A Proof of Results in Section 3

A.1 Proof of Theorem 3.3 (Global Integration Error, Infinite Time Version)

Proof. We write the solution of an SDE by $\boldsymbol{x}_{t_0,\boldsymbol{x}_{t_0}}(t_0+t)$ when the dependence on initialization needs highlight. Denote $t_k = kh$ and $\boldsymbol{x}_{t_k} = \boldsymbol{x}_k$ for better readability.

We will first make an easy observation that contraction and bounded 2nd-moment of the invariant distribution lead to bounded 2nd-moment of the SDE solution for all time: let $\boldsymbol{y}_0$ be a random variable following the invariant distribution of Eq. (1), i.e., $\boldsymbol{y}_0 \sim \mu$, then $\boldsymbol{y}_t \sim \mu$ and

$$\begin{aligned}
\mathbb{E}\left\|\boldsymbol{x}_t\right\|^2 \leq& 2\mathbb{E}\left\|\boldsymbol{x}_t - \boldsymbol{y}_t\right\|^2 + 2\mathbb{E}\left\|\boldsymbol{y}_t\right\|^2 \\
\leq& 2\mathbb{E}\left\|\boldsymbol{x}_0 - \boldsymbol{y}_0\right\|^2 \exp(-2\beta t) + 2\mathbb{E}\left\|\boldsymbol{y}_t\right\|^2 \\
\leq& 4\mathbb{E}(\left\|\boldsymbol{x}_0\right\|^2 + \left\|\boldsymbol{y}_0\right\|^2)\exp(-2\beta t) + 2\mathbb{E}\left\|\boldsymbol{y}_t\right\|^2 \\
=& 4\mathbb{E}\left\|\boldsymbol{x}_0\right\|^2 \exp(-2\beta t) + \big(2 + 4\exp(-2\beta t)\big)\,\mathbb{E}_{\boldsymbol{y}\sim\mu}\left\|\boldsymbol{y}\right\|^2 \\
\leq& 4\mathbb{E}\left\|\boldsymbol{x}_0\right\|^2 + 6\int_{\mathbb{R}^d}\left\|\boldsymbol{y}\right\|^2 d\mu \triangleq U^2
\end{aligned}$$

and then it follows that

$$\mathbb{E}\left\|\bar{\boldsymbol{x}}_k\right\|^2 \leq 2\mathbb{E}\left\|\bar{\boldsymbol{x}}_k - \boldsymbol{x}_k\right\|^2 + 2\mathbb{E}\left\|\boldsymbol{x}_k\right\|^2 \leq 2e_k^2 + 2U^2. \tag{13}$$

Denote $\langle \boldsymbol{x}, \boldsymbol{y}\rangle_A = \langle A\boldsymbol{x}, A\boldsymbol{y}\rangle$, $\left\|\boldsymbol{x}\right\|_A = \left\|A\boldsymbol{x}\right\|$ and

$$f_k = \left\{\mathbb{E}\left\|\boldsymbol{x}_k - \bar{\boldsymbol{x}}_k\right\|_A^2\right\}^{\frac{1}{2}} \tag{14}$$

where A is the non-singular matrix from Equation (4). Also denote that largest and smallest singular values of A by $\sigma_{\max}$ and $\sigma_{\min}$, respectively, and the condition number of A by $\kappa_A = \frac{\sigma_{\max}}{\sigma_{\min}}$. Recall $e_k = \mathbb{E}\left\|\boldsymbol{x}_k - \bar{\boldsymbol{x}}_k\right\|$, it is easy to see that

$$\sigma_{\min} e_k \leq f_k \leq \sigma_{\max} e_k. \tag{15}$$

Further, we have the following decomposition

$$\begin{aligned}
f_{k+1}^2 =& \mathbb{E}\left\|\boldsymbol{x}_{k+1} - \bar{\boldsymbol{x}}_{k+1}\right\|_A^2 \\
=& \mathbb{E}\left\|\boldsymbol{x}_{t_k,\boldsymbol{x}_{t_k}}(t_{k+1}) - \boldsymbol{x}_{t_k,\bar{\boldsymbol{x}}_k}(t_{k+1}) + \boldsymbol{x}_{t_k,\bar{\boldsymbol{x}}_k}(t_{k+1}) - \bar{\boldsymbol{x}}_{k+1}\right\|_A^2 \\
=& \underbrace{\mathbb{E}\left\|\boldsymbol{x}_{t_k,\boldsymbol{x}_{t_k}}(t_{k+1}) - \boldsymbol{x}_{t_k,\bar{\boldsymbol{x}}_k}(t_{k+1})\right\|_A^2}_{1} + \underbrace{\mathbb{E}\left\|\boldsymbol{x}_{t_k,\bar{\boldsymbol{x}}_k}(t_{k+1}) - \bar{\boldsymbol{x}}_{k+1}\right\|_A^2}_{2} \\
&+ 2\underbrace{\mathbb{E}\left\langle A\left(\boldsymbol{x}_{t_k,\boldsymbol{x}_{t_k}}(t_{k+1}) - \boldsymbol{x}_{t_k,\bar{\boldsymbol{x}}_k}(t_{k+1})\right), A\left(\boldsymbol{x}_{t_k,\bar{\boldsymbol{x}}_k}(t_{k+1}) - \bar{\boldsymbol{x}}_{k+1}\right)\right\rangle}_{3}.
\end{aligned} \tag{16}$$

Term 1 is taken care of the contraction property

$$\mathbb{E}\left\|\boldsymbol{x}_{t_k,\boldsymbol{x}_{t_k}}(t_{k+1}) - \boldsymbol{x}_{t_k,\bar{\boldsymbol{x}}_k}(t_{k+1})\right\|_A^2 \leq f_k^2 \exp(-2\beta h). \tag{17}$$

Term 2 is dealt with by the bound on local strong error

$$\mathbb{E}\left\|\boldsymbol{x}_{t_k,\bar{\boldsymbol{x}}_k}(t_{k+1}) - \bar{\boldsymbol{x}}_{k+1}\right\|_A^2 \leq \sigma_{\max}^2\left(C_2^2 + D_2^2\mathbb{E}\left\|\bar{\boldsymbol{x}}_k\right\|^2\right)h^{2p_2}. \tag{18}$$

