Convergence Rates for Langevin Monte Carlo in the Nonconvex Setting

Xiang Cheng X.CHENG@BERKELEY.EDU
Division of Computer Science
University of California Berkeley, CA, USA

Niladri S. Chatterji CHATTERJI@BERKELEY.EDU
Department of Physics
University of California Berkeley, CA, USA

Yasin Abbasi-Yadkori YASIN.ABBASI@GMAIL.COM
VinAI
Hanoi, Vietnam

Peter L. Bartlett PETER@BERKELEY.EDU
Division of Computer Science and Department of Statistics
University of California Berkeley, CA, USA

Michael I. Jordan JORDAN@CS.BERKELEY.EDU
Division of Computer Science and Department of Statistics
University of California Berkeley, CA, USA

Editor:

Abstract

We study the problem of sampling from a distribution $p^*(x) \propto \exp(-U(x))$, where the function U is L-smooth everywhere and m-strongly convex outside a ball of radius R, but potentially nonconvex inside this ball. We study both overdamped and underdamped Langevin MCMC and establish upper bounds on the number of steps required to obtain a sample from a distribution that is within ε of p^* in 1-Wasserstein distance. For the first-order method (overdamped Langevin MCMC), the iteration complexity is $\widetilde{\mathcal{O}}\left(e^{cLR^2} d/\varepsilon^2\right)$, where d is the dimension of the underlying space. For the second-order method (underdamped Langevin MCMC), the iteration complexity is $\widetilde{\mathcal{O}}\left(e^{cLR^2}\sqrt{d}/\varepsilon\right)$ for an explicit positive constant c. Surprisingly, the iteration complexity for both these algorithms is only polynomial in the dimension d and the target accuracy ε. It is exponential, however, in the problem parameter LR^2, which is a measure of non-log-concavity of the target distribution.

Keywords: Langevin Monte Carlo, Sampling algorithms, Nonconvex potentials

1. Introduction

We study the problem of sampling from a target distribution of the following form:

$$p^*(x) \propto \exp(-U(x)),$$

where $x \in \mathbb{R}^d$, and the *potential function* $U : \mathbb{R}^d \mapsto \mathbb{R}$ is L-smooth everywhere and m-strongly convex outside a ball of radius R (see detailed assumptions in Section 2.1).

Our focus is on theoretical rates of convergence of sampling algorithms, including analysis of the dependence of these rates on the dimension d. Much of the theory of convergence of sampling—for example, sampling based on Markov chain Monte Carlo (MCMC) algorithms—has focused on asymptotic convergence, and has stopped short of providing a detailed study of dimension dependence. In the allied field of optimization algorithms, a significant new literature has emerged in recent years on nonasymptotic rates, including tight characterizations of dimension dependence. The optimization literature, however, generally stops short of the kinds of inferential and decision-theoretic computations that are addressed by sampling, in domains such as Bayesian statistics (Robert and Casella, 2013), bandit algorithms (Cèsa-Bianchi and Lugosi, 2006) and adversarial online learning (Bubeck, 2011; Abbasi et al., 2013).

In both optimization and sampling, while the classical theory focused on convex problems, recent work focuses on the more broadly useful setting of nonconvex problems. While general nonconvex problems are infeasible, it is possible to make reasonable assumptions that allow theory to proceed while still making contact with practice.

We will consider the class of MCMC algorithms that have access to the gradients of the potential, $\nabla U(\cdot)$. A particular algorithm of this kind that has received significant recent attention from theoreticians is the *overdamped Langevin MCMC algorithm* (Parisi, 1981; Roberts and Tweedie, 1996). The underlying *first-order* stochastic differential equation (henceforth SDE) is given by:

$$dx_t = -\nabla U(x_t)dt + \sqrt{2}dB_t, \tag{1}$$

where B_t represents a standard Brownian motion in $\mathbb{R}^d$. Overdamped Langevin MCMC (Algorithm 1) is a discretization of this SDE. It is possible to show that under mild assumptions on U, the invariant distribution of the overdamped Langevin diffusion is given by $p^*(x)$.

The *second-order* generalization of the overdamped Langevin diffusion is *underdamped Langevin diffusion*, which can be represented by the following SDE:

$$\begin{aligned} dx_t &= u_t dt\,, \\ du_t &= -\lambda_1 u_t - \lambda_2 \nabla U(x_t)dt + \sqrt{2\lambda_1\lambda_2}dB_t, \end{aligned} \tag{2}$$

where $\lambda_1, \lambda_2 > 0$ are free parameters. This SDE can also be discretized appropriately to yield a corresponding MCMC algorithm (Algorithm 2). Second-order methods such as underdamped Langevin MCMC are particularly interesting as it has been previously observed both empirically (Neal, 2011) and theoretically (Cheng et al., 2017; Mangoubi and Smith, 2017) that these methods can be faster to converge than the classical first-order methods.

In this work, we show that it is possible to sample from p^* in time polynomial in the dimension d and the target accuracy ε (as measured in 1-Wasserstein distance). We also show that the convergence depends exponentially on the product LR^2. Intuitively, LR^2 is a measure of the nonconvexity of U. Our results establish rigorously that as long as the problem is not "too badly nonconvex," sampling is provably tractable.

Our main results are presented in Theorem 2 and Theorem 3, and can be summarized informally as follows:

Theorem 1 (informal) *Given a potential U that is L-smooth everywhere and strongly-convex outside a ball of radius R, we can output a sample from a distribution which is ε-close to $p^*(x) \propto \exp(-U(x))$ in W_1 distance by running $\widetilde{\mathcal{O}}\left(e^{cLR^2} d/\varepsilon^2\right)$ steps of overdamped Langevin MCMC (Algorithm 1), or $\widetilde{\mathcal{O}}\left(e^{cLR^2}\sqrt{d}/\varepsilon\right)$ steps of underdamped Langevin MCMC (Algorithm 2). Here, c is an explicit positive constant.*

For the case of strongly convex U, it has been shown by Cheng et al. (2017) that the iteration complexity of Algorithm 2 is $\widetilde{\mathcal{O}}(\sqrt{d}/\varepsilon)$, improving quadratically upon the best known iteration complexity of $\widetilde{\mathcal{O}}(d/\varepsilon^2)$ for Algorithm 1 (Durmus and Moulines, 2016). We will find this quadratic speed-up in d and ε in our setting as well (see Theorem 2 versus Theorem 3).

Related work: A convergence rate for overdamped Langevin diffusion, under assumptions (A1) – (A3) (see Section 2.1) has been established by Eberle (2016), but the continuous-time diffusion studied in that paper is not implementable algorithmically. In a more algorithmic line of work, Dalalyan (2017) bounded the discretization error of overdamped Langevin MCMC, and provided the first nonasymptotic convergence rate of overdamped Langevin MCMC under log-concavity assumptions. This was followed by a sequence of papers in the strongly log-concave setting (see, e.g., Durmus and Moulines, 2016; Cheng and Bartlett, 2017; Dalalyan and Karagulyan, 2017; Dwivedi et al., 2018).

Our result for overdamped Langevin MCMC is in line with this existing work; indeed, we combine the continuous-time convergence rate of Eberle (2016) with a variant of the discretization error analysis by Durmus and Moulines (2016). The final number of timesteps needed is $\widetilde{\mathcal{O}}(e^{cLR^2} d/\varepsilon^2)$, which is expected, as the rate of Eberle (2016) is $\mathcal{O}(e^{-cLR^2})$ (for the continuous-time process) and the iteration complexity established by Durmus and Moulines (2016) is $\widetilde{\mathcal{O}}(d/\varepsilon^2)$.

On the other hand, convergence of underdamped Langevin MCMC under (strongly) log-concave assumptions was first established by Cheng et al. (2017). Also very relevant to our results is the work of Eberle et al. (2017), who demonstrated a contraction property of the continuous-time process stated in Eq. (2). That result deals, however, with a much larger class of potential functions, and accordingly the distance to the invariant distribution scales exponentially with dimension d. Our analysis yields a more favorable result by combining ideas from both Eberle et al. (2017) and Cheng et al. (2017), under new assumptions; see Section 4 for a full discussion.

Also noteworthy is the fact that the problem of sampling from non-log-concave distributions has been studied by Raginsky et al. (2017), but under weaker assumptions, with a worst-case convergence rate that is exponential in d. In Xu et al. (2018), this technique is used to study the application of Stochastic Gradient Langevin Diffusion (and its variance-reduced version) to nonconvex optimization. Similarly, Durmus and Moulines (2017) analyze the overdamped Langevin MCMC algorithm under the assumption that U is superlinear outside a ball. This is more general than our assumption of "strong convexity outside a ball"; in this setting, the authors prove a rate that is exponential in dimension. On the other hand, Ge et al. (2017) established a $poly(d, 1/\varepsilon)$ convergence rate for sampling from a distribution that is close to a mixture of Gaussians, where the mixture components have the same variance (which is subsumed by our assumptions).

Finally, there is a large class of sampling algorithms known as Hamiltonian Monte Carlo (HMC), which involve Hamiltonian dynamics in some form. We refer to Ma et al. (2015) for a survey of the results in this area. Among these, the variant studied in this paper (Algorithm 2), based on the discretization of the SDE in Eq. (2), has a natural physical interpretation as the evolution of a particle's dynamics under a viscous force field. This model was first studied by Kramers (1940) in the context of chemical reactions. The continuous-time process has been studied extensively (Hérau, 2002; Villani, 2009; Eberle et al., 2017; Gorham et al., 2016; Baudoin, 2016; Bolley et al., 2010; Calogero, 2012; Dolbeault et al., 2015; Mischler and Mouhot, 2014). Four recent papers—Mangoubi and Smith (2017), Lee and Vempala (2017), Mangoubi and Vishnoi (2018) and Deligiannidis et al. (2018)—study the convergence rate of (variants of) HMC under log-concavity assumptions. In Eberle et al. (2019), the authors study the convergence of HMC on general metric state spaces. Bou-Rabee et al. (2018) study the convergence of HMC under assumptions similar to ours, and prove a convergence rate that depends on e^{cLR^2} for some constant c. We remark that the algorithm studied in this case is different from the underdamped Langevin MCMC algorithm, because of the incorporation of an accept-reject step.

