

430 A Proof of Results in Section 3

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

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

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

$$\begin{aligned} \mathbb{E} \|\mathbf{x}_t\|^2 &\leq 2\mathbb{E} \|\mathbf{x}_t - \mathbf{y}_t\|^2 + 2\mathbb{E} \|\mathbf{y}_t\|^2 \\ &\leq 2\mathbb{E}_{\mathbf{y} \sim \mu} \|\mathbf{x}_0 - \mathbf{y}_0\|^2 \exp(-2\beta t) + 2\mathbb{E} \|\mathbf{y}_t\|^2 \\ &\leq 4\|\mathbf{x}_0\|^2 \exp(-2\beta t) + (2 + 5\exp(-2\beta t)) \int_{\mathbb{R}^d} \|\mathbf{w}\|^2 d\mu \\ &\leq 4\|\mathbf{x}_0\|^2 + 5 \int_{\mathbb{R}^d} \|\mathbf{w}\|^2 d\mu \triangleq U^2 \end{aligned}$$

437 and then it follows that

$$\mathbb{E} \|\bar{\mathbf{x}}_k\|^2 \leq 2\mathbb{E} \|\bar{\mathbf{x}}_k - \mathbf{x}_k\|^2 + 2\mathbb{E} \|\mathbf{x}_k\|^2 \leq 2e_k^2 + 2U^2. \quad (13)$$

438 Denote $\langle \mathbf{x}, \mathbf{y} \rangle_A = \langle A\mathbf{x}, A\mathbf{y} \rangle$, $\|\mathbf{x}\|_A = \|A\mathbf{x}\|$ and

$$f_k = \mathbb{E} \|\mathbf{x}_k - \bar{\mathbf{x}}_k\|_A \quad (14)$$

439 where A is the non-singular matrix from Equation (4). Also denote that largest and smallest singular
440 values of A by σ_{\max} and σ_{\min} , respectively, and the condition number of A by $\kappa_A = \frac{\sigma_{\max}}{\sigma_{\min}}$. Recall
441 $e_k = \mathbb{E} \|\mathbf{x}_k - \bar{\mathbf{x}}_k\|$, it is easy to see that

$$\sigma_{\min} e_k \leq f_k \leq \sigma_{\max} e_k. \quad (15)$$

442 Further, we have the following decomposition

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

443 Term $\textcircled{1}$ is taken care of the contraction property

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

444 Term $\textcircled{2}$ is dealt with by the bound on local strong error

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

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

$$\begin{aligned}
& \mathbb{E} \langle (\mathbf{x}_{t_k, \mathbf{x}_{t_k}}(t_{k+1}) - \mathbf{x}_{t_k, \bar{\mathbf{x}}_k}(t_{k+1}), \mathbf{x}_{t_k, \bar{\mathbf{x}}_k}(t_{k+1}) - \bar{\mathbf{x}}_{k+1} \rangle_A \\
&= \mathbb{E} \langle \mathbf{x}_k - \bar{\mathbf{x}}_k, \mathbf{x}_{t_k, \bar{\mathbf{x}}_k}(t_{k+1}) - \bar{\mathbf{x}}_{k+1} \rangle_A + \mathbb{E} \langle \mathbf{z}, \mathbf{x}_{t_k, \bar{\mathbf{x}}_k}(t_{k+1}) - \bar{\mathbf{x}}_{k+1} \rangle_A \\
&\stackrel{(i)}{=} \mathbb{E} \langle \mathbf{x}_k - \bar{\mathbf{x}}_k, \mathbb{E}[\mathbf{x}_{t_k, \bar{\mathbf{x}}_k}(t_{k+1}) - \bar{\mathbf{x}}_{k+1} | \mathcal{F}_k] \rangle_A + \mathbb{E} \langle \mathbf{z}, \mathbf{x}_{t_k, \bar{\mathbf{x}}_k}(t_{k+1}) - \bar{\mathbf{x}}_{k+1} \rangle_A \\
&\stackrel{(ii)}{\leq} f_k \left(\mathbb{E} \|\mathbb{E}[\mathbf{x}_{t_k, \bar{\mathbf{x}}_k}(t_{k+1}) - \bar{\mathbf{x}}_{k+1} | \mathcal{F}_k]\|_A^2 \right)^{\frac{1}{2}} + \left(\mathbb{E} \|\mathbf{z}\|_A^2 \right)^{\frac{1}{2}} \left(\mathbb{E} \|\mathbf{x}_{t_k, \bar{\mathbf{x}}_k}(t_{k+1}) - \bar{\mathbf{x}}_{k+1}\|_A^2 \right)^{\frac{1}{2}} \\
&\stackrel{(iii)}{\leq} \sigma_{\max} f_k \left(\mathbb{E} \|\mathbb{E}[\mathbf{x}_{t_k, \bar{\mathbf{x}}_k}(t_{k+1}) - \bar{\mathbf{x}}_{k+1} | \mathcal{F}_k]\|_A^2 \right)^{\frac{1}{2}} + \sigma_{\max}^2 \left(\mathbb{E} \|\mathbf{z}\|^2 \right)^{\frac{1}{2}} \left(\mathbb{E} \|\mathbf{x}_{t_k, \bar{\mathbf{x}}_k}(t_{k+1}) - \bar{\mathbf{x}}_{k+1}\|^2 \right)^{\frac{1}{2}} \\
&\stackrel{(iv)}{\leq} \sigma_{\max} f_k \left(C_1 + D_1 \sqrt{\mathbb{E} \|\bar{\mathbf{x}}_k\|^2} \right) h^{p_1} + \kappa_A \sigma_{\max} C_0 f_k \sqrt{h} \left(C_2 + D_2 \sqrt{\mathbb{E} \|\bar{\mathbf{x}}_k\|^2} \right) h^{p_2} \\
&\stackrel{(v)}{\leq} \kappa_A \sigma_{\max} (C_1 + C_0 C_2) e_k h^{p_2 + \frac{1}{2}} + \kappa_A \sigma_{\max} (D_1 + C_0 D_2) \sqrt{\mathbb{E} \|\bar{\mathbf{x}}_k\|^2} f_k h^{p_2 + \frac{1}{2}} \tag{19}
\end{aligned}$$

