、(38)和(39), 得到

$$\begin{aligned}
 & \mathbb{E}^{1/2} \left[\sup_{nT \leq s < (n+1)T} \left| \int_{nT}^s \left[H(\bar{\zeta}_s^{\eta,n}, X_{\lfloor s \rfloor+1}) - h_{nT,s}(\bar{\zeta}_s^{\eta,n}) \right] ds \right|^2 1_{Z_i^{nT}} | \mathcal{H}_{nT} \right] \\
 & \leq \mathbb{E}^{1/3} \left[\sup_{nT \leq s < (n+1)T} \left| \int_{nT}^s \left[H(\bar{\zeta}_s^{\eta,n}, X_{\lfloor s \rfloor+1}) - h_{nT,s}(\bar{\zeta}_s^{\eta,n}) \right] ds \right|^3 1_{Z_i^{nT}} | \mathcal{H}_{nT} \right] \\
 & \leq 2C'(3) \sqrt{T} \left[L_1 j + L_2 M_3^{nT}(X) + L_2 \Gamma_3^{nT}(X) + H_* \right] 1_{Z_i^{nT}} \\
 & \leq 20\sqrt{T} \left[L_1 \left(1 + \sup_{nT \leq s < (n+1)T} |\bar{\zeta}_s^{\eta,n}| \right) + L_2 M_3^{nT}(X) + L_2 \Gamma_3^{nT}(X) + H_* \right] 1_{Z_i^{nT}}
 \end{aligned} \tag{40}$$

由文献[15]引理 3.3, 有

$$\sup_{n \in \mathbb{N}} \sup_{t \in (nT, (n+1)T)} \mathbb{E} |\bar{\zeta}_t^{\eta,n}|^2 \leq C_\zeta := 8(1-2\lambda)^{-1} \beta^{-1} \gamma^{-2} \left(\int_{\mathbb{R}^{2d}} \mathcal{V}(\theta, v) \mu_0(d\theta, dv) + 5\lambda^{-1}(d + A_c) \right)$$

故对(40)两边取期望, 结合 Minkowski 不等式, 得到

$$\begin{aligned}
 & \mathbb{E} \left[\sup_{nT \leq s < (n+1)T} \left| \int_{nT}^s \left[H(\bar{\zeta}_{s'}^{\eta,n}, X_{\lfloor s' \rfloor+1}) - h_{nT,s'}(\bar{\zeta}_{s'}^{\eta,n}) \right] ds' \right|^2 \right] \\
 & \leq 400T \left[L_1 (C_\zeta^{1/2} + 1) + L_2 \mathbb{E}^{1/2} \left[(M_3^{nT})^2 \right] + L_2 \mathbb{E}^{1/2} \left[(\Gamma_3^{nT})^2 \right] + H_* \right]^2
 \end{aligned} \tag{41}$$

最后, 求(34)中的第三项的上界。由引理 4.5 得到

$$\mathbb{E} \left[\int_{nT}^\infty |h_{k,nT}(\bar{\zeta}_{s'}^{\eta,n}) - h(\bar{\zeta}_{s'}^{\eta,n})|^2 ds' \right] \leq 4L_2^2 (\Gamma_2^0(X))^2 \tag{42}$$

由(34)、(35)、(41)和(42), 且 $\eta T \leq 1$, 最终得到

$$\sup_{nT \leq u \leq t} \mathbb{E} |\bar{\theta}_u^\eta - \bar{\zeta}_u^{\eta,n}|^2 \leq 4c_1 \sigma_v \eta + 6c_1 \eta L_1^2 \int_{nT}^t \sup_{nT \leq u \leq s'} \mathbb{E} |\bar{\theta}_u^\eta - \bar{\zeta}_u^{\eta,n}|^2 ds' + \eta(\sigma_1 + \sigma_2) \tag{43}$$

其中

$$\begin{aligned}
 \sigma_1 &= 2400c_1 \left[L_1 (C_\zeta^{1/2} + 1) + L_2 \mathbb{E}^{1/2} \left[(M_3^{nT})^2 \right] + L_2 \mathbb{E}^{1/2} \left[(\Gamma_3^{nT})^2 \right] + H_* \right]^2 \\
 \sigma_2 &= 12c_1 L_2 \Gamma_2^0(X)
 \end{aligned}$$