2. Notation, definitions and assumptions

In this section, we present the basic definitions, notational conventions and assumptions used throughout the paper. For $q \in \mathbb{N}$ we let $\|v\|_q$ denote the q-norm of a vector $v \in \mathbb{R}^d$. Throughout the paper we use B_t to denote standard Brownian motion (see, e.g., Mörters and Peres, 2010).

2.1 Assumptions on the potential U

We make the following assumptions on the *potential function* U:

(A1) The function U is continuously-differentiable on $\mathbb{R}^d$ and has Lipschitz-continuous gradients; that is, there exists a positive constant $L > 0$ such that for all $x, y \in \mathbb{R}^d$,

$$\|\nabla U(x) - \nabla U(y)\|_2 \leq L\|x - y\|_2.$$

(A2) The function has a stationary point at zero:

$$\nabla U(0) = 0.$$

(A3) The function is strongly convex outside of a ball; that is, there exist constants $m, R > 0$ such that for all $x, y \in \mathbb{R}^d$ with $\|x - y\|_2 > R$, we have:

$$\langle \nabla U(x) - \nabla U(y), x - y \rangle \geq m\|x - y\|_2^2.$$

Finally we define the condition number as $\kappa := L/m$. Observe that Assumption (A2) is imposed without loss of generality, because we can always find a stationary point in polynomial time and shift the coordinate system so that this stationary point of U is at zero. These conditions are similar to the assumptions made by Eberle (2016). Note that crucially Assumption (A3) is *strictly stronger* than the assumption made in recent papers by Durmus and Moulines (2017), Raginsky et al. (2017) and Zhang et al. (2017). To see this observe that these papers only require Assumption (A3) to hold for a fixed $y = 0$, while we require this condition to hold for all $y \in \mathbb{R}^d$. One can also think of the difference between these two conditions as being analogous to the difference between strong convexity (outside a ball) and one-point strong convexity (outside a ball).

2.2 Coupling and Wasserstein distance

Denote by $\mathcal{B}(\mathbb{R}^d)$ the Borel σ-field of $\mathbb{R}^d$. Given probability measures μ and ν on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, we define a *transference plan* ζ between μ and ν as a probability measure on $(\mathbb{R}^d \times \mathbb{R}^d, \mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d))$ such that for all sets $A \in \mathcal{B}(\mathbb{R}^d)$, $\zeta(A \times \mathbb{R}^d) = \mu(A)$ and $\zeta(\mathbb{R}^d \times A) = \nu(A)$. We denote by $\Gamma(\mu, \nu)$ the set of all transference plans. A pair of random variables (X, Y) is called a *coupling* if there exists a $\zeta \in \Gamma(\mu, \nu)$ such that (X, Y) are distributed according to ζ. (With some abuse of notation, we will also refer to ζ as the coupling.)

3. Overdamped Langevin diffusion

In this section, we study *overdamped Langevin diffusion*, given by the following stochastic differential equation (SDE):

$$dy_t = -\nabla U(y_t)dt + \sqrt{2}dB_t. \tag{3}$$

It can be readily verified that the invariant distribution of the SDE is $p^*(y) \propto e^{-U(y)}$, which ensures that the marginal along y is the distribution that we are interested in. Based on Eq. (3), we define the discretized overdamped Langevin diffusion as

$$dx_t = -\nabla U\left(x_{\lfloor \frac{t}{\delta} \rfloor}\right)dt + \sqrt{2}dB_t, \tag{4}$$

where δ is the step-size of the discretization and $\lfloor \cdot \rfloor$ denotes the floor function.

Our first result, stated as Theorem 2, establishes the rate at which the distribution of the solution of Eq. (4) converges to p^*. The SDE in Eq. (4) is implementable as Algorithm 1. It can be verified that $x_{i\delta}$ in Algorithm 1 and the solution to the SDE in Eq. (4) at time $t = i\delta$ have the same distribution. The following theorem establishes a convergence rate for Algorithm 1.

Theorem 2 *Assume that* $m \geq \frac{\exp\left(-LR^2/2\right)}{R^2}$, *and let* $0 < \varepsilon \leq \frac{dR^2}{\sqrt{d/m+R^2}}$ *be the desired accuracy. Also let the initial point* $x^{(0)}$ *be such that* $\|x^{(0)}\|_2 \leq R$. *Then if the step size scales as:*

$$\delta = \frac{\varepsilon^2 \exp\left(-LR^2\right)}{2^{10}R^2 d},$$

Algorithm 1: Overdamped Langevin MCMC

Input : Step-size $\delta < 1$, number of iterations n, initial point $x_0 = x^{(0)}$, and gradient oracle $\nabla U(\cdot)$.

```
for i = 0, 1, ..., n − 1 do
    Sample x_{(i+1)δ} ~ N(x_{iδ} − δ∇U(x_{iδ}), 2δ I_{d×d})
end
```

and number of iterations scales as:

$$n = \widetilde{\Omega}\left(\exp\left(\frac{3LR^2}{2}\right) \cdot \frac{d}{\varepsilon^2}\right),$$

we have the following guarantee:

$$W_1(p_{n\delta}, p^*) \leq \varepsilon,$$

where $p_{n\delta}$ is the distribution of $x_{n\delta}$ in Algorithm 1 and the distribution $p^*(y) \propto e^{-U(y)}$.

For potentials where LR^2 is a constant, the number of iterations taken by overdamped MCMC scales as $\widetilde{\Omega}(d/\varepsilon^2)$. This matches the rate obtained in the strongly log-concave setting by Durmus and Moulines (2016).

Intuitively, LR^2 measures the extent of nonconvexity. When this quantity is large, it is possible for U to contain numerous local minima that are deep. It is therefore reasonable that the runtime of the algorithm should be exponential in this quantity.

The assumption on the strong convexity parameter, m, is made to simplify the presentation of the theorem. Note that this assumption is without loss of generality, since we can always take the radius R to be sufficiently large in Assumption (A3). Similarly, our assumption on the target accuracy can also be easily removed, but we make this assumption in the interest of clarity.

The proof of Theorem 2 is relegated to Appendix C. The proof follows by carefully combining the continuous-time argument of Eberle (2016) together with the discretization bound of Durmus and Moulines (2016).

4. Underdamped Langevin diffusion

In this section, we present our results for *underdamped Langevin diffusion.* The underdamped Langevin diffusion is a second-order stochastic process described by the following SDE:

$$dy_t = v_t dt, \tag{5}$$
$$dv_t = -2v_t - \frac{c_\kappa}{L}\nabla U(y_t)dt + \sqrt{\frac{4c_\kappa}{L}}dB_t,$$

where we define the constant:

$$c_\kappa := 1/(1000\kappa), \tag{6}$$

where $\kappa = L/m$ is the condition number. Similar to the case of overdamped Langevin diffusion, it can be verified that the invariant distribution of the SDE is $p^*(y,v) \propto e^{-U(y)-\frac{L}{2c_\kappa}\|v\|_2^2}$. This ensures that the marginal along y is the distribution that we are interested in. Based on the SDE in Eq. (5), we define the discretized underdamped Langevin diffusion as:

$$\begin{aligned} dx_t &= u_t dt, \\ du_t &= -2u_t - \frac{c_\kappa}{L}\nabla U\left(x_{\lfloor \frac{t}{\delta} \rfloor \delta}\right)dt + \sqrt{\frac{4c_\kappa}{L}}dB_t, \end{aligned} \tag{7}$$

where δ is the step size of discretization. The SDE in Eq. (7) is implementable as the following algorithm:

Algorithm 2: Underdamped Langevin MCMC

Input : Step-size $\delta < 1$, number of iterations n, initial point $(x^{(0)}, 0)$, smoothness parameter L, condition number κ and gradient oracle $\nabla U(\cdot)$.

for $i = 0, 1, \ldots, n-1$ **do**

  Sample $(x_{(i+1)\delta}, u_{(i+1)\delta}) \sim Z^{(i)}(x_{i\delta}, u_{i\delta})$

end

In this algorithm $Z^{(i)}(x_{i\delta}, u_{i\delta}) \in \mathbb{R}^{2d}$ is a Gaussian random vector with the following mean and covariance (which are functions of the previous iterates $(x_{i\delta}, u_{i\delta})$):

$$\begin{aligned}
&\mathbb{E}\left[u_{(i+1)\delta}\right] = u_{i\delta}e^{-2\delta} - \frac{c_\kappa}{2L}(1-e^{-2\delta})\nabla U(x_{i\delta}), \\
&\mathbb{E}\left[x_{(i+1)\delta}\right] = x_{i\delta} + \frac{1}{2}(1-e^{-2\delta})u_{i\delta} - \frac{c_\kappa}{2L}\left(\delta - \frac{1}{2}\left(1-e^{-2\delta}\right)\right)\nabla U(x_{i\delta}), \\
&\mathbb{E}\left[\left(x_{(i+1)\delta} - \mathbb{E}\left[x_{(i+1)\delta}\right]\right)\left(x_{(i+1)\delta} - \mathbb{E}\left[x_{(i+1)\delta}\right]\right)^\top\right] = \frac{c_\kappa}{L}\left[\delta - \frac{1}{4}e^{-4\delta} - \frac{3}{4} + e^{-2\delta}\right]\cdot I_{d\times d}, \\
&\mathbb{E}\left[\left(u_{(i+1)\delta} - \mathbb{E}\left[u_{(i+1)\delta}\right]\right)\left(u_{(i+1)\delta} - \mathbb{E}\left[u_{(i+1)\delta}\right]\right)^\top\right] = \frac{c_\kappa}{L}(1-e^{-4\delta})\cdot I_{d\times d}, \\
&\mathbb{E}\left[\left(x_{(i+1)\delta} - \mathbb{E}\left[x_{(i+1)\delta}\right]\right)\left(u_{(i+1)\delta} - \mathbb{E}\left[u_{(i+1)\delta}\right]\right)^\top\right] = \frac{c_\kappa}{2L}\left[1 + e^{-4\delta} - 2e^{-2\delta}\right]\cdot I_{d\times d}.
\end{aligned}$$

We show that the iterates at round i of Algorithm 2 and the solution to the SDE in Eq. (7) at time $t = i\delta$ have the same distribution (see Lemma 40 in Appendix H).

