Supplementary Material of "Langevin Quasi-Monte Carlo"

Anonymous Author(s)
Affiliation
Address
email

1 Proof of Theorem 4.1

We start by decomposing the error $|\frac{1}{n}\sum_{k=1}^{n} f(\theta_k) - \pi(f)|$ into three parts

$$\begin{aligned}\left|\frac{1}{n}\sum_{k=1}^{n} f(\theta_k) - \pi(f)\right| &\le \left|\frac{1}{n}\sum_{k=\ell+1}^{n} f(\theta_k) - \frac{1}{n}\sum_{k=\ell+1}^{n} \bar{f}_\ell(\mathbf{w}_k^{(\ell)})\right| + \left|\frac{1}{n}\sum_{k=\ell+1}^{n} \bar{f}_\ell(\mathbf{w}_k^{(\ell)}) - \pi(f)\right| + \frac{\ell}{n}2\|f\|_\infty \\ &= (I) + (II) + \frac{2\ell}{n}\|f\|_\infty.\end{aligned}$$

We first upper bound (I).

Lemma 1 (Upper bound of (I); adapted from Lemma 6.1.4 of Chen (2011))**.** *If the transition map ψ is a contraction with parameter ρ and if f is 1-Lipschitz, then*

$$|\bar{f}_\ell(\mathbf{w}_k^{(\ell)}) - f(\theta_k)| \le \left(\max_{0\le i\le n}\|\theta_i\| + \mathbb{E}_\pi\left[\|\theta\|\right]\right)\rho^\ell.$$

Proof of Lemma 1. Note that

$$\begin{aligned}|\bar{f}_\ell(\mathbf{w}_k^{(\ell)}) - f(\theta_k)| &\le \int |f(\psi_\ell(\theta, \mathbf{w}_k^{(\ell)})) - f(\psi_\ell(\theta_{k-\ell}, \mathbf{w}_k^{(\ell)}))|\pi(\mathrm{d}\theta) \\ &\le \int \|(\psi_\ell(\theta, \mathbf{w}_k^{(\ell)})) - (\psi_\ell(\theta_{k-\ell}, \mathbf{w}_k^{(\ell)}))\|\pi(\mathrm{d}\theta) \\ &\le \rho\int \|(\psi_{\ell-1}(\theta, \mathbf{w}_{k-1}^{(\ell-1)})) - (\psi_{\ell-1}(\theta_{k-\ell}, \mathbf{w}_{k-1}^{(\ell-1)}))\|\pi(\mathrm{d}\theta) \\ &\le \rho^\ell\int \|\theta - \theta_{k-\ell}\|\pi(\mathrm{d}\theta) \\ &\le \rho^\ell(\max_{0\le i\le n}\|\theta_i\| + \mathbb{E}_\pi\left[\|\theta\|\right]).\end{aligned}$$

□

To bound (II), note that $\frac{1}{n-\ell}\sum_{k=\ell+1}^{n} \bar{f}_\ell(\mathbf{w}_k^{(\ell)})$ is estimating

$$\mathbb{E}\left[\bar{f}_\ell(\mathbf{w}^{(\ell)})\right] = \int \psi_\ell(\theta, \mathbf{w}^{(\ell)})\pi(\mathrm{d}\theta)\mathrm{d}\mathbf{w}^{(\ell)} =: \pi P_\ell(f).$$

Submitted to 37th Conference on Neural Information Processing Systems (NeurIPS 2023). Do not distribute.

Here, πP_ℓ denote the distribution of the ℓ-step state θ_ℓ starting from $\theta_0 \sim \pi$. So we have the further decomposition

$$\begin{aligned}(II) &\leq \left|\frac{1}{n-\ell}\sum_{k=\ell+1}^{n}\bar{f}_\ell(\mathbf{w}_k^{(\ell)}) - \pi(f)\right| + \frac{\ell}{n-\ell}\|f\|_\infty \\ &\leq |\pi(f) - \pi P_\ell(f)| + \left|\frac{1}{n-\ell}\sum_{k=\ell+1}^{n}\bar{f}_\ell(\mathbf{w}_k^{(\ell)}) - \pi P_\ell(f)\right| + \frac{\ell}{n-\ell}\|f\|_\infty \\ &\leq (II)' + (II)'' + \frac{\ell}{n-\ell}\|f\|_\infty.\end{aligned}$$

The first term $(II)'$ is due to the discretization in time. The second term $(II)''$ is the numerical integration error.

To bound $(II)'$, we use the following result.