Term 3 requires more efforts to cope with, and by the decomposition in Eq. (5) we have

$$
\begin{aligned}
&\mathbb{E}\langle(\boldsymbol{x}_{t_k,\boldsymbol{x}_{t_k}}(t_{k+1})-\boldsymbol{x}_{t_k,\bar{\boldsymbol{x}}_k}(t_{k+1}),\boldsymbol{x}_{t_k,\bar{\boldsymbol{x}}_k}(t_{k+1})-\bar{\boldsymbol{x}}_{k+1}\rangle_A\\
=&\mathbb{E}\langle\boldsymbol{x}_k-\bar{\boldsymbol{x}}_k,\boldsymbol{x}_{t_k,\bar{\boldsymbol{x}}_k}(t_{k+1})-\bar{\boldsymbol{x}}_{k+1}\rangle_A+\mathbb{E}\langle\boldsymbol{z}_h(\boldsymbol{x}_k,\bar{\boldsymbol{x}}_k),\boldsymbol{x}_{t_k,\bar{\boldsymbol{x}}_k}(t_{k+1})-\bar{\boldsymbol{x}}_{k+1}\rangle_A\\
\overset{(i)}{=}&\mathbb{E}\langle\boldsymbol{x}_k-\bar{\boldsymbol{x}}_k,\mathbb{E}[\boldsymbol{x}_{t_k,\bar{\boldsymbol{x}}_k}(t_{k+1})-\bar{\boldsymbol{x}}_{k+1}|\mathcal{F}_k]\rangle_A+\mathbb{E}\langle\boldsymbol{z}_h(\boldsymbol{x}_k,\bar{\boldsymbol{x}}_k),\boldsymbol{x}_{t_k,\bar{\boldsymbol{x}}_k}(t_{k+1})-\bar{\boldsymbol{x}}_{k+1}\rangle_A\\
\overset{(ii)}{\le}& f_k\left(\mathbb{E}\left\|\mathbb{E}[\boldsymbol{x}_{t_k,\bar{\boldsymbol{x}}_k}(t_{k+1})-\bar{\boldsymbol{x}}_{k+1}|\mathcal{F}_k]\right\|_A^2\right)^{\frac{1}{2}}+\left(\mathbb{E}\left\|\boldsymbol{z}_h(\boldsymbol{x}_k,\bar{\boldsymbol{x}}_k)\right\|_A^2\right)^{\frac{1}{2}}\left(\mathbb{E}\left\|\boldsymbol{x}_{t_k,\bar{\boldsymbol{x}}_k}(t_{k+1})-\bar{\boldsymbol{x}}_{k+1}\right\|_A^2\right)^{\frac{1}{2}}\\
\overset{(iii)}{\le}&\sigma_{\max}f_k\left(\mathbb{E}\left\|\mathbb{E}[\boldsymbol{x}_{t_k,\bar{\boldsymbol{x}}_k}(t_{k+1})-\bar{\boldsymbol{x}}_{k+1}|\mathcal{F}_k]\right\|^2\right)^{\frac{1}{2}}+\sigma_{\max}^2\left(\mathbb{E}\left\|\boldsymbol{z}_h(\boldsymbol{x}_k,\bar{\boldsymbol{x}}_k)\right\|^2\right)^{\frac{1}{2}}\left(\mathbb{E}\left\|\boldsymbol{x}_{t_k,\bar{\boldsymbol{x}}_k}(t_{k+1})-\bar{\boldsymbol{x}}_{k+1}\right\|^2\right)^{\frac{1}{2}}\\
\overset{(iv)}{\le}&\sigma_{\max}f_k\left(C_1+D_1\sqrt{\mathbb{E}\left\|\bar{\boldsymbol{x}}_k\right\|^2}\right)h^{p_1}+\kappa_A\sigma_{\max}C_0f_k\sqrt{h}\left(C_2+D_2\sqrt{\mathbb{E}\left\|\bar{\boldsymbol{x}}_k\right\|^2}\right)h^{p_2}\\
\overset{(v)}{\le}&\kappa_A\sigma_{\max}(C_1+C_0C_2)e_kh^{p_2+\frac{1}{2}}+\kappa_A\sigma_{\max}(D_1+C_0D_2)\sqrt{\mathbb{E}\left\|\bar{\boldsymbol{x}}_k\right\|^2}f_kh^{p_2+\frac{1}{2}}
\end{aligned}
\tag{19}
$$

where (i) uses the tower property of conditional expectation and $\mathcal{F}_k$ is the filtration at k-th iteration, (ii) uses Cauchy-Schwarz inequality, (iii) is due to the relationship between e_k and f_k, (iv) is due to local weak error, local strong error and Eq. (5), and (v) is due to $p_1 \ge p_2 + \frac{1}{2}$ and $0 < h \le h_0 \le 1$.