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 $p_1 \geq p_2 + \frac{1}{2}$ and $0 < h \leq h_0 \leq 1$.

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

$$\begin{aligned}
f_{k+1}^2 &\leq f_k^2 \exp(-2\beta h) + \sigma_{\max}^2 \left(C_2^2 + D_2^2 \mathbb{E} \|\bar{\mathbf{x}}_k\|^2 \right) h^{2p_2} + \kappa_A \sigma_{\max} (C_1 + C_0 C_2) f_k h^{p_2 + \frac{1}{2}} \\
&\quad + \kappa_A \sigma_{\max} (D_1 + C_0 D_2) \sqrt{\mathbb{E} \|\bar{\mathbf{x}}_k\|^2} f_k h^{p_2 + \frac{1}{2}} \\
&\stackrel{(i)}{\leq} \left(1 - \frac{7}{8} \beta h \right) f_k^2 + \sigma_{\max}^2 \left(C_2^2 + D_2^2 \mathbb{E} \|\bar{\mathbf{x}}_k\|^2 \right) h^{2p_2} + \kappa_A \sigma_{\max} (C_1 + C_0 C_2) f_k h^{p_2 + \frac{1}{2}} \\
&\quad + \kappa_A \sigma_{\max} (D_1 + C_0 D_2) \sqrt{\mathbb{E} \|\bar{\mathbf{x}}_k\|^2} f_k h^{p_2 + \frac{1}{2}} \\
&\stackrel{(ii)}{\leq} \left(1 - \frac{7}{8} \beta h \right) f_k^2 + \kappa_A \sigma_{\max} \left(C_1 + C_0 C_2 + \sqrt{2} U (D_1 + C_0 D_2) \right) f_k h^{p_2 + \frac{1}{2}} + 2\kappa_A^2 D_2^2 f_k^2 h^{2p_2} \\
&\quad + \sqrt{2} \kappa_A^2 (D_1 + C_0 D_2) f_k^2 h^{p_2 + \frac{1}{2}} + \sigma_{\max}^2 \left(C_2^2 + 2D_2^2 U^2 \right) h^{2p_2} \\
&\stackrel{(iii)}{\leq} \left(1 - \frac{7}{8} \beta h \right) f_k^2 + \kappa_A \sigma_{\max} \left(C_1 + C_0 C_2 + \sqrt{2} U (D_1 + C_0 D_2) \right) f_k h^{p_2 + \frac{1}{2}} + \frac{3\beta}{8} f_k^2 h \\
&\quad + \sigma_{\max}^2 \left(C_2^2 + 2D_2^2 U^2 \right) h^{2p_2} \\
&= \left(1 - \frac{1}{2} \beta h \right) f_k^2 + \kappa_A \sigma_{\max} \left(C_1 + C_0 C_2 + \sqrt{2} U (D_1 + C_0 D_2) \right) f_k h^{p_2 + \frac{1}{2}} \\
&\quad + \sigma_{\max}^2 \left(C_2^2 + 2D_2^2 U^2 \right) h^{2p_2} \\
&\stackrel{(iv)}{\leq} \left(1 - \frac{1}{2} \beta h \right) f_k^2 + \frac{\beta}{4} f_k^2 h + \frac{\kappa_A^2 \sigma_{\max}^2 \left(C_1 + C_0 C_2 + \sqrt{2} U (D_1 + C_0 D_2) \right)^2}{\beta} h^{2p_2} \\
&\quad + \sigma_{\max}^2 \left(C_2^2 + 2D_2^2 U^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_0 C_2 + \sqrt{2} U (D_1 + C_0 D_2) \right)^2}{\beta} + C_2^2 + 2D_2^2 U^2 \right) h^{p_2}
\end{aligned}$$

where (i) is due to the assumption $0 < h \leq \frac{1}{4\beta}$ and $e^{-x} \leq 1 - x + \frac{x^2}{2}$ for $0 < x < 1$,