Lemma 2 (Upper bound on discretization error $(II)'$)**.** *Under Assumption 1, we have for f 1-Lipschitz,*

$$|\pi(f) - \pi P_\ell(f)| \leq \frac{3\sqrt{2}}{2}\frac{L}{M}h^{1/2}d.$$

Proof of Lemma 2. We let $\theta(t)$ be the continuous-time Langevin diffusion with $\theta(0) = \theta_0 \sim \pi$, $W_{t_{k+1}} - W_{t_k} = \sqrt{h}\xi_{k+1}$, where $\xi_{k+1} \overset{iid}{\sim} \mathcal{N}(0, I_d)$, $t_k = kh$. So we have

$$\theta(t_{k+1}) = \theta(t_k) - \int_{t_k}^{t_{k+1}} \nabla U(\theta(s))\mathrm{d}s + \sqrt{2h}\xi_{k+1}$$

and

$$\theta_{k+1} = \theta_k - h\nabla U(\theta_k) + \sqrt{2h}\xi_{k+1}.$$

Combing the previous two equations gives

$$\theta(t_{k+1}) - \theta_{k+1} = \theta(t_k) - \theta_k - h[\nabla U(\theta(t_k)) - \nabla U(\theta_k)] - \int_{t_k}^{t_{k+1}} \nabla U(\theta(s)) - \nabla U(\theta(t_k))\mathrm{d}s.$$

Let $\Delta_k = \theta(t_k) - \theta_k$. The last display reads

$$\Delta_{k+1} = \Delta_k - h[\nabla U(\theta_k + \Delta_k) - \nabla U(\theta_k)] - \int_{t_k}^{t_{k+1}} \nabla U(\theta(s)) - \nabla U(\theta(t_k))\mathrm{d}s.$$

By the contracting property (6) in the main paper,

$$\|\Delta_k - h[\nabla U(\theta_k + \Delta_k) - \nabla U(\theta_k)]\| \leq \rho\|\Delta_k\|.$$

Taking expectation and use L-smoothness of U, we have

$$\mathbb{E}\left[\|\Delta_{k+1}\|\right] \leq \rho\mathbb{E}\left[\|\Delta_k\|\right] + L\int_{t_k}^{t_{k+1}} \mathbb{E}\left[\|\theta(s) - \theta(t_k)\|\right]\mathrm{d}s.$$

By Lemma 3 of Dalalyan and Karagulyan (2019), $\mathbb{E}\left[\|\nabla U(\theta)\|_2^2\right] \leq Ld$. So we have $\mathbb{E}\left[\|\nabla U(\theta)\|\right] \leq \sqrt{d\mathbb{E}\left[\|\nabla U(\theta)\|_2^2\right]} \leq \sqrt{L}d$. Because $\theta(t)$ is a stationary process,

$$\begin{aligned}\int_{t_k}^{t_{k+1}} \mathbb{E}\left[\|\theta(s) - \theta(t_k)\|\right]\mathrm{d}s &= \int_0^h \mathbb{E}\left[\|\theta(t) - \theta(0)\|\right]\mathrm{d}t \\ &= \int_0^h \mathbb{E}\left[\|-\int_0^t \nabla U(\theta(s))\mathrm{d}s + \sqrt{2}W_t\|\right]\mathrm{d}t \\ &\leq \int_0^h\int_0^t \mathbb{E}\left[\|\nabla U(\theta(s))\|\right]\mathrm{d}s\mathrm{d}t + \int_0^h \sqrt{2}\mathbb{E}\left[\|W_t\|\right]\mathrm{d}t \\ &= \frac{h^2}{2}\sqrt{L}d + \int_0^h \sqrt{2t}\mathbb{E}\left[\|\xi_1\|\right]\mathrm{d}t.\end{aligned}$$

Note that

$$\mathbb{E}\left[\|\xi_1\|\right] = \sqrt{2}\frac{\Gamma(d/2+1/2)}{\Gamma(d/2)} \leq \sqrt{2}(\frac{d+1}{2})^{1/2} = \sqrt{d+1}.$$

Thus,

$$\begin{aligned}\int_{t_k}^{t_{k+1}} \mathbb{E}\left[\|\theta(s)-\theta(t_k)\|\right] \mathrm{d}s &\leq \frac{1}{2}L^{1/2}h^2 d + \frac{3\sqrt{2}}{2}h^{3/2}d^{1/2} \\ &\leq \frac{\sqrt{2}}{2}h^{3/2}d + \frac{3\sqrt{2}}{2}h^{3/2}d^{1/2} \\ &\leq \frac{3\sqrt{2}}{2}h^{3/2}d.\end{aligned}$$