In Theorem 3, we establish a bound on the rate at which the distribution of the iterates produced by this algorithm converge to the target distribution p^*.

Theorem 3 *Assume that* $m \geq \frac{\exp\left(-6LR^2\right)}{64R^2}$ *and let* $0 < \varepsilon \leq \frac{dR^2}{\sqrt{d/m+R^2}}$ *be the desired accuracy. Also let the initial point* $x^{(0)}$ *be such that* $\|x^{(0)}\|_2 \leq R$. *Assume also that* $e^{72LR^2} \geq 2$.

Then if the step size scales as:

$$\delta = \frac{\varepsilon}{R + \sqrt{d/m}} \cdot e^{-12LR^2} \cdot 2^{-35}\min\left(\frac{1}{LR^2}, \frac{1}{\kappa}\right),$$

and the number of iterations as:

$$n = \widetilde{\Omega}\left(\frac{\sqrt{d}}{\varepsilon}\exp\left(18LR^2\right)\right),$$

we have the guarantee that

$$W_1(p_{n\delta}, p^*) \leq \varepsilon,$$

where $p_{n\delta}$ *is the distribution of* $x_{n\delta}$ *and we have* $p^*(y) \propto e^{-U(y)}$*.*

If we consider potentials for which LR^2 is a constant, the iteration complexity of underdamped Langevin MCMC grows as $\widetilde{\mathcal{O}}(\sqrt{d}/\varepsilon)$, which is a quadratic improvement over the first-order overdamped Langevin MCMC algorithm. Again, the iteration complexity grows exponentially in LR^2 which is to be expected. As before, the condition on the strong convexity parameter and the target accuracy is made in the interest of clarity and can be removed.

The heart of the proof of this theorem is a somewhat intricate coupling argument. We begin by defining two processes, (x_t, u_t) and (y_t, v_t), and then couple them appropriately. The first set of variables, (x_t, u_t), represent a solution to the discretized SDE in Eq. (7). On the other hand, the variables (y_t, v_t) represent a solution of the continuous-time SDE in Eq. (5) with the initial conditions being $(y_0, v_0) \sim p^*(y, v)$. Thus the variables (y_t, v_t) evolve according to the invariant distribution for all $t > 0$. The noise that underlies both processes is *coupled*, and with an appropriate choice of a Lyapunov function we are able to demonstrate that the distributions of these variables converge in 1-Wasserstein distance.

We present the coupling construction and a proof sketch in the subsequent sections. We relegate most of the technical details to the appendix.

4.1 A coupling construction

Let $\beta = 1/poly(L, 1/m, d, R, 1/C_m)$ be a small constant (see proof of Theorem 3 for the exact value), and let $\ell(x) = q(\|x\|_2)$ be a smoothed approximation of $\|x\|_2$ at a scale of β, as defined in (29).

Additionally, let $\nu = 1/poly(L, 1/m, d, R, 1/C_m)$ be another small constant (see proof of Theorem 3 for the exact value). In designing our coupling, we ensure that certain values are only updated at intervals of size ν. These are needed to ensure that the stochastic process that we work with is sufficiently regular.

While reading the proofs it might be convenient for the reader to think of both β and ν to be arbitrarily close to zero, and to think of $\ell(x)$ as equal to $\|x\|_2$; β and ν do not impact the bound on the iteration complexity in Theorem 3. For a detailed discussion see Appendix B.

We define a time T_{sync} as

$$T_{sync} := \frac{3\log 100}{c_\kappa^2}. \tag{8}$$

We then choose ν to be such that $\frac{T_{sync}}{\nu}$ is a positive integer, and define the constant

$$
\begin{aligned}
C_m :=& \min \left\{ \frac{e^{-6LR^2}}{6000\kappa(1+LR^2)}, \frac{e^{-6LR^2}}{200T_{sync}}, \frac{c_\kappa^2}{3} \right\} \\
=& \min \left\{ \frac{e^{-6LR^2}}{2^{13}\kappa(1+LR^2)}, \frac{e^{-6LR^2}}{2^{29}\cdot \log(100)\cdot \kappa^2}, \frac{1}{2^{22}\kappa^2} \right\}.
\end{aligned}
\tag{9}
$$

This constant C_m will be the rate at which our Lyapunov function contracts.

With these definitions in place we are ready to define a coupling between variables (x_t, u_t) that evolve according to the discretized process described in Eq. (11), and variables (y_t, v_t) that evolve according to the SDE in Eq. (13).

Let the initial conditions for these processes be given by,

$$
\begin{aligned}
(x_0, u_0) &= (x^{(0)}, 0), \\
(y_0, v_0) &\sim p^*(y, v).
\end{aligned}
\tag{10}
$$

Define a variable τ_t that will be useful in determining how the noise underlying the processes is coupled. We initialize this variable as follows: $\tau_0 = 0$, if $\sqrt{\|x_0 - y_0\|_2^2 + \|x_0 - y_0 + u_0 - w_0\|_2^2} \geq \sqrt{5}R$, and $\tau_0 = -T_{sync}$ otherwise.

Let A_t and B_t denote independent d-dimensional Brownian motions. We then let the complete set of variables $\left(x_t, u_t, y_t, v_t, \tau_{\lfloor \frac{t}{\nu} \rfloor}\right)$ evolve according to the following stochastic dynamics:

$$
dx_t = u_t dt \tag{11}
$$

$$
du_t = -2u_t dt - \frac{c_\kappa}{L}\nabla U\left(x_{\lfloor \frac{t}{\delta} \rfloor \delta}\right)dt + 2\sqrt{\frac{c_\kappa}{L}}dB_t \tag{12}
$$

$$
dy_t = v_t dt \tag{13}
$$

$$
\begin{aligned}
dv_t =& -2v_t - \frac{c_\kappa}{L}\nabla U(y_t)dt + 2\sqrt{\frac{c_\kappa}{L}}dB_t \\
&- \mathbb{1}\left\{k\nu \geq \tau_{\lfloor \frac{t}{\nu} \rfloor} + T_{sync}\right\} \cdot \left(4\sqrt{\frac{c_\kappa}{L}}\gamma_t\gamma_t^T dB_t + 2\sqrt{\frac{c_\kappa}{L}}\bar{\gamma}_t\bar{\gamma}_t^T dA_t\right),
\end{aligned}
\tag{14}
$$

where the functions $\mathcal{M}$, γ_t and $\bar{\gamma}_t$ are defined as follows:

$$
\begin{aligned}
\mathcal{M}(r) :=& \begin{cases} 1, & \text{for } r \in [\beta, \infty) \\ \frac{1}{2} + \frac{1}{2}\cos\left(r \cdot \frac{2\pi}{\beta}\right), & \text{for } r \in [\beta/2, \beta] \\ 0, & \text{for } r \in [0, \beta/2] \end{cases} \\
\gamma_t :=& (\mathcal{M}(\|z_t + w_t\|_2))^{1/2}\frac{z_t + w_t}{\|z_t + w_t\|_2} \\
\bar{\gamma}_t :=& \left(1 - (1 - 2\mathcal{M}(\|z_t + w_t\|_2))^2\right)^{1/4}\frac{z_t + w_t}{\|z_t + w_t\|_2},
\end{aligned}
\tag{15}
$$

and where for convenience we have defined

$$\begin{aligned} z_t &:= x_t - y_t \\ w_t &:= u_t - v_t. \end{aligned} \tag{16}$$

Note that the function $\mathcal{M}$ essentially is a Lipschitz approximation to the indicator function $\mathbb{1}\{r > 0\}$.

Let us unpack the definition of the SDE. First, note that when the indicator $\mathbb{1}\left\{k\nu \geq \tau_{\lfloor \frac{t}{\nu} \rfloor} + T_{sync}\right\}$ is equal to zero, then both (x_t, u_t) and (y_t, v_t) are evolved by the same Brownian motion B_t. This is called a *synchronous coupling* between the processes.

Second, when this indicator is equal to one, the processes are evolved by the same Brownian motion in the directions perpendicular to $z_t + w_t$, and (roughly) by the reflected Brownian motion along the direction $z_t + w_t$. This is called a *reflection coupling* between the two processes.

In the following lemma, we show that the variables (y_t, v_t) have the same marginal distributions as the solution to the SDE defined in Eq. (13).

Lemma 4 *The dynamics in defined by Eq.* (13) *and Eq.* (14) *is distributionally equivalent to the dynamics defined by Eq.* (5)*.*

We give the proof in Appendix H. It is easy to verify that (x_t, u_t) have the same marginal distribution as the solution to the SDE defined in Eq. (11) so we omit the proof.

Finally, we define an update rule for τ which dictates how the noise is coupled. For any $k \in \mathbb{Z}^+$, τ_k is defined as follows:

$$\tau_k := \begin{cases} k\nu & \text{if } \left(k\nu - \tau_{k-1} \geq T_{sync} \text{ AND } \sqrt{\|z_{k\nu}\|_2^2 + \|z_{k\nu} + w_{k\nu}\|_2^2} \geq \sqrt{5}R\right) \\ \tau_{k-1} & \textit{otherwise}. \end{cases} \tag{17}$$

From the dynamics in Eq. (14), we see that τ_k is used for determining whether (x_t, y_t, u_t, v_t) evolves by synchronous or reflection coupling over the interval $t \in [k\nu, (k+1)\nu)$. From its definition in Eq. (17), we see that, roughly speaking, τ_k is "the last time (up to $k\nu$) that (z_t, w_t) ends up outside the ball $\sqrt{\|z_t\|_2^2 + \|z_t + w_t\|_2^2} = \sqrt{5}R$," but with a caveat: we do not update the value of τ_k more than once in a T_{sync} interval of time.