Now plug Eq. (17), (18) and (19) in Eq. (16), we obtain

$$
\begin{aligned}
f_{k+1}^2\le& f_k^2\exp(-2\beta h)+\sigma_{\max}^2\left(C_2^2+D_2^2\mathbb{E}\left\|\bar{\boldsymbol{x}}_k\right\|^2\right)h^{2p_2}+\kappa_A\sigma_{\max}(C_1+C_0C_2)f_kh^{p_2+\frac{1}{2}}\\
&+\kappa_A\sigma_{\max}(D_1+C_0D_2)\sqrt{\mathbb{E}\left\|\bar{\boldsymbol{x}}_k\right\|^2}f_kh^{p_2+\frac{1}{2}}\\
\overset{(i)}{\le}&\left(1-\frac{7}{8}\beta h\right)f_k^2+\sigma_{\max}^2\left(C_2^2+D_2^2\mathbb{E}\left\|\bar{\boldsymbol{x}}_k\right\|^2\right)h^{2p_2}+\kappa_A\sigma_{\max}(C_1+C_0C_2)f_kh^{p_2+\frac{1}{2}}\\
&+\kappa_A\sigma_{\max}(D_1+C_0D_2)\sqrt{\mathbb{E}\left\|\bar{\boldsymbol{x}}_k\right\|^2}f_kh^{p_2+\frac{1}{2}}\\
\overset{(ii)}{\le}&\left(1-\frac{7}{8}\beta h\right)f_k^2+\kappa_A\sigma_{\max}\left(C_1+C_0C_2+\sqrt{2}U(D_1+C_0D_2)\right)f_kh^{p_2+\frac{1}{2}}+2\kappa_A^2D_2^2f_k^2h^{2p_2}\\
&+\sqrt{2}\kappa_A^2(D_1+C_0D_2)f_k^2h^{p_2+\frac{1}{2}}+\sigma_{\max}^2\left(C_2^2+2D_2^2U^2\right)h^{2p_2}\\
\overset{(iii)}{\le}&\left(1-\frac{7}{8}\beta h\right)f_k^2+\kappa_A\sigma_{\max}\left(C_1+C_0C_2+\sqrt{2}U(D_1+C_0D_2)\right)f_kh^{p_2+\frac{1}{2}}+\frac{3\beta}{8}f_k^2h\\
&+\sigma_{\max}^2\left(C_2^2+2D_2^2U^2\right)h^{2p_2}\\
=&\left(1-\frac{1}{2}\beta h\right)f_k^2+\kappa_A\sigma_{\max}\left(C_1+C_0C_2+\sqrt{2}U(D_1+C_0D_2)\right)f_kh^{p_2+\frac{1}{2}}\\
&+\sigma_{\max}^2\left(C_2^2+2D_2^2U^2\right)h^{2p_2}\\
\overset{(iv)}{\le}&\left(1-\frac{1}{2}\beta h\right)f_k^2+\frac{\beta}{4}f_k^2h+\frac{\kappa_A^2\sigma_{\max}^2\left(C_1+C_0C_2+\sqrt{2}U(D_1+C_0D_2)\right)^2}{\beta}h^{2p_2}\\
&+\sigma_{\max}^2\left(C_2^2+2D_2^2U^2\right)h^{2p_2}\\
=&\left(1-\frac{1}{4}\beta h\right)f_k^2+\kappa_A^2\sigma_{\max}^2\left(\frac{\left(C_1+C_0C_2+\sqrt{2}U(D_1+C_0D_2)\right)^2}{\beta}+C_2^2+2D_2^2U^2\right)h^{2p_2}
\end{aligned}
$$

where (i) is due to the assumption $0 < h \le \frac{1}{4\beta}$ and $e^{-x} \le 1 - x + \frac{x^2}{2}$ for $0 < x < 1$, (ii) is due to the upper bound on $\mathbb{E}\left\|\bar{\boldsymbol{x}}_k\right\|^2$ in Eq. (13), (iii) holds provided when $h \le$

$\min\left\{\left(\frac{\sqrt{\beta}}{4\sqrt{2}\kappa_A D_2}\right)^{\frac{1}{p_2-\frac{1}{2}}}, \left(\frac{\beta}{8\sqrt{2}\kappa_A^2(D_1+C_0D_2)}\right)^{\frac{1}{p_2-\frac{1}{2}}}\right\}$ and (iv) is due to Cauchy-Schwarz inequality.

Unfolding the above inequality gives us

$$f_{k+1}^2 \leq \frac{4}{\beta}\kappa_A^2\sigma_{\max}^2\left(\frac{\left(C_1+C_0C_2+\sqrt{2}U(D_1+C_0D_2)\right)^2}{\beta}+C_2^2+2D_2^2U^2\right)h^{2p_2-1}.$$

Taking square root on both sides and using $\sqrt{a^2+b^2+c^2} \leq a+b+c, \forall a,b,c \geq 0$ yields

$$f_{k+1} \leq \frac{2}{\sqrt{\beta}}\kappa_A\sigma_{\max}\left(\frac{C_1+C_0C_2+\sqrt{2}U(D_1+C_0D_2)}{\sqrt{\beta}}+C_2+\sqrt{2}D_2U\right)h^{p_2-\frac{1}{2}}.$$

Finally using the relationship between e_k and f_k, we obtain

$$e_k \leq \frac{2}{\sqrt{\beta}}\kappa_A^2\left(\frac{C_1+C_0C_2+\sqrt{2}U(D_1+C_0D_2)}{\sqrt{\beta}}+C_2+\sqrt{2}D_2U\right)h^{p_2-\frac{1}{2}}.$$

□

A.2 Proof of Theorem 3.4 (Non-Asymptotic Sampling Error Bound: General Case)

Proof. Let $\boldsymbol{y}_0 \sim \mu$ and $(\boldsymbol{x}_0, \boldsymbol{y}_0)$ are coupled such that $\mathbb{E}\left\|\boldsymbol{x}_0-\boldsymbol{y}_0\right\|^2 = W_2^2(\mathrm{Law}(\boldsymbol{x}_0),\mu)$. Denote the solution of Eq. (1) starting from $\boldsymbol{x}_0, \boldsymbol{y}_0$ by $\boldsymbol{x}_t, \boldsymbol{y}_t$ respectively, and $t_k = kh$. We have

$$\begin{aligned}
W_2(\mathrm{Law}(\bar{\boldsymbol{x}}_k),\mu) \leq& W_2(\mathrm{Law}(\bar{\boldsymbol{x}}_k),\mathrm{Law}(\boldsymbol{x}_{t_k})) + W_2(\mathrm{Law}(\boldsymbol{x}_{t_k}),\mu)\\
\leq& \sqrt{\mathbb{E}\left\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_{t_k}\right\|^2}+\sqrt{\mathbb{E}\left\|\boldsymbol{x}_{t_k}-\boldsymbol{y}_{t_k}\right\|^2}\\
\overset{(i)}{\leq}& e_k + \sqrt{\mathbb{E}\left\|\boldsymbol{x}_0-\boldsymbol{y}_0\right\|^2\exp\left(-2\beta t_k\right)}\\
=& e_k + \exp\left(-\beta t_k\right)W_2(\mathrm{Law}(\boldsymbol{x}_0),\mu)
\end{aligned}$$

where (i) is due to the contraction assumption on Eq. (1). Invoking the conclusion of Theorem 3.3 completes the proof. □

A.3 Proof of Corollary 3.5 (Upper Bound of Mixing Time: General Case)

Proof. Given any tolerance $\epsilon > 0$, we know from Theorem 3.4 that if k is large enough and h is small enough such that

$$\exp\left(-\beta kh\right)W_2(\mathrm{Law}(\boldsymbol{x}_0),\mu) \leq \frac{\epsilon}{2}. \tag{20}$$

$$Ch^{p_2-\frac{1}{2}} \leq \frac{\epsilon}{2} \tag{21}$$

we then have $W_2(\mathrm{Law}(\bar{\boldsymbol{x}}_k),\mu) \leq \epsilon$. Solving Inequality (20) yields

$$k \geq \frac{1}{\beta h}\log\frac{2W_2(\mathrm{Law}(\boldsymbol{x}_0),\mu)}{\epsilon} \triangleq k^\star \tag{22}$$