再次应用 Grönwall 不等式, 且由 $\eta T \leq 1$, 得到

$$\sup_{nT \leq u \leq t} \mathbb{E} |\bar{\theta}_u^\eta - \bar{\zeta}_u^{\eta,n}|^2 \leq \exp(6c_1 L_1^2) (4c_1 \sigma_v + \sigma_1 + \sigma_2) \eta \tag{44}$$

故

$$\mathbb{E}^{1/2} |\bar{\theta}_u^\eta - \bar{\zeta}_u^{\eta,n}|^2 \leq C_{1,1} \eta^{1/2} \tag{45}$$

其中 $C_{1,1} := \sqrt{\exp(6c_1 L_1^2) (4c_1 \sigma_v + \sigma_1 + \sigma_2)}$ 。注意到, $\sigma_v = \mathcal{O}(d)$ 且 $\sigma_1 = \mathcal{O}(d)$, 故 $C_{1,1} = \mathcal{O}(d^{1/2})$ 。

下面求(31)右式第二项的上界。由(11)中 \bar{V}_t^η 的定义和(13)中 $\bar{Z}_t^{\eta,n}$ 的定义, 有

$$\begin{aligned}
 |\bar{V}_t^\eta - \bar{Z}_t^{\eta,n}| & \leq \left| \eta \gamma \int_{nT}^t [\bar{V}_{\lfloor s \rfloor}^\eta - \bar{Z}_s^{\eta,n}] ds \right| + \eta \left| \int_{nT}^t \left[H(\bar{\theta}_{\lfloor s \rfloor}^\eta, X_{\lfloor s \rfloor+1}) - h(\bar{\zeta}_s^{\eta,n}) \right] ds \right| \\
 & \leq \eta \gamma \int_{nT}^t |\bar{V}_{\lfloor s \rfloor}^\eta - \bar{Z}_s^{\eta,n}| ds + \eta \int_{nT}^t \left| H(\bar{\theta}_{\lfloor s \rfloor}^\eta, X_{\lfloor s \rfloor+1}) - H(\bar{\zeta}_s^{\eta,n}, X_{\lfloor s \rfloor+1}) \right| ds \\
 & \quad + \eta \left| \int_{nT}^t H(\bar{\zeta}_s^{\eta,n}, X_{\lfloor s \rfloor+1}) - h_{k,nT}(\bar{\zeta}_s^{\eta,n}) ds \right| + \eta \int_{nT}^t |h_{k,nT}(\bar{\zeta}_s^{\eta,n}) - h(\bar{\zeta}_s^{\eta,n})| ds
 \end{aligned}$$

类似地, 由前面的讨论及 $\eta T \leq 1$, 得到

$$\begin{aligned} \mathbb{E}|\bar{V}_t^\eta - \bar{Z}_t^{\eta,n}|^2 &\leq 2\eta\gamma^2 \int_{nT}^t \mathbb{E}|\bar{V}_{[s]}^\eta - \bar{Z}_s^{\eta,n}|^2 ds + 6\eta \int_{nT}^t \mathbb{E} \left| H(\bar{\theta}_{[s]}^\eta, X_{[s]}^\eta) - H(\bar{\zeta}_s^{\eta,n}, X_{[s]}^\eta) \right|^2 ds \\ &\quad + 6\eta^2 \mathbb{E} \left| \int_{nT}^t H(\bar{\zeta}_s^{\eta,n}, X_{[s]}^\eta) - h_{k,nT}(\bar{\zeta}_s^{\eta,n}) ds \right|^2 + 6\eta \int_{nT}^t \mathbb{E} \left| h_{k,nT}(\bar{\zeta}_s^{\eta,n}) - h(\bar{\zeta}_s^{\eta,n}) \right|^2 ds \\ &\leq 2\eta\gamma^2 \int_{nT}^t \mathbb{E}|\bar{V}_{[s]}^\eta - \bar{Z}_s^{\eta,n}|^2 ds + 6\eta L_1^2 \int_{nT}^t \mathbb{E}|\bar{\theta}_{[s]}^\eta - \bar{\zeta}_s^{\eta,n}|^2 ds \\ &\quad + 6\eta^2 \mathbb{E} \left| \int_{nT}^t H(\bar{\zeta}_s^{\eta,n}, X_{[s]}^\eta) - h_{k,nT}(\bar{\zeta}_s^{\eta,n}) ds \right|^2 + 6\eta \int_{nT}^t \mathbb{E} \left| h_{k,nT}(\bar{\zeta}_s^{\eta,n}) - h(\bar{\zeta}_s^{\eta,n}) \right|^2 ds \\ &\leq 2\eta\gamma^2 \int_{nT}^t \mathbb{E}|\bar{V}_{[s]}^\eta - \bar{Z}_s^{\eta,n}|^2 ds + 6\eta L_1^2 \int_{nT}^t \mathbb{E}|\bar{\theta}_{[s]}^\eta - \bar{\zeta}_s^{\eta,n}|^2 ds + (\sigma_1 + \sigma_2)\eta \end{aligned} \tag{46}$$