Let $(\Omega, \mathcal{F}_t, P)$ be the probability space, where $\mathcal{F}_t$ is the σ-algebra generated by (y_0, v_0), B_s and A_s for all $s \in [0, t)$. In the following Lemma, we prove that $\left(x_t, u_t, v_t, y_t, \tau_{\lfloor \frac{t}{\nu} \rfloor}\right)$ has a unique strong solution $(x_t, u_t, y_t, v_t, \tau_{\lfloor \frac{t}{\nu} \rfloor})(\omega)$ $(\omega \in \Omega)$, which is adapted to the filtration $\mathcal{F}_t$. Furthermore, with probability one, $(x_t, u_t, y_t, v_t)(\omega)$ is t-continuous:

Lemma 5 *Let B_t and A_t be two independent Brownian motions, and let $\mathcal{F}_t$ be the σ-algebra generated by B_s, A_s; $s \leq t$, and (x_0, u_0, y_0, v_0).*

For all $t \geq 0$, the stochastic process $(x_t, u_t, y_t, v_t, \tau_{\lfloor \frac{t}{\nu} \rfloor})(\omega)$ defined in Eqs. (11)–(17) *has a unique solution such that (x_s, u_s, y_s, v_s) is t-continuous with probability one, and satisfies the following, for all $s \geq 0$,*

1. $(x_s, u_s, y_s, v_s, \tau_{\lfloor \frac{s}{\nu} \rfloor})$ *is adapted to the filtration* $\mathcal{F}_s$.

2. $\mathbb{E}\left[\|x_s\|_2^2 + \|y_s\|_2^2 + \|u_s\|_2^2 + \|v_s\|_2^2\right] \leq \infty.$

We defer the proof of this lemma to Appendix G.

Finally, for notational convenience, we define the following quantities, for any $k \in \mathbb{Z}^+$:

$$\mu_k := \mathbb{1}\left\{k\nu \geq \tau_k + T_{sync}\right\} \tag{18}$$

$$r_t := (1 + 2c_\kappa)\ell(z_t) + \ell(z_t + w_t) \tag{19}$$

$$\nabla_t := \nabla U(x_t) - \nabla U(y_t)$$

$$\Delta_t := \nabla U(x_{\lfloor \frac{t}{\delta} \rfloor \delta}) - \nabla U(x_t). \tag{20}$$

As described above, when $\mu_k = 0$ the processes are synchronously coupled, and when $\mu_k = 1$ they are coupled via reflection coupling. Roughly, r_t corresponds to the sum of $\|z_t\|_2$ and $\|z_t + w_t\|_2$. ∇_t is the difference of the gradients of U at x_t and y_t, while Δ_t is the difference of the gradients at $x_{\lfloor \frac{t}{\delta} \rfloor \delta}$ and x_t.

4.2 Lyapunov Function

In this section, we define a Lyapunov function that will be useful in demonstrating that the distributions of (x_t, u_t) and (y_t, v_t) converge in 1-Wasserstein distance.

We follow Eberle (2016) in our specification of the *distance function* f that is used in the definition of our Lyapunov function. We define two constants,

$$\alpha_f := \frac{L}{4}, \qquad \text{and,} \qquad \mathcal{R}_f := 12R, \tag{21}$$

and auxiliary functions $\psi(r)$, $\Psi(r)$ and $g(r)$, all from $\mathbb{R}^+$ to $\mathbb{R}^+$:

$$h(r) := \begin{cases} 1, & \text{for } r \in [0, \mathcal{R}_f] \\ 1 - \frac{1}{\mathcal{R}_f}(r - \mathcal{R}_f), & \text{for } r \in [\mathcal{R}_f, 2\mathcal{R}_f] \\ 0, & \text{for } r \in [2\mathcal{R}_f, \infty) \end{cases}$$

$$\psi(r) := e^{-2\alpha_f \int_0^r h(s)ds}, \qquad \Psi(r) := \int_0^r \psi(s)ds, \tag{22}$$

$$g(r) := 1 - \frac{1}{2}\frac{\int_0^r h(s)\frac{\Psi(s)}{\psi(s)}ds}{\int_0^\infty h(s)\frac{\Psi(s)}{\psi(s)}ds}.$$

Let us summarize some important properties of the functions ψ and g:

- ψ is decreasing, $\psi(0) = 1$, and $\psi(r) = \psi(2\mathcal{R}_f)$ for any $r > 2\mathcal{R}_f$.
- g is decreasing, $g(0) = 1$, and $g(r) = \frac{1}{2}$ for any $r > 2\mathcal{R}_f$.

Finally we define f as

$$f(r) := \int_0^r \psi(s)g(s)ds. \tag{23}$$

In Lemma 31 in Appendix E, we state and prove various several useful properties of the distance function f.

Additionally define the stochastic processes:

$$\xi_t = \int_0^t e^{-C_m(t-s)} c_\kappa \left\| x_s - x_{\lfloor \frac{s}{\delta} \rfloor \delta} \right\|_2 ds, \tag{24}$$

$$\sigma_t = \int_0^t \mu_{\lfloor \frac{s}{\nu} \rfloor} \cdot e^{-C_m(t-s)} \cdot \mathbb{1}\left\{ r_s \geq \sqrt{12} R \right\} 4 r_s ds, \tag{25}$$

$$\phi_t = \int_0^t \mu_{\lfloor \frac{s}{\nu} \rfloor} \cdot e^{-C_m(t-s)} \left\langle \nabla_{w_s}(f(r_s)), 4\sqrt{\frac{c_\kappa}{L}} \left(\gamma_s \gamma_s^T dB_s + \frac{1}{2} \bar{\gamma}_s \bar{\gamma}_s^T dA_s \right) \right\rangle. \tag{26}$$

These processes essentially track the discretization error arising due to a finite step size δ and ν. We refer to Lemma 38 in Appendix G for a proof of existence of ϕ_t.

Then following stochastic process $\mathcal{L}_t$ acts as our Lyapunov function:

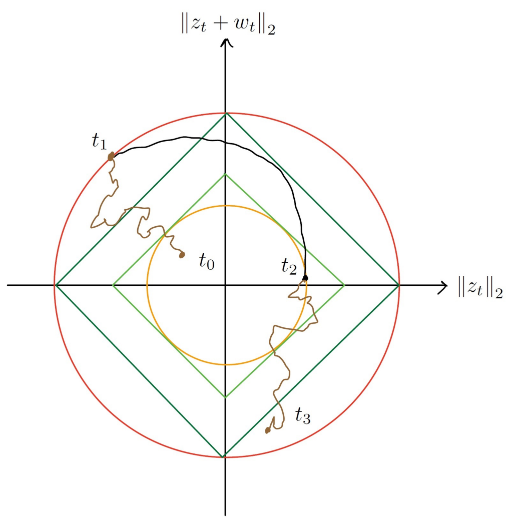
$$\mathcal{L}_t := \mu_k \cdot (f(r_t) - \xi_t) + (1 - \mu_k) \cdot \exp\left(-C_m(t - \tau_k)\right) \cdot (f(r_{\tau_k}) - \xi_{\tau_k}) - (\sigma_t + \phi_t), \tag{27}$$

where $k := \left\lfloor \frac{t}{\nu} \right\rfloor$. Note that $\mathcal{L}_t$ (the Lyapunov function at time t) depends on r_{τ_k} (at time τ_k). In Lemma 26, we demonstrate that this function contracts at a rate of $e^{-C_m t}$. The convergence bound then follows by showing that the convergence of this Lyapunov function implies convergence of the distributions in 1-Wasserstein distance.

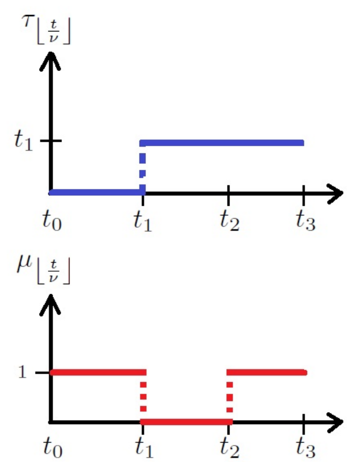
4.3 Proof Sketch

We present a full proof of Theorem 3 in Appendix D. In this section we provide a high-level sketch of our proof.

The proof proceeds by a path-wise analysis of the evolution of the Lyapunov function. In Figure 1b, we illustrate a sample path of the process.



(a) Illustration of Coupling



(b) Update for τ and ν

First, let us highlight the features of the figure.

1. The red circle represents the set $\sqrt{\|z_t\|_2^2 + \|z_t + w_t\|_2^2} = \sqrt{5}R$. It affects the updates of $\tau_{\lfloor \frac{t}{\nu} \rfloor}$, which, in turn, dictates how the processes are coupled.
2. The orange circle represents $\sqrt{\|z_t\|_2^2 + \|z_t + w_t\|_2^2} = \frac{23}{50} \cdot \sqrt{5}R$. In relation to the red circle, it represents the contraction of $\sqrt{\|z_t\|_2^2 + \|z_t + w_t\|_2^2}$ when evolved according to synchronous coupling.
3. The dark green diamond represents $(1 + 2c_\kappa)\|z_t\|_2 + \|z_t + w_t\|_2 = \sqrt{5}R$. It is a lower bound on $(1 + 2c_\kappa)\|z_t\|_2 + \|z_t + w_t\|_2 = \sqrt{5}R$ when $\sqrt{\|z_t\|_2^2 + \|z_t + w_t\|_2^2} = \sqrt{5}R$.
4. The light green diamond represents $2((1 + 2c_\kappa)\|z_t\|_2 + \|z_t + w_t\|_2) = 2 \cdot \frac{23}{50}\sqrt{5}R$. It represents an upper bound on $(1+2c_\kappa)\|z_t\|_2+\|z_t + w_t\|_2 = \sqrt{5}R$ when $\sqrt{\|z_t\|_2^2 + \|z_t + w_t\|_2^2} = \frac{23}{50} \cdot \sqrt{5}R$.
5. It is not drawn, but note that the red circle is contained in $(1+2c_\kappa)\|z_t\|_2+\|z_t + w_t\|_2 \le \sqrt{12}R$, which is the radius used for defining f in Eq. (21).
6. The brown squiggly lines $(t_0 \to t_1)$ and $(t_2 \to t_3)$ represent the evolution of the process under *reflection coupling*.
7. The black line $t_1 \to t_2$ represents the evolution of the process under *synchronous coupling*.