To minimize the lower bound, we want pick step size h as large as possible. Besides $h \leq h_1$, Eq. (21) poses further constraint on h, hence we have

$$h \leq \min\left\{h_1, \left(\frac{\epsilon}{2C}\right)^{\frac{1}{p_2-\frac{1}{2}}}\right\}.$$

Plug the upper bound of h in Eq. (22), we have

$$k^\star = \max\left\{\frac{1}{\beta h_1}, \frac{1}{\beta}\left(\frac{2C}{\epsilon}\right)^{\frac{1}{p_2-\frac{1}{2}}}\right\} \log \frac{2W_2(\mathrm{Law}(\boldsymbol{x}_0), \mu)}{\epsilon}.$$

When high accuracy is needed, i.e., $\epsilon < 2Ch_1^{p_2-\frac{1}{2}}$, we have

$$k^\star = \frac{(2C)^{\frac{1}{p_2-\frac{1}{2}}}}{\beta} \frac{1}{\epsilon^{\frac{1}{p_2-\frac{1}{2}}}} \log \frac{2W_2(\mathrm{Law}(\boldsymbol{x}_0), \mu)}{\epsilon} = \widetilde{\mathcal{O}}\left(\frac{C^{\frac{1}{p_2-\frac{1}{2}}}}{\beta} \frac{1}{\epsilon^{\frac{1}{p_2-\frac{1}{2}}}}\right).$$

□

B Proof of Results in Section 4

B.1 Proof of Theorem 4.1 (Non-Asymptotic Error Bound: LMC)

Proof. From Lemma C.1 we know that Langevin dynamics is a member of the family of contractive SDE, and with a contraction rate of strong-convexity coefficient $\beta = m$ (w.r.t. identity matrix $I_{d\times d}$).

Next, we will need to work out the constants C_0, C_1, D_1, D_2, C_2 needed in Theorem 3.3. We have $C_0 = \frac{\sqrt{m}}{2}$, implied from Lemma C.3.

The local strong error and local weak error are bounded in Lemma D.1 and D.2 respectively. Note that the coefficient $\widetilde{C}_1/\widetilde{C}_2$ in the bound for local strong/weak error depends on initial value, which changes from iteration to iteration. Combined with Lemma D.3, we would obtain C_1 and C_2, namely

$$\widetilde{C}_1 \le 2(L^2+G)\left(\frac{d}{4\kappa L} + \mathbb{E}\left\|\boldsymbol{x}_0\right\|^2 + \frac{8d}{7m} + 1\right)^{\frac{1}{2}} \le 2(L^2+G)\sqrt{\frac{2d}{m} + \mathbb{E}\left\|\boldsymbol{x}_0\right\|^2 + 1} \triangleq C_1$$

and

$$\widetilde{C}_2 \le 2L\left(d + \frac{m}{2}\left(\mathbb{E}\left\|\boldsymbol{x}_0\right\|^2 + \frac{8d}{7m}\right)\right)^{\frac{1}{2}} \le 2L\sqrt{m}\sqrt{\frac{2d}{m} + \mathbb{E}\left\|\boldsymbol{x}_0\right\|^2 + 1} \triangleq C_2.$$

We collect all constants here in the proof for easier reference

$$\begin{aligned}
&A = I_{d\times d}, \kappa_A = 1, \beta = m,\ h_0 = \frac{1}{4\kappa L},\ C_0 = \frac{\sqrt{m}}{2},\\
&C_1 = 2(L^2+G)\sqrt{\frac{2d}{m} + \mathbb{E}\left\|\boldsymbol{x}_0\right\|^2 + 1},\ D_1 = 0\\
&C_2 = 2L\sqrt{m}\sqrt{\frac{2d}{m} + \mathbb{E}\left\|\boldsymbol{x}_0\right\|^2 + 1},\ D_2 = 0.
\end{aligned}$$

Then the constant in Theorem 3.3 for LMC algorithm simplifies to

$$\begin{aligned}
C &= \frac{2}{\sqrt{\beta}}\left(\frac{C_1 + C_0 C_2}{\sqrt{\beta}} + C_2\right),\\
&\le \frac{10(L^2+G)}{m^{\frac{3}{2}}}\sqrt{2d + m\left(\mathbb{E}\left\|\boldsymbol{x}_0\right\|^2 + 1\right)} \triangleq C_{\mathrm{LMC}}.
\end{aligned}$$

Assuming L, m, G are all constants and independent of d, then clearly $C_{\mathrm{LMC}} = \mathcal{O}(\sqrt{d})$. Then applying Theorem 3.4 to LMC, we have

$$W_2(\mathrm{Law}(\bar{\boldsymbol{x}}_k), \mu) \le e^{-mkh} W_2(\mathrm{Law}(\boldsymbol{x}_0), \mu) + C_{\mathrm{LMC}} h \tag{23}$$

for $0 < h \le \frac{1}{4\kappa L}$. □

B.2 PROOF OF THEOREM 4.3 (LOWER BOUND OF MIXING TIME)

Proof. If we start from $\boldsymbol{x}_0 = \mathbf{1}_{2d}$ and run LMC for the potential function in Eq. (11), we then have

$$(\bar{\boldsymbol{x}}_k)_i = \begin{cases} (1-mh)^k(\boldsymbol{x}_0)_i + \sqrt{2h}\sum_{l=1}^k (1-mh)^{k-l}(\boldsymbol{\xi}_l)_i, \ 1 \le i \le d \\ (1-Lh)^k(\boldsymbol{x}_0)_i + \sqrt{2h}\sum_{l=1}^k (1-Lh)^{k-l}(\boldsymbol{\xi}_l)_i, \ d+1 \le i \le 2d \end{cases}$$

and hence

$$(\bar{\boldsymbol{x}}_k)_i \sim \begin{cases} \mathcal{N}\left((1-mh)^k, \frac{2}{m(2-mh)}\left(1-(1-mh)^{2k}\right)\right), \ 1 \le i \le d \\ \mathcal{N}\left((1-Lh)^k, \frac{2}{L(2-Lh)}\left(1-(1-Lh)^{2k}\right)\right), \ d+1 \le i \le 2d \end{cases}$$

Clearly, stability requires $h < \frac{2}{L}$.