(46)右边关于 t 是递增的, 故由(44)和 $\eta T \leq 1$, 得到

$$\begin{aligned} \sup_{nT \leq u \leq t} \mathbb{E}|\bar{V}_u^\eta - \bar{Z}_u^{\eta,n}|^2 &\leq 2\eta\gamma^2 \int_{nT}^t \sup_{nT \leq u \leq s} \mathbb{E}|\bar{V}_u^\eta - \bar{Z}_u^{\eta,n}|^2 ds + 6\eta L_1^2 \int_{nT}^t \sup_{nT \leq u \leq s} \mathbb{E}|\bar{\theta}_u^\eta - \bar{\zeta}_u^{\eta,n}|^2 ds + (\sigma_1 + \sigma_2)\eta \\ &\leq 2\eta\gamma^2 \int_{nT}^t \sup_{nT \leq u \leq s} \mathbb{E}|\bar{V}_u^\eta - \bar{Z}_u^{\eta,n}|^2 ds + (6L_1^2 \exp(6c_1 L_1^2)(4c_1 \sigma_V + \sigma_1 + \sigma_2) + \sigma_1 + \sigma_2)\eta \end{aligned}$$

应用 Grönwall 不等式, 且由 $\eta T \leq 1$, 得到

$$\sup_{nT \leq u \leq t} \mathbb{E}|\bar{V}_u^\eta - \bar{Z}_u^{\eta,n}|^2 \leq \exp(2\gamma^2) \left(6L_1^2 \exp(6c_1 L_1^2)(4c_1 \sigma_V + \sigma_1 + \sigma_2) + \sigma_1 + \sigma_2 \right) \eta$$

故

$$\mathbb{E}^{1/2} |\bar{\theta}_u^\eta - \bar{\zeta}_u^{\eta,n}|^2 \leq C_{1,1} \eta^{1/2} \tag{47}$$

其中 $C_{1,2} := \sqrt{\exp(2\gamma^2) \left[(6L_1^2 \exp(6c_1 L_1^2)(4c_1 \sigma_V + \sigma_1 + \sigma_2) + \sigma_1 + \sigma_2) \right]}$, 且 $C_{1,2} = \mathcal{O}(d^{1/2})$ 。

最后, 由(31)、(45)和(47), 得到

$$W_2(\mathcal{L}(\bar{\theta}_t^\eta, \bar{V}_t^\eta), \mathcal{L}(\bar{\zeta}_t^{\eta,n}, \bar{Z}_t^{\eta,n})) \leq C_1 \eta^{1/2}$$

其中 $C_1 = C_{1,1} + C_{1,2}$, 且 $C_1 = \mathcal{O}(d^{1/2})$ 。

定理 4.7 (文献[7]推论 2.6) 若假设 2.1~2.5 成立, 且欠阻尼 Langevin 随机微分方程 (θ_t, V_t) 和 (θ'_t, V'_t) 的分布分别开始于 $(\theta_0, V_0) \sim \mu_0$ 和 $(\theta'_0, V'_0) \sim \nu_0$, 则, 存在常数 $\dot{c}, \dot{C} \in (0, \infty)$ 使得

$$W_2(\mathcal{L}(\theta_t, V_t), \mathcal{L}(\theta'_t, V'_t)) \leq \sqrt{\dot{C}} e^{-\dot{c}t/2} \sqrt{\mathcal{W}_\rho(\mu_0, \nu_0)}$$