(ii) is due to the upper bound on $\mathbb{E} \|\bar{\mathbf{x}}_k\|^2$ in Eq. (13), (iii) holds provided when $h \leq$

$\min \left\{ \left(\frac{\sqrt{\beta}}{4\sqrt{2}\kappa_A D_2} \right)^{\frac{1}{p_2 - \frac{1}{2}}}, \left(\frac{\beta}{8\sqrt{2}\kappa_A (D_1 + C_0 D_2)} \right)^{\frac{1}{p_2 - \frac{1}{2}}} \right\}$ and (iv) is due to Cauchy-Schwarz inequality.

453 Unfolding the above inequality gives us

$$f_{k+1}^2 \leq \frac{4}{\beta} \kappa_A^2 \sigma_{\max}^2 \left(\frac{(C_1 + C_0 C_2 + \sqrt{2} U (D_1 + C_0 D_2))^2}{\beta} + C_2^2 + 2 D_2^2 U^2 \right) h^{2p_2-1}.$$

454 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

$$f_{k+1} \leq \frac{2}{\sqrt{\beta}} \kappa_A \sigma_{\max} \left(\frac{C_1 + C_0 C_2 + \sqrt{2} U (D_1 + C_0 D_2)}{\sqrt{\beta}} + C_2 + \sqrt{2} D_2 U \right) h^{p_2-\frac{1}{2}}.$$

455 Finally using the relationship between e_k and f_k , we obtain

$$e_k \leq \frac{2}{\sqrt{\beta}} \kappa_A^2 \left(\frac{C_1 + C_0 C_2 + \sqrt{2} U (D_1 + C_0 D_2)}{\sqrt{\beta}} + C_2 + \sqrt{2} D_2 U \right) h^{p_2-\frac{1}{2}}.$$

456

□

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

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

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

460 where (i) is due to the contraction assumption on Eq. (1).

461 Taking square roots on both sides, we obtain

$$\begin{aligned} W_2(\text{Law}(\bar{\mathbf{x}}_k), \mu) &\leq \sqrt{2e_k^2 + 2\exp(-2\beta t_k) W_2^2(\text{Law}(\mathbf{x}_0), \mu)} \\ &\leq \sqrt{2} \exp(-\beta kh) W_2(\text{Law}(\mathbf{x}_0), \mu) + \sqrt{2} e_k \end{aligned}$$

462 Invoking the conclusion of Theorem 3.3 completes the proof.

□

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

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

$$\sqrt{2} \exp(-\beta kh) W_2(\text{Law}(\mathbf{x}_0), \mu) \leq \frac{\epsilon}{2}. \quad (20)$$

$$Ch^{p_2-\frac{1}{2}} \leq \frac{\epsilon}{2} \quad (21)$$

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

$$k \geq \frac{1}{\beta h} \log \frac{2\sqrt{2} W_2(\text{Law}(\mathbf{x}_0), \mu)}{\epsilon} \triangleq k^* \quad (22)$$

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

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

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

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

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

$$k^* = \frac{(2C)^{\frac{1}{p_2 - \frac{1}{2}}}}{\beta} \frac{1}{\epsilon^{\frac{1}{p_2 - \frac{1}{2}}}} \log \frac{2\sqrt{2}W_2(\text{Law}(\mathbf{x}_0), \mu)}{\epsilon} = \tilde{\mathcal{O}}\left(\frac{C^{\frac{1}{p_2 - \frac{1}{2}}}}{\beta} \frac{1}{\epsilon^{\frac{1}{p_2 - \frac{1}{2}}}}\right)$$

471

□

472 B Proof of Results in Section 4

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

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

476 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
477 $C_0 = \frac{\sqrt{m}}{2}$, implied from Lemma C.3.