The squared 2-Wasserstein distance between the law of the k-th iterate of LMC and target distribution is

$$\begin{aligned} W_2^2(\mathrm{Law}(\bar{\boldsymbol{x}}_k), \mu) =& d(1-mh)^{2k} + \frac{d}{m}\left(\sqrt{\frac{2}{2-mh}}\sqrt{1-(1-mh)^{2k}} - 1\right)^2 \\ &+ d(1-Lh)^{2k} + \frac{d}{L}\left(\sqrt{\frac{2}{2-Lh}}\sqrt{1-(1-Lh)^{2k}} - 1\right)^2. \end{aligned}$$

Suppose $W_2(\mathrm{Law}(\bar{\boldsymbol{x}}_k), \mu) \le \epsilon$, we then must have

$$d(1-mh)^{2k} \le \epsilon^2 \tag{24}$$

$$\frac{d}{m}\left(\sqrt{\frac{2}{2-mh}}\sqrt{1-(1-mh)^{2k}} - 1\right)^2 \le \epsilon^2. \tag{25}$$

A necessary condition of Eq. (25) is that

$$1 + \frac{\sqrt{m}}{\sqrt{d}}\epsilon \ge \sqrt{\frac{2}{2-mh}}\sqrt{1-(1-mh)^{2k}} \overset{(i)}{\ge} \sqrt{\frac{2}{2-mh}}\sqrt{1-\frac{\epsilon^2}{d}} \tag{26}$$

where (i) is due to Eq. (24). It follows from Eq. (26) and $m = 1$ that

$$h \le \frac{4}{1+\frac{\epsilon}{\sqrt{d}}}\frac{\epsilon}{\sqrt{d}} \le \frac{4\epsilon}{\sqrt{d}}. \tag{27}$$

Revisiting Eq. (24) yields

$$\begin{aligned} &\epsilon^2 \ge d(1-mh)^{2k} \overset{(i)}{\ge} d\left(1 - 2mh + \frac{(2mh)^2}{2}\right)^{2k} \overset{(ii)}{\ge} de^{-4mkh} \\ \Longleftrightarrow\ & k \ge \frac{1}{2hm}\log\frac{\sqrt{d}}{\epsilon} \end{aligned} \tag{28}$$

where (i) is due to $mh < \frac{2}{\kappa} < \frac{1}{2}$ and (ii) is due to $e^{-x} \le 1 - x + \frac{x^2}{2}, 0 < x < 1$.

Combine Eq. (27) and (28), we then obtain a lower bound of the mixing time

$$k \ge \frac{\sqrt{d}}{8m\epsilon}\log\frac{\sqrt{d}}{\epsilon} = \frac{\sqrt{d}}{8\epsilon}\log\frac{\sqrt{d}}{\epsilon} = \widetilde{\Omega}\left(\frac{\sqrt{d}}{\epsilon}\right).$$

$\square$

C Some Properties of Langevin Dynamics

C.1 Contraction of Langevin Dynamics

Lemma C.1. *Suppose Assumption 1 holds. Then two copies of overdamped Langevin dynamics have the following contraction property*

$$\left\{\mathbb{E}\left\|\boldsymbol{y}_t-\boldsymbol{x}_t\right\|^2\right\}^{\frac{1}{2}} \leq \left\{\mathbb{E}\left\|\boldsymbol{y}-\boldsymbol{x}\right\|^2\right\}^{\frac{1}{2}} \exp(-mt)$$

where $\boldsymbol{x}, \boldsymbol{y}$ are the initial values of $\boldsymbol{x}_t, \boldsymbol{y}_t$.

Proof. First assume $\boldsymbol{x}, \boldsymbol{y}$ are deterministic. Suppose $\boldsymbol{x}_t, \boldsymbol{y}_t$ are respectively the solutions to

$$\begin{aligned} d\boldsymbol{x}_t &= -\nabla f(\boldsymbol{x}_t)dt + \sqrt{2}d\boldsymbol{B}_t \\ d\boldsymbol{y}_t &= -\nabla f(\boldsymbol{y}_t)dt + \sqrt{2}d\boldsymbol{B}_t \end{aligned}$$

where $\boldsymbol{B}_t$ is a standard d-dimensional Brownian motion. Denote $L_t = \frac{1}{2}\mathbb{E}\left\|\boldsymbol{y}_t-\boldsymbol{x}_t\right\|^2$ and take time derivative, we obtain

$$\frac{d}{dt}L_t = -\mathbb{E}\langle \boldsymbol{y}_t-\boldsymbol{x}_t, \nabla f(\boldsymbol{y}_t)-\nabla f(\boldsymbol{x}_t)\rangle \overset{(i)}{\leq} -m\mathbb{E}\left\|\boldsymbol{y}_t-\boldsymbol{x}_t\right\|^2 = -2mL_t$$

where (i) is due to the strong-convexity assumption made on f. We then obtain $L_t \leq L_0 \exp(-2mt)$ and it follows by Gronwall's inequality that

$$\left\{\mathbb{E}\left\|\boldsymbol{y}_t-\boldsymbol{x}_t\right\|^2\right\}^{\frac{1}{2}} \leq \left\|\boldsymbol{y}-\boldsymbol{x}\right\| \exp(-mt).$$

When $\boldsymbol{x}, \boldsymbol{y}$ are random, by the conditioning version of the above inequality and Jensen's inequality, we have

$$\left\{\mathbb{E}\left[\mathbb{E}\left\|\boldsymbol{y}_t-\boldsymbol{x}_t\right\|^2 \middle| \boldsymbol{x},\boldsymbol{y}\right]\right\}^{\frac{1}{2}} \leq \left\{\mathbb{E}\left\|\boldsymbol{y}-\boldsymbol{x}\right\|^2 \exp(-2mt)\right\}^{\frac{1}{2}} = \left\{\mathbb{E}\left\|\boldsymbol{y}-\boldsymbol{x}\right\|^2\right\}^{\frac{1}{2}} \exp(-mt).$$

$\square$

C.2 Growth Bound of Langevin Dynamics

Lemma C.2. *Suppose Assumption 1 holds, then when $0 \leq h \leq \frac{1}{4\kappa L}$, the solution of overdamped Langevin dynamics $\boldsymbol{x}_t$ satisfies*

$$\mathbb{E}\left\|\boldsymbol{x}_h-\boldsymbol{x}\right\|^2 \leq 6\left(d + \frac{m}{2}\mathbb{E}\left\|\boldsymbol{x}\right\|^2\right)h$$

where $\boldsymbol{x}$ is the initial value at $t=0$.