其中常数 $\dot{c} = \mathcal{O}(e^{-d})$, $\dot{C} = \mathcal{O}(e^d)$ 分别由文献[7]定理 2.3 和推论 2.6 给出, \mathcal{W}_ρ 由文献[7]中式(2.12)定义。

定理 4.8 若假设 2.1~2.5 成立, 且 $0 < \eta \leq \eta_{\max}$, 则

$$W_2(\mathcal{L}(\bar{\zeta}_t^{\eta,n}, \bar{Z}_t^{\eta,n}), \mathcal{L}(\zeta_t^\eta, Z_t^\eta)) \leq C_2 \eta^{1/4}$$

其中 $C_2 = \mathcal{O}(e^d d^{3/2})$, 具体形式在证明中给出。

证明类似文献[15]定理 4.2 的证明, 由定理 4.7 和文献[7]引理 5.4, 得到

$$\begin{aligned} &W_2(\mathcal{L}(\bar{\zeta}_t^{\eta,n}, \bar{Z}_t^{\eta,n}), \mathcal{L}(\zeta_t^\eta, Z_t^\eta)) \\ &\leq \sqrt{3 \max\{1 + \alpha, \gamma^{-1}\} \dot{C} \sum_{k=1}^n e^{-\eta c(t-kT)/2} \sqrt{1 + \varepsilon \mathbb{E}^{1/2} [\nu^2(\theta_{kT}^\eta, V_{kT}^\eta)] + \varepsilon \mathbb{E}^{1/2} [\nu^2(\bar{\zeta}_{kT}^{\eta,(k-1)T}, \bar{Z}_{kT}^{\eta,(k-1)T})]}} \\ &\quad \times \sqrt{W_2(\mathcal{L}(\theta_{kT}^\eta, V_{kT}^\eta), \mathcal{L}(\bar{\zeta}_{kT}^{\eta,(k-1)T}, \bar{Z}_{kT}^{\eta,(k-1)T}))} \end{aligned}$$

其中常数 α 由文献[7]定理 2.3 给出。

注意 $(\bar{\zeta}_k^{\eta,k-1}, \bar{Z}_k^{\eta,k-1})$ 初值为 $(\bar{\theta}_{(k-1)T}^\eta, \bar{V}_{(k-1)T}^\eta)$, 且 $\mathcal{L}(\theta_{nT}^\eta, V_{nT}^\eta) = \mathcal{L}(\bar{\theta}_{nT}^\eta, \bar{V}_{nT}^\eta)$, 因此由引理 4.3, 得到

$$\mathbb{E}[\mathcal{V}^2(\bar{\zeta}_k^{\eta,k-1}, \bar{Z}_k^{\eta,k-1})] \leq \mathbb{E}[\mathcal{V}^2(\theta_{nT}^\eta, V_{nT}^\eta)] + \frac{4D}{\gamma\lambda} \leq \mathbb{E}[\mathcal{V}^2(\theta_0, V_0)] + \frac{8D}{\gamma\lambda}$$

故由引理 4.3 和定理 4.6, 得到

$$W_2(\mathcal{L}(\bar{\zeta}_t^{\eta,n}, \bar{Z}_t^{\eta,n}), \mathcal{L}(\zeta_t^\eta, Z_t^\eta)) \leq \sqrt{3 \max\{1 + \alpha, \gamma^{-1}\} \dot{C}C_1(1 + \varepsilon(C_{2,1} + C_{2,2}))} \sum_{k=1}^n e^{-\eta\dot{c}(t-kT)/2} \eta^{1/4}$$

其中

$$C_{2,1} := \mathbb{E}^{1/2}[\mathcal{V}^2(\theta_0, v_0)] + 2\sqrt{\frac{D}{\gamma\lambda}}, \quad C_{2,2} := \mathbb{E}^{1/2}[\mathcal{V}^2(\theta_0, v_0)] + 2\sqrt{\frac{2D}{\gamma\lambda}}$$