Below, we describe how (z_t, w_t) evolves over $t \in [t_0, t_3]$, and illustrate the main ideas behind the proof. To simplify matters, assume that

1. $k_i := t_i/\nu$ are integers, for $i = 0, 1, 2, 3$.
2. $t_3 - t_2 = T_{sync}$.
3. $\xi_t = \sigma_t = 0$ as these terms correspond to discretization errors.
4. $r_t \approx \|z_t\|_2 + (1 + 2c_\kappa)\|z_t + w_t\|_2$.

Then

- From $t_0 \to t_1$:
  Suppose that the process starts somewhere inside the red circle and stays inside for until time t_1, then $\tau_{\lfloor \frac{t}{\nu} \rfloor} = t_0$ and $\mu_{\lfloor \frac{t}{\nu} \rfloor} = 1$ for $t \in [t_0, t_1)$, and the process (z_t, w_t) undergoes reflection coupling.

  In this case, we can show that when $r_t \le \sqrt{12}R$ then $f(r_t) - \phi_t$ contracts at a rate of $\exp(-C_m t)$ with probability one (see Lemma 9). This in turn implies that our Lyapunov function $\mathcal{L}_t$ also contracts at the same rate with probability one (see Lemma 29 and Lemma 30).
- From $t_1 \to t_2$:
  At $t = t_1$, we update τ_{k_1} so that $\tau_{k_1} = t_1$. Thus $\mu_s = 0$ for all $s \in [t_1, t_2)$. Dur-

ing this period, (z_t, w_t) evolves under synchronous coupling. In Lemma 13, we show that $\sqrt{\|z_{t_2}\|_2^2 + \|z_{t_2} + w_{t_2}\|_2^2} \leq \frac{23}{50}\sqrt{\|z_{t_1}\|_2^2 + \|z_{t_1} + w_{t_1}\|_2^2}$. This implies that $f(r_{t_2}) \leq e^{-C_m(t_2-t_1)} f(r_{t_1})$ (Lemma 10). Again, this contraction is with probability one. Intuitively, we use synchronous coupling because when the value of $\|z_t\|_2 + \|z_t + w_t\|_2$ is large, Assumption (A3) guarantees contraction even in the absence of noise.

This contraction in f consequently results in a contraction of the Lyapunov function (see Lemma 28).

- After a duration T_{sync} of synchronous coupling, we have $\mu_{k_2} = 1$ and we resume reflection coupling over $[t_2, t_3]$. Note that at $t = t_2$, the Lyapunov function $\mathcal{L}_t$, undergoes a jump in value, from $\exp(-C_m(t_2 - t_1)) f(r_{t_1})$ to $f(r_{t_2})$ (see (27)). We show in Lemma 27 that this jump is negative with probability one.

5. Discussion

In this paper, we study algorithms for sampling from distributions which satisfy a more general structural assumption than log-concavity, in time polynomial in dimension and accuracy. We also demonstrate that when using underdamped dynamics the runtime can be improved, mirroring the strongly convex case.

There are a few natural questions that we hope to answer in further investigation of non-log-concave sampling problems. First, it would be interesting to determine other structural assumptions that may be imposed on the target distribution that are more general than log-concavity but still admit tractable sampling guarantees; for example, we would like to uncover assumptions that may alleviate the exponential dependence on LR^2. Conversely, existing guarantees may be extended to weaker assumptions, such as weak convexity outside a ball. Secondly, one might also wish to consider algorithms which have access to more than a gradient oracle, such as the Metropolis Hastings filter, or discretizations which use higher-order information.

Acknowledgements

This work was supported in part by the Mathematical Data Science program of the Office of Naval Research under grant number N00014-18-1-2764.

References

Yasin Abbasi, Peter L. Bartlett, Varun Kanade, Yevgeny Seldin, and Csaba Szepesvári. Online learning in Markov decision processes with adversarially chosen transition probability distributions. In *Advances in Neural Information Processing Systems*, pages 2508–2516, 2013.

Fabrice Baudoin. Wasserstein contraction properties for hypoelliptic diffusions. *arXiv preprint arXiv:1602.04177*, 2016.

Michael Betancourt, Simon Byrne, Sam Livingstone, and Mark Girolami. The geometric foundations of Hamiltonian Monte Carlo. *Bernoulli*, 23(4A):2257–2298, 2017.

Francois Bolley, Arnaud Guillin, and Florent Malrieu. Trend to equilibrium and particle approximation for a weakly self-consistent Vlasov-Fokker-Planck equation. *ESAIM: Mathematical Modelling and Numerical Analysis*, 44(5):867–884, 2010.

Nawaf Bou-Rabee, Andreas Eberle, and Raphael Zimmer. Coupling and convergence for Hamiltonian Monte Carlo. *ArXiv e-prints*, May 2018.

Sébastien Bubeck. Lecture notes: Introduction to online optimization, 2011.

Simone Calogero. Exponential convergence to equilibrium for kinetic Fokker-Planck equations. *Communications in Partial Differential Equations*, 37(8):1357–1390, 2012.

Nicolo Cèsa-Bianchi and Gábor Lugosi. *Prediction, Learning, and Games.* Cambridge University Press, 2006.

Xiang Cheng and Peter Bartlett. Convergence of Langevin MCMC in KL-divergence. *arXiv preprint arXiv:1705.09048*, 2017.

Xiang Cheng, Niladri S. Chatterji, Peter L. Bartlett, and Michael I. Jordan. Underdamped Langevin MCMC: A non-asymptotic analysis. *arXiv preprint arXiv:1707.03663*, 2017.

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

Arnak S. Dalalyan and Avetik G. Karagulyan. User-friendly guarantees for the Langevin Monte Carlo with inaccurate gradient. *arXiv preprint arXiv:1710.00095*, 2017.

Arnak S Dalalyan and Alexandre B Tsybakov. Sparse regression learning by aggregation and Langevin Monte-Carlo. *Journal of Computer and System Sciences*, 78(5):1423–1443, 2012.

George Deligiannidis, Daniel Paulin, Alexandre Bouchard-Côté, and Arnaud Doucet. Randomized Hamiltonian Monte Carlo as scaling limit of the bouncy particle sampler and dimension-free convergence rates. *arXiv preprint arXiv:1808.04299*, 2018.

Jean Dolbeault, Clément Mouhot, and Christian Schmeiser. Hypocoercivity for linear kinetic equations conserving mass. *Transactions of the American Mathematical Society*, 367(6): 3807–3828, 2015.

Sever S. Dragomir. *Some Gronwall Type Inequalities and Applications.* Nova Science Publishers, 2003.

Alain Durmus and Eric Moulines. Sampling from strongly log-concave distributions with the Unadjusted Langevin Algorithm. *arXiv preprint arXiv:1605.01559*, 2016.

Alain Durmus and Eric Moulines. Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *The Annals of Applied Probability*, 27(3):1551–1587, 2017.

Raaz Dwivedi, Yuansi Chen, Martin Wainwright, and Bin Yu. Log-concave sampling: Metropolis-Hastings algorithms are fast! *arXiv preprint arXiv:1801.02309*, 2018.

Andreas Eberle. Reflection couplings and contraction rates for diffusions. *Probability Theory and Related Fields*, 166(3-4):851–886, 2016. URL `https://doi-org.libproxy.berkeley.edu/10.1007/s00440-015-0673-1`.

Andreas Eberle, Arnaud Guillin, and Raphael Zimmer. Couplings and quantitative contraction rates for Langevin dynamics. *arXiv preprint arXiv:1703.01617*, 2017.

Andreas Eberle, Mateusz B Majka, et al. Quantitative contraction rates for markov chains on general state spaces. *Electronic Journal of Probability*, 24, 2019.

Rong Ge, Holden Lee, and Andrej Risteski. Beyond log-concavity: Provable guarantees for sampling multi-modal distributions using simulated tempering Langevin Monte Carlo. *arXiv preprint arXiv:1710.02736*, 2017.

Saul B. Gelfand and Sanjoy K. Mitter. Recursive stochastic algorithms for global optimization in $\mathbb{R}^d$. *SIAM Journal on Control and Optimization*, 29(5):999–1018, 1991.

Jack Gorham, Andrew B. Duncan, Sebastian J. Vollmer, and Lester Mackey. Measuring sample quality with diffusions. *arXiv preprint arXiv:1611.06972*, 2016.

Thomas H. Gronwall. Note on the derivatives with respect to a parameter of the solutions of a system of differential equations. *Annals of Mathematics*, 20:292–296, 1919.

Frédéric Hérau. Isotropic hypoellipticity and trend to the equilibrium for the Fokker-Planck equation with high degree potential. pages 1–13, 2002.

Hendrik A. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284–304, 1940.

Walid Krichene, Alexandre Bayen, and Peter Bartlett. Accelerated mirror descent in continuous and discrete time. In *Advances in Neural Information Processing Systems*, pages 2845–2853, 2015.

Yin Tat Lee and Santosh Vempala. Convergence rate of Riemannian Hamiltonian Monte Carlo and faster polytope volume computation. *arXiv preprint arXiv:1710.06261*, 2017.