Proof. We have

$$\begin{aligned}
\mathbb{E}\left\|\boldsymbol{x}_h-\boldsymbol{x}\right\|^2 =&\mathbb{E}\left\|-\int_0^h \nabla f(\boldsymbol{x}_t)dt+\sqrt{2}\int_0^h d\boldsymbol{B}_t\right\|^2\\
\leq&2\mathbb{E}\left\|\int_0^h \nabla f(\boldsymbol{x}_t)dt\right\|^2+4\mathbb{E}\left\|\int_0^h d\boldsymbol{B}_t\right\|^2\\
\overset{(i)}{=}&2\mathbb{E}\left\|\int_0^h \nabla f(\boldsymbol{x}_t)dt\right\|^2+4hd\\
\leq&2\mathbb{E}\left[\left(\int_0^h \left\|\nabla f(\boldsymbol{x}_t)-\nabla f(\boldsymbol{x})\right\|dt+\int_0^h\left\|\nabla f(\boldsymbol{x})\right\|dt\right)^2\right]+4hd\\
\leq&2\mathbb{E}\left[\left(L\int_0^h \left\|\boldsymbol{x}_t-\boldsymbol{x}\right\|dt+h\left\|\nabla f(\boldsymbol{x})\right\|\right)^2\right]+4hd\\
\leq&4\mathbb{E}\left[L^2\left(\int_0^h \left\|\boldsymbol{x}_t-\boldsymbol{x}\right\|dt\right)^2+h^2\left\|\nabla f(\boldsymbol{x})\right\|^2\right]+4hd\\
\overset{(ii)}{\leq}&4hd+4h^2\mathbb{E}\left\|\nabla f(\boldsymbol{x})\right\|^2+4L^2h\int_0^h\mathbb{E}\left\|\boldsymbol{x}_t-\boldsymbol{x}\right\|^2dt
\end{aligned}$$

where (i) is due to Ito's isometry, (ii) is due to Cauchy-Schwarz inequality. By Gronwall's inequality, we obtain

$$\mathbb{E}\left\|\boldsymbol{x}_h-\boldsymbol{x}\right\|^2\leq 4h\left(d+h\mathbb{E}\left\|\nabla f(\boldsymbol{x})\right\|^2\right)\exp\left\{4L^2h^2\right\}.$$

Since $\left\|\nabla f(\boldsymbol{x})\right\|=\left\|\nabla f(\boldsymbol{x})-\nabla f(\boldsymbol{0})\right\|\leq L\left\|\boldsymbol{x}\right\|$, when $0<h\leq\frac{1}{4\kappa L}$, we finally reach at

$$\mathbb{E}\left\|\boldsymbol{x}_h-\boldsymbol{x}\right\|^2\leq 4e^{\frac{1}{4}}\left(d+2hL^2\mathbb{E}\left\|\boldsymbol{x}\right\|^2\right)h\leq 6\left(d+\frac{m}{2}\mathbb{E}\left\|\boldsymbol{x}\right\|^2\right)h.$$

□

C.3 Bound on Evolved Deviation

Lemma C.3. *Suppose Assumption 1 holds. Let $\boldsymbol{x}_t,\boldsymbol{y}_t$ be two solutions of overdamped Langevin dynamics starting from $\boldsymbol{x},\boldsymbol{y}$ respectively, for $0<h\leq\frac{1}{4\kappa L}$, we have the following representation*

$$\boldsymbol{x}_h-\boldsymbol{y}_h=\boldsymbol{x}-\boldsymbol{y}+\boldsymbol{z}$$

with

$$E\left\|\boldsymbol{z}\right\|^2\leq\frac{m}{4}\mathbb{E}\left\|\boldsymbol{x}-\boldsymbol{y}\right\|^2h.$$

Proof. Let $\boldsymbol{z}=(\boldsymbol{x}_h-\boldsymbol{y}_h)-(\boldsymbol{x}-\boldsymbol{y})=-\int_0^h\nabla f(\boldsymbol{x}_s)-\nabla f(\boldsymbol{y}_s)ds$. Ito's lemma readily implies that

$$\begin{aligned}
\mathbb{E}\left\|\boldsymbol{x}_h-\boldsymbol{y}_h\right\|^2=&\mathbb{E}\left\|\boldsymbol{x}-\boldsymbol{y}\right\|^2-2\mathbb{E}\int_0^h\langle\boldsymbol{x}_s-\boldsymbol{y}_s,\nabla f(\boldsymbol{x}_s)-\nabla f(\boldsymbol{y}_s)\rangle ds\\
\overset{(i)}{\leq}&\mathbb{E}\left\|\boldsymbol{x}-\boldsymbol{y}\right\|^2-2m\int_0^h\mathbb{E}\left\|\boldsymbol{x}_s-\boldsymbol{y}_s\right\|^2ds\\
\leq&\mathbb{E}\left\|\boldsymbol{x}-\boldsymbol{y}\right\|^2
\end{aligned}$$

where (i) is due to strong-convexity of f. We then have that

$$\begin{aligned}
\mathbb{E}\left\|\boldsymbol{z}\right\|^2 =& \left\|\mathbb{E}\left[\int_0^h \nabla f(\boldsymbol{x}_s) - \nabla f(\boldsymbol{y}_s) ds\right]\right\|^2 \\
\leq & \left(\int_0^h \left\|\mathbb{E}\left[\nabla f(\boldsymbol{x}_s) - \nabla f(\boldsymbol{y}_s)\right]\right\| ds\right)^2 \\
\leq & \int_0^h 1^2 ds \int_0^h \left\|\mathbb{E}\left[\nabla f(\boldsymbol{x}_s) - \nabla f(\boldsymbol{y}_s)\right]\right\|^2 ds \\
\leq & h \int_0^h \mathbb{E}\left\|\nabla f(\boldsymbol{x}_s) - \nabla f(\boldsymbol{y}_s)\right\|^2 ds \\
\leq & L^2 h \int_0^h \mathbb{E}\left\|\boldsymbol{x}_s - \boldsymbol{y}_s\right\|^2 ds \\
\leq & L^2 \mathbb{E}\left\|\boldsymbol{x} - \boldsymbol{y}\right\|^2 h^2 \\
\overset{(i)}{\leq} & \frac{m}{4}\mathbb{E}\left\|\boldsymbol{x} - \boldsymbol{y}\right\|^2 h
\end{aligned}$$