Denote $r = \frac{3\sqrt{2}}{2}Lh^{3/2}d$. So

$$\begin{aligned}\mathbb{E}\left[\|\Delta_{k+1}\|\right] &\leq \rho\mathbb{E}\left[\|\Delta_k\|\right] + r \leq \rho^{k+1}\mathbb{E}\left[\|\Delta_0\|\right] + \sum_{i=0}^{k}\rho^i r \\ &\leq \frac{r}{1-\rho} = \frac{3\sqrt{2}}{2}\frac{L}{M}h^{1/2}d\end{aligned}$$

Therefore, for any $k \geq 1$,

$$\begin{aligned}|\pi(f) - \pi P_k(f)| &= |\mathbb{E}\left[f(\theta(t_k)) - \mathbb{E}\left[f(\theta_k)\right]\right] \leq \mathbb{E}\left[|f(\theta(t_k)) - f(\theta_k)|\right] \\ &\leq \mathbb{E}\left[\|\Delta_k\|\right] \leq \frac{3\sqrt{2}}{2}\frac{L}{M}h^{1/2}d.\end{aligned}$$

□

Theorem 1.1 (Theorem 9.8 of Niederreiter (1992))**.** *Let $v_0, v_1, \ldots$ be an LFSR with offset s and period $n = 2^m - 1$ which satisfy gcd$(m, n) = 1$. Then the sequence $\{\mathbf{u}_i\}_{i=0}^{n-1} \subset [0,1]^s$ with $\mathbf{u}_i = (v_i, v_{i+1}, \ldots, v_{i+s-1})$ has, on average, star-discrepancy*

$$O(n^{-1}(\log n)^{d+1}\log\log n)$$

with an implied constant depending only on d and the average is taken over all primitive polynomials over GF(2) of degree m.

Proof of Theorem 4.1. The error on the left-hand-side is bounded by

$$(I) + (II)' + (II)'' + \frac{4\ell}{n}\|f\|_\infty.$$

Lemma 1 shows that $(I) \leq (\max_{0\leq i<n}\|\theta_i\| + \mathbb{E}_\pi\left[\|\theta\|\right])\rho^\ell \leq (\max_{0\leq i\leq n}\|\theta_i\| + \mathbb{E}_\pi\left[\|\theta\|\right])h^{1/2}$ since $\ell = \lceil(1/2)\log_\rho h\rceil$. Lemma 2 shows that $(II)' \leq \frac{3\sqrt{2}}{2}\frac{L}{M}dh^{1/2}$. Denote $C_2 = \max_{0\leq i\leq n}\|\theta_i\| + \mathbb{E}_\pi\left[\|\theta\|\right] + \frac{3\sqrt{2}}{2}\frac{L}{M}d$. So $(I) + (II)' \leq C_2 h^{1/2}$.

By Theorem 1.1 and the condition that $\gcd(d\ell, n) = 1$, the star-discrepancy $D^*(\{\bar{w}_k^{(\ell)}\}_{k\geq 1})$ is upper bounded by $O(n^{-1}(\log n)^{d\ell+1}\log\log n)$. Finally, by Koksma-Hlawka inequality, we have $(II)'' \leq \|\bar{f}_\ell\|_{\mathrm{HK}} \cdot D^*(\{\bar{w}_k^{(\ell)}\}_{k\geq 1})$. Thus, $(II)'' + \frac{4\ell}{n}\|f\|_\infty \leq C_1 n^{-1+\delta}$, where δ hides the poly-logarithmic terms in $\log n$ and C_1 depends on $d, \ell, \|\bar{f}_\ell\|_{\mathrm{HK}}$.

Therefore, the upper bound becomes

$$(I) + (II)' + (II)'' + \frac{4\ell}{n}\|f\|_\infty \leq C_1 n^{-1+\delta} + C_2 h^{1/2}.$$

□

References

Chen, S. (2011). *Consistency and convergence rate of Markov chain quasi Monte Carlo with examples*. PhD thesis, Stanford University.

Dalalyan, A. S. and Karagulyan, A. (2019). User-friendly guarantees for the Langevin Monte Carlo with inaccurate gradient. *Stochastic Processes and their Applications*, 129(12):5278–5311.

Niederreiter, H. (1992). *Random Number Generation and Quasi-Monte Carlo Methods*. SIAM.