Yi-An Ma, Tianqi Chen, and Emily Fox. A complete recipe for stochastic gradient MCMC. In *Advances in Neural Information Processing Systems*, pages 2917–2925, 2015.

Oren Mangoubi and Aaron Smith. Rapid mixing of Hamiltonian Monte Carlo on strongly log-concave distributions. *arXiv preprint arXiv:1708.07114*, 2017.

Oren Mangoubi and Nisheeth K Vishnoi. Dimensionally tight running time bounds for second-order Hamiltonian Monte Carlo. *arXiv preprint arXiv:1802.08898*, 2018.

Stéphane Mischler and Clément Mouhot. Exponential stability of slowly decaying solutions to the kinetic Fokker-Planck equation. *arXiv preprint arXiv:1412.7487*, 2014.

Peter Mörters and Yuval Peres. *Brownian Motion*. Cambridge University Press, 2010.

Radford M. Neal. MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo*. CRC Press, 2011.

Yurii Nesterov. A method of solving a convex programming problem with convergence rate $\mathcal{O}(1/k^2)$. *Soviet Mathematics Doklady*, 27(2):372–376, 1983.

Bernt Øksendal. *Stochastic Differential Equations: An Introduction with Wpplications*. Springer Science & Business Media, 2013.

Giorgio Parisi. Correlation functions and computer simulations. *Nuclear Physics B*, 180(3): 378–384, 1981.

Grigorios A. Pavliotis. *Stochastic Processes and Applications*. Springer, 2016.

Boris T Polyak. Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics*, 4(5):1–17, 1964.

Philip E. Protter. Stochastic differential equations. In *Stochastic Integration and Differential Equations*, pages 249–361. Springer, 2005.

Maxim Raginsky, Alexander Rakhlin, and Matus Telgarsky. Non-convex learning via stochastic gradient Langevin dynamics: A nonasymptotic analysis. In *Conference on Learning Theory*, pages 1674–1703, 2017.

Christian Robert and George Casella. *Monte Carlo Statistical Methods*. Springer Science & Business Media, 2013.

Gareth Roberts and Richard Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996.

John R Silvester. Determinants of block matrices. *The Mathematical Gazette*, 84(501): 460–467, 2000.

Weijie Su, Stephen Boyd, and Emmanuel Candès. A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights. In *Advances in Neural Information Processing Systems*, pages 2510–2518, 2014.

Cédric Villani. *Optimal Transport: Old and New*. Springer Science and Business Media, 2008.

Cédric Villani. *Hypocoercivity*. American Mathematical Society, 2009.

Andre Wibisono, Ashia C. Wilson, and Michael I. Jordan. A variational perspective on accelerated methods in optimization. *Proceedings of the National Academy of Sciences*, 113 (47):E7351–E7358, 2016. URL http://www.pnas.org/content/113/47/E7351.abstract.

Pan Xu, Jinghui Chen, Difan Zou, and Quanquan Gu. Global convergence of Langevin dynamics based algorithms for nonconvex optimization. In *Advances in Neural Information Processing Systems*, pages 3122–3133, 2018.

Yuchen Zhang, Percy Liang, and Moses Charikar. A hitting time analysis of stochastic gradient Langevin dynamics. In *Conference on Learning Theory*, pages 1980–2022, 2017.

Appendix

We outline here the organization of the Appendix.

In Appendix A, we list the variables used in this paper, in alphabetical order, with references to their definitions. In Appendix B, we give a description of two small constants, β and ν, which are used throughout our analysis to ensure regularity in time and space.

In Appendix C, we give a proof of Theorem 2. In Appendix D, we give a proof of Theorem 3.

In Appendix E, we specify the construction of the distance function f, which is used to demonstrate contraction. In Appendix F, we bound the moments of some of the relevant quantities; these are used in discretization bounds of Appendix C and D. In Appendix G, we gives proofs of the existence of our coupling constructions. In Appendix H, we prove that our coupling constructions have the correct marginals. We also prove that Algorithm 2 exactly implements (7).

Appendix A. Index of notation

α_f	Parameter of f (23). See (34)(overdamped) and (21) (underdamped).
β	Constant in defining ℓ. See also Section B.
c_κ	See (6)
C_m	Underdamped contraction rate, see (9).
C_o	Overdamped contraction rate, see (35).
d	Dimension of x
f	See (23)
κ	Condition number, defined after Assumption (A3)
ℓ	Twice continuously differentiable approximation to $\|\cdot\|_2$ with β error. See Lemma 6.
L	Lipschitz gradient parameter, see Assumption (A1).
$\mathcal{L}$	Lyapunov function. See (38) (overdamped) and (27) (underdamped)
m	Contraction parameter outside the R ball. See Assumption (A3)
$\mathcal{M}$	See (32) (overdamped) (15) (underdamped)
q	See Lemma 7
r	See (18).
R	See Assumption (A3)
$\mathcal{R}_f$	Parameter of f (23). See (34)(overdamped) and (21) (underdamped).
T_{sync}	See (8).
τ_k	See (17)
w_t	Short for $u_t - v_t$, defined in (16)
z_t	Short for $x_t - y_t$, defined in (16)
γ	See (15)
μ	See (18)
ν	Coupling stepsize in underdamped Coupling, used in (17)). See also Section B.
ξ	See (24).
σ	See (25).
ϕ	See (26).
∇_t and Δ_t	See (33) and (20)

Appendix B. Two Small Constants

On ℓ and β:
In this paper, we will take $\beta = 1/poly(L, 1/m, d, R)$ to be a small constant. See the proofs of Theorem 2 and Theorem 3 for the exact values of β. Intuitively, β is a radius inside of which we perform the following smoothing:

We define a function $q(r)$ in (28), which is a smoothed approximation of $|r|$, such that it has continuous second derivatives everywhere. Specifically, for $r \leq \beta/2$, $q(r)$ is a cubic spline.

$$q(r) = \begin{cases} \frac{\beta}{3} + \frac{8}{3\beta^2} \cdot r^3, & \text{for } r \in [0, \beta/4] \\ \frac{5\beta}{12} - r + \frac{4}{\beta} \cdot r^2 - \frac{8}{3\beta^2} \cdot r^3, & \text{for } r \in [\beta/4, \beta/2] \\ r, & \text{for } r \in [\beta/2, \infty]. \end{cases} \tag{28}$$

This allows us to define a smoothed version of $\|x\|_2$, which has continuous second derivatives everywhere:

$$\ell(x) = q(\|x\|_2). \tag{29}$$

In various parts of our proof, we replace $\|\cdot\|_2$ by its smooth approximation $\ell(\cdot)$, defined in Lemma 6, parametrized by β; a small β means that $\ell(\cdot)$ and $\|\cdot\|_2$ are close. We need to be careful as $\ell(\cdot)$ is strongly convex, with parameter $1/\beta^2$, in a $\beta/2$ radius around zero. We thus need to design our dynamics to ensure that the coupling has no noise in this region (see Eq. (15)).

When reading the proofs, it helps to think of $\ell(\cdot) = \|\cdot\|_2$ and $\beta = 0$, as we can take β to be arbitrarily small without additional computation costs. In our proof, it suffices to let $\beta = 1/poly(L, 1/m, d, R)$.

On ν:
In order to demonstrate the existence of a strong solution to the coupling presented in Section 4.1 (Lemma 5), we switch between synchronous and reflection coupling at deterministic, finite intervals of width ν.

This is not necessary strictly speaking, as there are results that ensure the existence of solutions of an SDE when the diffusion and drift coefficients are discontinuous but have finite variation. However, we choose to use a discretized coupling as the existence of its solution can be verified by using standard results.

This discretized coupling scheme adds an error term σ_t (see Eq. (25)). We show in Lemma 18 that this is $o(\nu^2)$.

When reading the proofs, it helps to think of $\nu = 0$ and $\sigma_t = 0$, as we can take ν to be arbitrarily small without additional computation costs. In the proof, it suffices to let $\nu = 1/poly(L, 1/m, d, R)$. See the proof and Theorem 3 for the exact value of ν.

Note that ν is distinct from (and unrelated to) δ, which is the step-size of the underdamped Langevin MCMC algorithm (Algorithm 2). δ, and the corresponding discretization error ξ_t, cannot be made arbitrarily small without additional computation costs.

Lemma 6 *For a given $\beta > 0$, let $m(r)$ be as defined in Lemma 7. Let $\ell(x) : \mathbb{R}^n \to \mathbb{R}^+$ be defined as in* (29)*, reproduced below for ease of reference:*

$$\ell(x) = q(\|x\|_2).$$

Then,

1. *For all x, $\ell(x)$ satisfies $\beta/3 \le \ell(x)$ and $|\ell(x) - \|x\|_2| \le \beta/3$. In addition, for $\|x\|_2 \ge \beta/2$, $\ell(x) = \|x\|_2$.*
2. *$\nabla \ell(x) = q'(\|x\|_2)\frac{x}{\|x\|_2}$, for all x, $\|\nabla \ell(x)\|_2 \le 1$, for $\|x\|_2 \ge \beta/2$, $\nabla \ell(x) = \frac{x}{\|x\|_2}$.*
3. *for $\|x\|_2 \ge \beta/2$, $\nabla^2 \ell(x) = q''(\|x\|_2)\frac{xx^T}{\|x\|_2^2} + q'(\|x\|_2)\frac{1}{\|x\|_2}\left(I - \frac{xx^T}{\|x\|_2^2}\right)$.*
4. *$\nabla \ell(x)$ and $\nabla^2 \ell(x)$ are defined everywhere and continuous. In particular, for $\|x\|_2 \le \beta/2$,*

$$\|\nabla \ell(x)\|_2 \le 4, \text{ and } \left\|\nabla^2 \ell(x)\right\|_2 \le \frac{8}{\beta}.$$

Proof

1. Immediate from Lemma 7.2.
2. By Chain rule, $\nabla \ell(x) = q'(\|x\|_2)\frac{x}{\|x\|_2}$. Furthermore, From the Lemma 7.1, we verify that $\nabla \ell(x)$ is defined everywhere, including at 0. The remaining claims follow from Lemma 7.3
3. This is just chain rule, together with Lemma 7.1, which guarantees the existence of $q''(\|x\|_2)/\|x\|_2^2$ for all x.
4. Existence and continuity follow from Lemma 7.