where (i) is due to $h \leq \frac{1}{4\kappa L}$. □

D Some Properties of LMC Algorithm

D.1 Local Strong Error

Lemma D.1. *Suppose Assumption 1 holds. Denote the one-step iteration of LMC algorithm with step size h by $\bar{\boldsymbol{x}}_1$ and the solution of overdamped Langevin dynamics at time $t = h$ by $\boldsymbol{x}_h$. Both the discrete algorithm and the continuous dynamics start from the same initial value $\boldsymbol{x}$. If $0 \leq h \leq \frac{1}{4\kappa L}$, then the local strong error of LMC algorithm satisfies*

$$\left\{\mathbb{E}\left\|\bar{\boldsymbol{x}}_1 - \boldsymbol{x}_h\right\|^2\right\}^{\frac{1}{2}} \leq \widetilde{C}_2 h^{\frac{3}{2}}$$

with $\widetilde{C}_2 = 2L\left(d + \frac{m}{2}\mathbb{E}\left\|\boldsymbol{x}\right\|^2\right)^{\frac{1}{2}}$.

Proof. We have for $0 \leq h \leq \frac{1}{4\kappa L}$,

$$\begin{aligned}
\mathbb{E}\left\|\bar{\boldsymbol{x}}_1 - \boldsymbol{x}_h\right\|^2 =& \mathbb{E}\left\|\int_0^h \nabla f(\boldsymbol{x}_s) - \nabla f(\boldsymbol{x}) ds\right\|^2 \\
\leq & \mathbb{E}\left(\int_0^h \left\|\nabla f(\boldsymbol{x}_s) - \nabla f(\boldsymbol{x})\right\| ds\right)^2 \\
\leq & L^2 \mathbb{E}\left(\int_0^h \left\|\boldsymbol{x}_s - \boldsymbol{x}\right\| ds\right)^2 \\
\overset{(i)}{\leq} & L^2 h \int_0^h \mathbb{E}\left\|\boldsymbol{x}_s - \boldsymbol{x}\right\|^2 ds \\
\overset{(ii)}{\leq} & 3L^2\left(d + \frac{m}{2}\mathbb{E}\left\|\boldsymbol{x}\right\|^2\right) h^3
\end{aligned}$$

where (i) is due to Cauchy-Schwartz inequality and (ii) is due to Lemma C.2. Taking square roots on both side completes the proof. □

D.2 Local Weak Error

Lemma D.2. *Suppose Assumption 1 and 2 hold. Denote the one-step iteration of LMC algorithm with step size h by $\bar{\boldsymbol{x}}_1$ and the solution of overdamped Langevin dynamics at time $t = h$ by $\boldsymbol{x}_h$. Both the discrete algorithm and the continuous dynamics start from the same initial value $\boldsymbol{x}$. If $0 \le h \le \frac{1}{4\kappa L}$, then the local weak error of LMC algorithm satisfies*

$$\|\mathbb{E}\bar{\boldsymbol{x}}_1 - \mathbb{E}\boldsymbol{x}_h\| \le \widetilde{C}_1 h^2$$

with $\widetilde{C}_1 = 2(L^2+G)\left(\frac{d}{4\kappa L} + \mathbb{E}\|\boldsymbol{x}\|^2 + 1\right)^{\frac{1}{2}}$.

Proof. By Ito's lemma, we have

$$d\nabla f(\boldsymbol{x}_t) = -\nabla^2 f(\boldsymbol{x}_t)\nabla f(\boldsymbol{x}_t)dt + \nabla(\Delta f(\boldsymbol{x}_t))dt + \sqrt{2}\int_0^t \nabla^2 f(\boldsymbol{x}_t)d\boldsymbol{B}_t.$$

It follows that

$$\begin{aligned}
\|\mathbb{E}\bar{\boldsymbol{x}}_1 - \mathbb{E}\boldsymbol{x}_h\| &= \left\|\mathbb{E}\int_0^h \nabla f(\boldsymbol{x}_s) - \nabla f(\boldsymbol{x})ds\right\| \\
&= \left\|\mathbb{E}\left\{\int_0^h\int_0^s -\nabla^2 f(\boldsymbol{x}_r)\nabla f(\boldsymbol{x}_r) + \nabla(\Delta f(\boldsymbol{x}_r))drds + \sqrt{2}\int_0^h\int_0^s \nabla^2 f(\boldsymbol{x}_r)d\boldsymbol{B}_r ds\right\}\right\| \\
&= \left\|\mathbb{E}\left\{\int_0^h\int_0^s -\nabla^2 f(\boldsymbol{x}_r)\nabla f(\boldsymbol{x}_r) + \nabla(\Delta f(\boldsymbol{x}_r))drds\right\}\right\| \\
&\le \int_0^h\int_0^s \mathbb{E}\left\|\nabla^2 f(\boldsymbol{x}_r)\nabla f(\boldsymbol{x}_r)\right\| drds + \int_0^h\int_0^s \mathbb{E}\left\|\nabla(\Delta f(\boldsymbol{x}_r))\right\| drds \\
&\le L\int_0^h\int_0^s \mathbb{E}\left\|\nabla f(\boldsymbol{x}_r)\right\| drds + \int_0^h\int_0^s \mathbb{E}\left\|\nabla(\Delta f(\boldsymbol{x}_r))\right\| drds \\
&\overset{(i)}{\le} (L^2+G)\int_0^h\int_0^s \mathbb{E}\|\boldsymbol{x}_r\| drds + \frac{G}{2}h^2 \\
&\le (L^2+G)\left(\int_0^h\int_0^s \mathbb{E}\|\boldsymbol{x}_r - \boldsymbol{x}\| drds + \frac{h^2}{2}\mathbb{E}\|\boldsymbol{x}\|\right) + \frac{G}{2}h^2 \\
&\overset{(ii)}{\le} (L^2+G)\left(\int_0^h\int_0^s \sqrt{\mathbb{E}\|\boldsymbol{x}_r - \boldsymbol{x}\|^2} drds + \frac{h^2}{2}\mathbb{E}\|\boldsymbol{x}\|\right) + \frac{G}{2}h^2 \\
&\overset{(iii)}{\le} (L^2+G)\left(\int_0^h\int_0^s \sqrt{6\left(d + \frac{m}{2}\mathbb{E}\|\boldsymbol{x}\|^2\right)r} drds + \frac{h^2}{2}\mathbb{E}\|\boldsymbol{x}\|\right) + \frac{G}{2}h^2 \\
&= (L^2+G)\left(\frac{4\sqrt{6}}{15}\sqrt{\left(d + \frac{m}{2}\mathbb{E}\|\boldsymbol{x}\|^2\right)h} + \frac{1}{2}\mathbb{E}\|\boldsymbol{x}\|\right)h^2 + \frac{G}{2}h^2 \\
&\overset{(iv)}{\le} (L^2+G)h^2\sqrt{\left(d + \frac{m}{2}\mathbb{E}\|\boldsymbol{x}\|^2\right)h + \frac{1}{2}\mathbb{E}\|\boldsymbol{x}\|^2} + \frac{G}{2}h^2 \\
&\overset{(v)}{\le} (L^2+G)h^2\sqrt{\frac{d}{4\kappa L} + \mathbb{E}\|\boldsymbol{x}\|^2} + \frac{G}{2}h^2 \\
&\le (L^2+G)\left(\sqrt{\frac{d}{4\kappa L} + \mathbb{E}\|\boldsymbol{x}\|^2} + 1\right)h^2 \\
&\le 2(L^2+G)\left(\frac{d}{4\kappa L} + \mathbb{E}\|\boldsymbol{x}\|^2 + 1\right)^{\frac{1}{2}} h^2
\end{aligned}$$