进一步计算, 得到

$$W_2(\mathcal{L}(\bar{\zeta}_t^{\eta,n}, \bar{Z}_t^{\eta,n}), \mathcal{L}(\zeta_t^\eta, Z_t^\eta)) \leq \sqrt{3 \max\{1 + \alpha, \gamma^{-1}\} \dot{C}C_1(1 + \varepsilon(C_{2,1} + C_{2,2}))} \frac{e^{-\eta\dot{c}(t-nT)/2}}{1 - e^{-\eta\dot{c}T/2}} \eta^{1/4} \leq C_2 \eta^{1/4}$$

其中 $C_2 := \sqrt{3 \max\{1 + \alpha, \gamma^{-1}\} \dot{C}C_1(1 + \varepsilon(C_{2,1} + C_{2,2}))} / (1 - e^{-\dot{c}/2})$, 且 $C_2 = \mathcal{O}(e^{d/2} d^{3/4})$ 。

定理 3.5 的证明 由定理 4.6、定理 4.7 和定理 4.8 及(30), 有

$$W_2(\mathcal{L}(\bar{\theta}_t^\eta, \bar{V}_t^\eta), \bar{\pi}_\beta) \leq C_1 \eta^{1/2} + C_2 \eta^{1/4} + \sqrt{\dot{C}W_\rho(\mu_0, v_0)} e^{-\dot{c}nt/2}$$

注意到, 对任意 $n \in \mathbb{N}$, $\mathcal{L}(\theta_n^\eta, V_n^\eta) = \mathcal{L}(\bar{\theta}_n^\eta, \bar{V}_n^\eta)$, 故

$$W_2(\mathcal{L}(\theta_n^\eta, V_n^\eta), \bar{\pi}_\beta) \leq C_1 \eta^{1/2} + C_2 \eta^{1/4} + C_3 e^{-C_4 n}$$

其中 $C_3 := \sqrt{\dot{C}W_\rho(\mu_0, v_0)}$, $C_4 := \dot{c}/2$ 。特别地, $C_3 = \mathcal{O}(e^{d/2})$, $C_4 = \mathcal{O}(e^{-d})$ 。

5. 结语

本文研究了非凸优化问题中近似采样 SGHMC 算法的非渐进估计。SGHMC 算法是随机梯度下降算法的一个变体, 该算法结合了随机梯度下降算法和哈密尔顿蒙特卡罗算法的思想, 一方面通过使用随机样本帮助算法逃离局部极小值从而实现算法的全局收敛, 另一方面通过引入哈密尔顿动力学思想在参数空间中模拟粒子的运动从而进行高效采样。通常的研究认为样本数据是独立的, 然而实际的数据往往存在相关性。本文放宽了对数据流独立性的限制, 在损失函数 U 非凸的情形下, 以耗散性条件放宽损失函数强凸的假设, 分析了全局 Lipschitz 条件下样本数据满足条件 L 混合性质时 SGHMC 算法的迭代分布与目标分布 π_β 之间 Wasserstein 距离的上界, 得到 SGHMC 算法收敛到其目标分布的速度为 $\eta^{1/2}$, 这一工作将 SGHMC 算法的收敛结果推广到了相关数据流的情形, 说明了在样本数据独立性失效的情况下 SGHMC 算法的收敛性仍旧成立, 从而拓宽了 SGHMC 算法的应用场景。研究的关键思想是将时间轴划分为长度依赖于步长 η 的等长区间, 在此之上定义合适的辅助连续时间过程, 从而将证明目标进行拆解, 通过比较 SGHMC 算法与所定义的辅助过程之间的距离, 并利用一个收缩估计的结论得到辅助过程与目标分布之间的距离, 从而得到 SGHMC 算法的迭代分布与目标分布 π_β 之间的 Wasserstein 距离上界, 其中条件 L 混合的概念及其相关矩不等式的引入为算法迭代分布与辅助过程之间距离的有界性提供了保障。本文的研究拓宽 SGHMC 算法的应用范围限制, 进一步可以在算法本身的结构上进行优化和改造, 以提高算法的效率, 另外还可以考虑将该算法应用于实际场景中, 通过数值实验来检验算法性能。