■

Lemma 7 *Let β be any positive real. Let $q(r)$ be defined as in* (28)*, reproduced below for ease of reference:*

$$q(r) = \begin{cases} \frac{\beta}{3} + \frac{8}{3\beta^2} \cdot r^3, & \text{for } r \in [0, \beta/4] \\ \frac{5\beta}{12} - r + \frac{4}{\beta} \cdot r^2 - \frac{8}{3\beta^2} \cdot r^3, & \text{for } r \in [\beta/4, \beta/2] \\ r, & \text{for } r \in [\beta/2, \infty]. \end{cases}$$

Then,

1. *$q(r)$, $q'(r)/r$ and $q''(r)/r^2$ exist for all r, and are continuous.*
2. *For all r, $q(r)$ satisfies $\beta/3 \le q(r)$ and $|r - q(r)| \le \beta/3$. In addition, $q(r) = r$ for $r \ge \beta/2$.*
3. *$q'(r)$ is monotonically nondecreasing, $q'(r) = 1$ for $r \ge \beta/2$, and $q'(r) = 0$ for $r = 0$.*
4. *$q''(r) = 0$ for all $r \ge \beta/2$.*

Proof Taking derivatives, we verify that

$$q'(r) = \begin{cases} \frac{8}{\beta^2} \cdot r^2, & \text{for } r \in [0, \beta/4] \\ -1 + \frac{8}{\beta} \cdot r - \frac{8}{\beta^2} \cdot r^2, & \text{for } r \in [\beta/4, \beta/2] \\ 1, & \text{for } r \in [\beta/2, \infty]; \end{cases}$$

$$q''(r) = \begin{cases} \frac{16}{\beta^2} \cdot r, & \text{for } r \in [0, \beta/4] \\ \frac{8}{\beta} - \frac{16}{\beta^2} \cdot r, & \text{for } r \in [\beta/4, \beta/2] \\ 0, & \text{for } r \in [\beta/2, \infty]. \end{cases}$$

All the claims can then be verified algebraically. ■

Appendix C. Proofs for overdamped Langevin Monte Carlo

C.1 Coupling construction for overdamped Langevin MCMC

Let β be a small constant (see proof of Theorem 2 for the exact value), and let $\ell(x) = q(\|x\|_2)$ be a smoothed approximation of $\|x\|_2$ as defined in (29). See Appendix B for a detailed discussion.

We begin by establishing the convergence of the continuous-time process in Eq. (1) to the invariant distribution. Similar to Eberle (2016), we construct a coupling between the SDEs described by Eq. (3) and Eq. (4). We initialize the coupling at

$$\begin{aligned} x_0 &= 0 \\ y_0 &\sim p^*(y), \end{aligned}$$

and evolve the pair (x_t, y_t) according to the dynamics

$$dx_t = -\nabla U\Big(x_{\lfloor \frac{t}{\delta} \rfloor \delta}\Big) dt + \sqrt{2} dB_t \tag{30}$$

$$dy_t = \nabla U(y_t) dt + \sqrt{2} dB_t - 2\sqrt{2}\gamma_t \gamma_t^T dB_t + \sqrt{2}\bar{\gamma}_t \bar{\gamma}_t^T dA_t, \tag{31}$$

where the terms γ_t and $\bar{\gamma}_t$ are defined as:

$$\begin{aligned} \gamma_t :=& (\mathcal{M}(\|z_t\|_2))^{1/2} \frac{z_t}{\|z_t\|_2} \\ \bar{\gamma}_t :=& \Big(1 - (1 - 2\mathcal{M}(\|z_t\|_2))^2\Big)^{1/4} \frac{z_t}{\|z_t\|_2}, \\ \text{with} \qquad z_t :=& x_t - y_t, \end{aligned}$$

$$\mathcal{M}(r) := \begin{cases} 1, & \text{for } r \in [\beta, \infty) \\ \frac{1}{2} + \frac{1}{2}\cos\Big(r \cdot \frac{2\pi}{\beta}\Big), & \text{for } r \in [\beta/2, \beta] \\ 0, & \text{for } r \in [0, \beta/2]. \end{cases} \tag{32}$$

We use the convention that $0/0 = 0$ when $\|z_t\|_2 = 0$. It can be verified that γ_t and $\bar{\gamma}_t$ are Lipschitz and gradient-Lipschitz for all $z_t \in \mathbb{R}^d$.

In the following Lemma, we show that y_t evolved according to Eq. (31) has the same marginal distributions as y_t evolved according to the SDE in Eq. (3).

Lemma 8 *The dynamics in Eq.* (31) *is distributionally equivalent to the dynamics defined in Eq.* (3).

We defer the proof to Appendix H.

For notational convenience, we define

$$\begin{aligned} \nabla_t :=& \nabla U(x_t) - \nabla U(y_t) \\ \Delta_t :=& \nabla U(x_{\lfloor \frac{t}{\delta} \rfloor \delta}) - \nabla U(x_t). \end{aligned} \tag{33}$$

Finally, we construct the Lyapunov function that we will use to show convergence. Let $f(r_t)$ be as defined in Eq. (23), with

$$\alpha_f := \frac{L}{4}, \quad \text{and,} \quad \mathcal{R}_f := R. \tag{34}$$

Define a constant,

$$C_o := \min\left\{\frac{1}{8R^2} e^{-LR^2/2}, m\right\}, \tag{35}$$

and finally, define two stochastic processes

$$\xi_t := L \int_0^t e^{-C_o(t-s)} \left\| x_s - x_{\lfloor \frac{s}{\delta} \rfloor \delta} \right\|_2 ds \tag{36}$$

$$\phi_t := \int_0^t e^{-C_o(t-s)} f'(\|z_s\|_2) \left\langle \frac{z_s}{\|z_s\|_2}, \left(2\sqrt{2}\gamma_s\gamma_s^T dB_s + \sqrt{2}\bar{\gamma}_s\bar{\gamma}_s^T dA_s\right) \right\rangle. \tag{37}$$

With these definitions, the following stochastic process $\mathcal{L}_t$ acts as our Lyapunov function:

$$\mathcal{L}_t := f(\ell(z_t)) - \xi_t - \phi_t. \tag{38}$$

C.2 Proof of Theorem 2

The proof follows in three steps. In Step 1 we analyze the evolution of $f(\ell(z_t))$ using Itô's Lemma. In Step 2 we use this to show that the Lyapunov function $\mathcal{L}_t$ which is defined in Eq. (38) contracts at a sufficiently fast rate. Finally in Step 3 we relate this contraction in the Lyapunov function to a bound on the iteration complexity of Algorithm 1.

We note that the technique in establishing Step 1 is essentially taken from Eberle (2016).

Step 1: By Itô's Lemma applied to $f(\ell(z_t))$,

$$df(\ell(z_t)) = \underbrace{\langle \nabla_z f(\ell(z_t)), -\nabla_t - \Delta_t \rangle}_{=:\spadesuit} dt + \underbrace{\frac{1}{2}\mathrm{tr}\big(\nabla_z^2 f(\ell(z_t))\big(8\gamma_t\gamma_t^T + 2\bar{\gamma}_t\bar{\gamma}_t^T\big)\big)}_{=:\heartsuit} dt$$
$$+ \left\langle \nabla_z f(\ell(z_t)), 2\sqrt{2}\gamma_t\gamma_t^T dB_t + \sqrt{2}\bar{\gamma}_t\bar{\gamma}_t^T dA_t \right\rangle.$$

We first bound the term $\spadesuit$. We can verify using Lemma 6 that $\nabla_z f(\ell(z_t)) = 0$ when $z_t = 0$. For the case when $\|z_t\|_2 \neq 0$ we have,

$$\nabla f(\ell(z_t)) = f'(\ell(z_t))q'(\|z_t\|_2)\frac{z_t}{\|z_t\|_2},$$

where $q(\cdot)$ is the function used to define ℓ (see Lemma 7). Thus

$$\begin{aligned}
\spadesuit &= \langle \nabla f(\ell(z_t)), -\nabla_t - \Delta_t \rangle \\
&= f'(\ell(z_t)) \cdot q'(z_t) \cdot \left\langle \frac{z_t}{\|z_t\|_2}, -\nabla_t - \Delta_t \right\rangle \\
&\overset{(i)}{\leq} f'(\ell(z_t)) \cdot q'(z_t) \left\langle \frac{z_t}{\|z_t\|_2}, -\nabla_t \right\rangle + \|\Delta_t\|_2 \\
&\overset{(ii)}{\leq} \mathbb{1}\left\{\|z_t\|_2 \in [0,\beta]\right\} \cdot L\beta + \mathbb{1}\left\{\|z_t\|_2 \in [R,\infty]\right\} \cdot (-m\|z_t\|_2) \\
&\quad + \mathbb{1}\left\{\|z_t\|_2 \in [\beta,R)\right\} \cdot f'(\ell(z_t)) \cdot (L\|z_t\|_2) + \|\Delta_t\|_2,
\end{aligned}$$

where (i) is by the Cauchy-Schwarz inequality, along with the fact that $|f'(r)| \leq 1$ (see (F2) of Lemma 31), and Lemma 7.3. The inequality in (ii) can be verified by considering three disjoint events. When $\|z_t\|_2 \in [0,\beta]$, the bound follows by Cauchy-Schwarz, (F2) of Lemma 31, combined with Lemma 7.3. While when $\|z_t\|_2 \in [R,\infty]$ the bound follows from Assumption (A3). When $\|z_t\|_2 \in [\beta,R]$, we bound the term using Cauchy-Schwarz, Assumption (A1), and Lemma 7.3.