where (i) is due to Assumption 2, (ii) is due to Jensen's inequality, (iii) is due to Lemma C.2, (iv) is due to $\sqrt{a}+\sqrt{b} \leq \sqrt{2}\sqrt{a^2+b^2}$ and (v) is due to $h \leq \frac{1}{4\kappa L}$. It is worth noting in the third equation that the Ito's correction term $\nabla\Delta f$ can also be written as $\Delta\nabla f$ as the two operators commute for $\mathcal{C}^3$ functions. $\square$

D.3 Boundedness of LMC Algorithm

Lemma D.3. *Suppose Assumption 1 holds. Denote the iterates of LMC by $\bar{\boldsymbol{x}}_k$. If $0 \leq h \leq \frac{1}{4\kappa L}$ we then have the iterates of LMC algorithm are uniformly upper bounded by*

$$\mathbb{E}\left\|\bar{\boldsymbol{x}}_k\right\|^2 \leq \mathbb{E}\left\|\boldsymbol{x}_0\right\|^2 + \frac{8d}{7m}, \quad \forall k \geq 0$$

Proof. We have

$$\begin{aligned}
\mathbb{E}\left\|\bar{\boldsymbol{x}}_{k+1}\right\|^2 =& \mathbb{E}\left\|\bar{\boldsymbol{x}}_k - h\nabla f(\bar{\boldsymbol{x}}_k) + \sqrt{2h}\boldsymbol{\xi}_{k+1}\right\|^2 \\
\stackrel{(i)}{=}& \mathbb{E}\left\|\bar{\boldsymbol{x}}_k\right\|^2 + h^2\mathbb{E}\left\|\nabla f(\bar{\boldsymbol{x}}_k)\right\|^2 + 2hd - 2h\mathbb{E}\langle\bar{\boldsymbol{x}}_k, \nabla f(\bar{\boldsymbol{x}}_k)\rangle \\
=& \mathbb{E}\left\|\bar{\boldsymbol{x}}_k\right\|^2 + h^2\mathbb{E}\left\|\nabla f(\bar{\boldsymbol{x}}_k) - \nabla f(0)\right\|^2 + 2hd - 2h\mathbb{E}\langle\bar{\boldsymbol{x}}_k, \nabla f(\bar{\boldsymbol{x}}_k)\rangle \\
\stackrel{(ii)}{\leq}& \mathbb{E}\left\|\bar{\boldsymbol{x}}_k\right\|^2 + h^2L^2\mathbb{E}\left\|\bar{\boldsymbol{x}}_k\right\|^2 + 2hd - 2h\mathbb{E}\langle\bar{\boldsymbol{x}}_k, \nabla f(\bar{\boldsymbol{x}}_k)\rangle \\
\stackrel{(iii)}{\leq}& \mathbb{E}\left\|\bar{\boldsymbol{x}}_k\right\|^2 + h^2L^2\mathbb{E}\left\|\bar{\boldsymbol{x}}_k\right\|^2 + 2hd - 2mh\mathbb{E}\left\|\bar{\boldsymbol{x}}_k\right\|^2 \\
\stackrel{(iv)}{\leq}& \left(1 - \frac{7}{4}mh\right)\mathbb{E}\left\|\bar{\boldsymbol{x}}_k\right\|^2 + 2hd
\end{aligned}$$

where (i) is due to the independence between $\boldsymbol{\xi}_{k+1}$ and $\bar{\boldsymbol{x}}_k$, (ii) is due to Assumption 1, (iii) is due to the property of m-strongly-convex functions, $\langle\nabla f(\boldsymbol{y}) - \nabla f(\boldsymbol{x}), \boldsymbol{y}-\boldsymbol{x}\rangle \geq m\left\|\boldsymbol{y}-\boldsymbol{x}\right\|^2\ \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, and (iv) uses the assumption $h \leq \frac{1}{4\kappa L}$.

Unfolding the inequality, we obtain

$$\mathbb{E}\left\|\bar{\boldsymbol{x}}_k\right\|^2 \leq (1 - \frac{7}{4}mh)^k\mathbb{E}\left\|\bar{\boldsymbol{x}}_0\right\|^2 + 2hd\left(1 + \frac{7}{4}mh + \cdots + (\frac{7}{4}mh)^{k-1}\right) \leq \mathbb{E}\left\|\boldsymbol{x}_0\right\|^2 + \frac{8d}{7m}$$

$\square$