参考文献

- [1] Raginsky, M., Rakhlin, A. and Telgarsky, M. (2017) Non-Convex Learning via Stochastic Gradient Langevin Dynamics: A Nonasymptotic Analysis. *Conference on Learning Theory*, **65**, 1674-1703.
- [2] Hwang, C.R. (1980) Laplace's Method Revisited: Weak Convergence of Probability Measures. *The Annals of Probability*, **8**, 1177-1182. <https://doi.org/10.1214/aop/1176994579>
- [3] Chiang, T.S., Hwang, C.R. and Sheu, S.J. (1987) Diffusion for Global Optimization in \mathbb{R}^n . *SIAM Journal on Control and Optimization*, **25**, 737-753. <https://doi.org/10.1137/0325042>
- [4] Welling, M. and Teh, Y.W. (2011) Bayesian Learning via Stochastic Gradient Langevin Dynamics. *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, Bellevue, Washington, 28 June 2011, 681-688.
- [5] Chen, T., Fox, E. and Guestrin, C. (2014) Stochastic Gradient Hamiltonian Monte Carlo. *International Conference on Machine Learning*, Beijing, 22-24 June 2014, 1683-1691.
- [6] Chau, H.N. and Rasonyi, M. (2022) Stochastic Gradient Hamiltonian Monte Carlo for Non-Convex Learning. *Stochastic Processes and Their Applications*, **149**, 341-368. <https://doi.org/10.1016/j.spa.2022.04.001>
- [7] Eberle, A., Guillin, A. and Zimmer, R. (2019) Couplings and Quantitative Contraction Rates for Langevin Dynamics. *The Annals of Probability*, **47**, 1982-2010. <https://doi.org/10.1214/18-AOP1299>
- [8] Chau, N.H., Moulines, É., Rásonyi, M., Sabanis, S. and Zhang, Y. (2021) On Stochastic Gradient Langevin Dynamics with Dependent Data Streams: The Fully Nonconvex Case. *SIAM Journal on Mathematics of Data Science*, **3**, 959-986. <https://doi.org/10.1137/20M1355392>
- [9] Majka, M.B., Mijatović, A. and Szpruch, Ł. (2020) Nonasymptotic Bounds for Sampling Algorithms without Log-Concavity. *The Annals of Applied Probability*, **30**, 1534-1581. <https://doi.org/10.1214/19-AAP1535>
- [10] Cont, R. (2001) Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues. *Quantitative Finance*, **1**, 223-226. <https://doi.org/10.1080/713665670>
- [11] Gao, X., Gürbüzbalaban, M. and Zhu, L. (2022) Global Convergence of Stochastic Gradient Hamiltonian Monte Carlo for Nonconvex Stochastic Optimization: Nonasymptotic Performance Bounds and Momentum-Based Acceleration. *Operations Research*, **70**, 2931-2947. <https://doi.org/10.1287/opre.2021.2162>
- [12] Gerencsér, L. (1989). On a Class of Mixing Processes. *Stochastics: An International Journal of Probability and Stochastic Processes*, **26**, 165-191. <https://doi.org/10.1080/17442508908833555>
- [13] Chau, H.N., Kumar, C., Rásonyi, M. and Sabanis, S. (2019) On Fixed Gain Recursive Estimators with Discontinuity in the Parameters. *ESAIM: Probability and Statistics*, **23**, 217-244. <https://doi.org/10.1051/ps/2018019>
- [14] Barkhagen, M., Chau, N.H., Moulines, É., Rásonyi, M., Sabanis, S. and Zhang, Y. (2021) On Stochastic Gradient Langevin Dynamics with Dependent Data Streams in the Logconcave Case. *Bernoulli*, **27**, 1-33. <https://doi.org/10.3150/19-BEJ1187>
- [15] Akyildiz, Ö.D. and Sabanis, S. (2020) Nonasymptotic Analysis of Stochastic Gradient Hamiltonian Monte Carlo under Local Conditions for Nonconvex Optimization.