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

$$\tilde{C}_1 \leq 2(L^2 + G) \left(\frac{d}{4\kappa L} + \|\mathbf{x}_0\|^2 + \frac{8d}{7m} + 1 \right)^{\frac{1}{2}} \leq 2(L^2 + G) \sqrt{\frac{2d}{m} + \|\mathbf{x}_0\|^2 + 1} \triangleq C_1$$

481 and

$$\tilde{C}_2 \leq 2L \left(d + \frac{m}{2} \left(\|\mathbf{x}_0\|^2 + \frac{8d}{7m} \right) \right)^{\frac{1}{2}} \leq 2L\sqrt{m} \sqrt{\frac{2d}{m} + \|\mathbf{x}_0\|^2 + 1} \triangleq C_2$$

482 We collect all constants here in the proof for easier reference

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

483 Then the constant in Theorem 3.3 for LMC algorithm simplifies to

$$\begin{aligned} C &= \frac{2}{\sqrt{\beta}} \left(\frac{C_1 + C_0 C_2}{\sqrt{\beta}} + C_2 \right) \\ &\leq \frac{10(L^2 + G)}{m^{\frac{3}{2}}} \sqrt{2d + m(\|\mathbf{x}_0\|^2 + 1)} \triangleq C_{\text{LMC}} \end{aligned}$$

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

$$W_2(\text{Law}(\bar{\mathbf{x}}_k), \mu) \leq \sqrt{2}e^{-m\kappa h} W_2(\text{Law}(\mathbf{x}_0), \mu) + \sqrt{2}C_{\text{LMC}}h \quad (23)$$

486 for $0 < h \leq \frac{1}{4\kappa L}$.

□

487 **B.2 Proof of Theorem 4.3 (Lower Bound of Mixing Time)**

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

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

489 and hence

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

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

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

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

493 Suppose $W_2(\text{Law}(\bar{\mathbf{x}}_k), \mu) \leq \epsilon$, we then must have

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

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

494 A necessary condition of Eq. (25) is that

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

495 where (i) is due to Eq. (24). It follows from Eq. (26) that

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

496 Revisiting Eq. (24) yields

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

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

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

$$k \geq \frac{\sqrt{d}}{8m\epsilon} \log \frac{\sqrt{d}}{\epsilon} = \frac{\sqrt{d}}{8\epsilon} \log \frac{\sqrt{d}}{\epsilon} = \tilde{\Theta} \left(\frac{\sqrt{d}}{\epsilon} \right).$$

499

□

500 C Some Properties of Langevin Dynamics

501 C.1 Contraction of Langevin Dynamics

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

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

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

505 *Proof.* $\mathbf{x}_t, \mathbf{y}_t$ are respectively the solutions to

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

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

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

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

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

510 □

511 C.2 Growth Bound of Langevin Dynamics

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

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

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

515 *Proof.* We have

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

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

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

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

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

519

□

520 C.3 Bound on Evolved Deviation

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

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

523 with

$$\mathbb{E} \|\mathbf{z}\|^2 \leq \frac{m}{4} \|\mathbf{x} - \mathbf{y}\|^2 h$$

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

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

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

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

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

□

528 D Some Properties of LMC Algorithm

529 D.1 Local Strong Error

530 **Lemma D.1.** Suppose Assumption 1 holds. Denote the one-step iteration of LMC algorithm with
 531 step size h by $\bar{\mathbf{x}}_1$ and the solution of overdamped Langevin dynamics at time $t = h$ by \mathbf{x}_h . Both the

532 discrete algorithm and the continuous dynamics start from the same initial value \mathbf{x} . If $0 \leq h \leq \frac{1}{4\kappa L}$,
 533 then the local strong error of LMC algorithm satisfies

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

534 with $\tilde{C}_2 = 2L \left(d + \frac{m}{2} \|\mathbf{x}\|^2 \right)^{\frac{1}{2}}$.

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

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

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

538 D.2 Local Weak Error

539 **Lemma D.2.** Suppose Assumption 1 and 2 hold. Denote the one-step iteration of LMC algorithm
 540 with step size h by $\bar{\mathbf{x}}_1$ and the solution of overdamped Langevin dynamics at time $t = h$ by \mathbf{x}_h .
 541 Both the discrete algorithm and the continuous dynamics start from the same initial value \mathbf{x} . If
 542 $0 \leq h \leq \frac{1}{4\kappa L}$, then the local weak error of LMC algorithm satisfies

$$\|\mathbb{E}\bar{\mathbf{x}}_1 - \mathbb{E}\mathbf{x}_h\| \leq \tilde{C}_1 h^2$$

543 with $\tilde{C}_1 = 2(L^2 + G) \left(\frac{d}{4\kappa L} + \|\mathbf{x}\|^2 + 1 \right)^{\frac{1}{2}}$.

544 *Proof.* By Ito's lemma, we have

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

545 It follows that

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

546 where (i) is due to Assumption 2, (ii) is due to Jensen's inequality, (iii) is due to Lemma C.2, (iv)
547 is due to $\sqrt{a} + \sqrt{b} \leq \sqrt{2}\sqrt{a^2 + b^2}$ and (v) is due to $h \leq \frac{1}{4\kappa L}$. \square

548 D.3 Boundedness of LMC Algorithm

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

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

551 *Proof.* We have

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

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

555 Unfolding the inequality, we obtain

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

556

□