Next, we consider the other term $\heartsuit = \frac{1}{2}\mathrm{tr}\big(\nabla_z^2 f(\ell(z_t))\big(8\gamma_t\gamma_t^T + 2\bar{\gamma}_t\bar{\gamma}_t^T\big)\big)$. We can verify using Lemma 6.4 that $\nabla^2 f(\ell(z_t)) = 0$ when $\|z_t\|_2 = 0$. Thus for $\|z_t\|_2 = 0$, the following holds:

$$\heartsuit = 0 = \mathbb{1}\left\{\|z_t\|_2 \in [\beta,R]\right\} 4f''(\ell(z_t)).$$

Alternatively, when $\|z_t\|_2 \neq 0$,

$$\begin{aligned}
\nabla^2 f(\ell(z_t)) =& f''(\ell(z_t))q'(\|z_t\|_2)^2\frac{z_t z_t^T}{\|z_t\|_2^2} + f'(\ell(z_t))q'(\|z_t\|_2)\frac{1}{\|z_t\|_2}\left(I - \frac{z_t z_t^T}{\|z_t\|_2^2}\right) \\
&+ f'(\ell(z_t))q''(\|z_t\|_2)\frac{z_t z_t^T}{\|z_t\|_2^2}.
\end{aligned}$$

Exapanding using the definition of $\heartsuit$,

$$\begin{aligned}\heartsuit =& \frac{1}{2}\mathrm{tr}\big(\nabla_z^2 f(\ell(z_t))8\gamma_t\gamma_t^T + 2\bar{\gamma}_t\bar{\gamma}_t^T\big)\\ \stackrel{(i)}{=}& \underbrace{\frac{1}{2}\mathrm{tr}\left(f''(\ell(z_t))\cdot q'(\|z_t\|_2)^2\cdot\frac{z_tz_t^T}{\|z_t\|_2^2}\big(8\gamma_t\gamma_t^T + 2\bar{\gamma}_t\bar{\gamma}_t^T\big)\right)}_{=:\heartsuit_1}\\ &+\underbrace{\frac{1}{2}\mathrm{tr}\left(f'(\ell(z_t))\cdot q'(\|z_t\|_2)\cdot\frac{1}{\|z_t\|_2}\left(I-\frac{z_tz_t^T}{\|z_t\|_2^2}\right)\big(8\gamma_t\gamma_t^T + 2\bar{\gamma}_t\bar{\gamma}_t^T\big)\right)}_{=:\heartsuit_2}\\ &+\underbrace{\frac{1}{2}\mathrm{tr}\left(f'(\ell(z_t))q''(z_t)\frac{z_tz_t^T}{\|z_t\|_2^2}\big(8\gamma_t\gamma_t^T + 2\bar{\gamma}_t\bar{\gamma}_t^T\big)\right)}_{=:\heartsuit_3},\end{aligned}$$

where (i) is by the expression for $\nabla^2 f(\ell(z_t))$ above.

Before proceeding, we verify by definition of γ_t and $\bar{\gamma}_t$ in Eq. (15) that

$$\mathrm{tr}\left(\frac{z_tz_t^T}{\|z_t\|_2^2}\big(8\gamma_t\gamma_t^T + 2\bar{\gamma}_t\bar{\gamma}_t^T\big)\right) = 8\|\gamma_t\|_2^2 + 2\|\bar{\gamma}_t\|_2^2. \tag{39}$$

First we simplify $\heartsuit_1$:

$$\begin{aligned}\heartsuit_1 =& \frac{1}{2}f''(\ell(z_t))\cdot q'(\|z_t\|_2)^2\cdot\Big(8\|\gamma_t\|_2^2 + 2\|\bar{\gamma}_t\|_2^2\Big)\\ \stackrel{(i)}{\leq}& \mathbb{1}\left\{\|z_t\|_2\in[\beta,R]\right\}\Big(f''(\ell(z_t))\cdot\Big(4\|\gamma_t\|_2^2 + \|\bar{\gamma}_t\|_2^2\Big)\Big)\\ \stackrel{(ii)}{=}& \mathbb{1}\left\{\|z_t\|_2\in[\beta,R]\right\}4f''(\ell(z_t)),\end{aligned}$$

where the inequality (i) is because $f''(r)\geq 0$ for all $r>0$ (by Lemma 31.(F5)), $q'(r)\geq 0$ for all r (by Lemma 7.3) and $q'(r)=1$ for all $r\geq\beta/2$ (Lemma 7.3). The equality in (ii) is because $\gamma_t = 1$ for $\|z_t\|_2\geq\beta$ (by its definition in Eq. (32)).

Next, using Eq. (39), we can immediately verify that $\heartsuit_2 = 0$.

Finally, we focus on $\heartsuit_3$,

$$\heartsuit_3 = \frac{1}{2}f'(\ell(z_t))q''(z_t)\Big(8\|\gamma_t\|_2^2 + 2\|\bar{\gamma}_t\|_2^2\Big) = 0,$$

where we use the fact that $q''(\|z_t\|_2) = 0$ if $\|z_t\|_2\geq\beta/2$ (by Lemma 7.4) and $\gamma_t=\bar{\gamma}_t=0$ if $\|z_t\|_2\leq\beta/2$ (by its definition in Eq. (32)).

Putting together the bounds on $\heartsuit_1$, $\heartsuit_2$ and $\heartsuit_3$, we can upper bound $\heartsuit$ as

$$\heartsuit\leq\mathbb{1}\left\{\|z_t\|_2\in[\beta,R]\right\}4f''(\ell(z_t)).$$

Combining the upper bounds on ♠ and ♡,

$$\spadesuit + \heartsuit \le \underbrace{\mathbb{1}\left\{\|z_t\|_2 \in [\beta, R]\right\}\big(L\|z_t\|_2 f'(\ell(z_t)) + 4f''(\ell(z_t))\big)}_{=:\clubsuit} \\ + \mathbb{1}\left\{\|z_t\|_2 \in [R, \infty]\right\} \cdot (-m\|z_t\|_2) + \|\Delta_t\|_2 + L\beta.$$

Let us now focus on ♣. By Lemma 31,

$$\begin{aligned}
\clubsuit =& \mathbb{1}\left\{\|z_t\|_2 \in [\beta, R]\right\} \cdot \big(L\|z_t\|_2 f'(\ell(z_t)) + 4f''(\ell(z_t))\big) \\
\overset{(i)}{\le}& \mathbb{1}\left\{\|z_t\|_2 \in [\beta, R]\right\} \cdot \big(L \cdot \ell(z_t) f'(\ell(z_t)) + 4f''(\ell(z_t)) + L\beta/3\big) \\
\overset{(ii)}{\le}& \mathbb{1}\left\{\|z_t\|_2 \in [\beta, R]\right\} \cdot (-C_o f(\ell(z_t))) + L\beta/3 \\
\overset{(iii)}{\le}& \mathbb{1}\left\{\|z_t\|_2 \in [0, R]\right\}(-C_o f(\ell(z_t))) + (L + 8C_o)\beta \\
\overset{(iv)}{\le}& \mathbb{1}\left\{\|z_t\|_2 \in [0, R]\right\}(-C_o f(\ell(z_t))) + 10L\beta,
\end{aligned}$$

where (i) is because $|\|z_t\|_2 - \ell(z_t)| \le \beta/3$ (by Lemma 6.1) and because $|f'(r)| \le 1$ for all $r > 0$ (by Lemma 31.(F2)). The inequality in (ii) is by Lemma 31 (F4), our definition of C_o in (35), and the fact that $\|z_t\|_2 \in [\beta, R]$ implies $\ell(z_t) \le R$ (Lemma 6.1). Inequality (iii) is again by Lemma 6.1 and Lemma 31 (F3). Finally, (iv) is by (35) and $m \le L$, a consequence of Assumptions (A1) and (A3).

Thus,

$$\begin{aligned}
\spadesuit + \heartsuit \le& \mathbb{1}\left\{\|z_t\|_2 \in [0, R]\right\}(-C_o f(\ell(z_t))) + \mathbb{1}\left\{\|z_t\|_2 \in [R, \infty]\right\} \cdot (-m\|z_t\|_2) + 11L\beta + \|\Delta_t\|_2 \\
\le& - C_o f(\ell(z_t)) + 12L\beta + \|\Delta_t\|_2,
\end{aligned}$$

where the second line is by Lemma 6.1 and 31.(F3), and by $m \le L$.

Putting this together with the expression for $df(\ell(z_t))$,

$$\begin{aligned}
df(\ell(z_t)) \le& (-C_o f(\ell(z_t)) + 12L\beta + \|\Delta_t\|_2)dt + \left\langle \nabla_z f(\ell(z_t)), 2\sqrt{2}\gamma_t\gamma_t^T dB_t + \sqrt{2}\bar{\gamma}_t\bar{\gamma}_t^T dA_t \right\rangle \\
\le& \Big(-C_o f(\ell(z_t)) + 12L\beta + L\left\|x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta}\right\|_2\Big)dt \\
&+ \left\langle \nabla_z f(\ell(z_t)), 2\sqrt{2}\gamma_t\gamma_t^T dB_t + \sqrt{2}\bar{\gamma}_t\bar{\gamma}_t^T dA_t \right\rangle.
\end{aligned}$$

The second inequality uses the definition of Δ_t in Eq. (33) and Assumption (A1).

Step 2: If we consider the evolution of the Lyapunov function $\mathcal{L}_t$ (defined in Eq. (38)), we can verify that

$$\begin{aligned}
d\mathcal{L}_t =& d(f(\ell(z_t)) - \phi_t - \xi_t) \\
\overset{(i)}{\le}& - C_o(f(\ell(z_t)) - \phi_t - \xi_t)dt + 12L\beta dt \\
=& - C_o\mathcal{L}_t dt + 12L\beta dt,
\end{aligned}$$

where the simplification in inequality (i) can be verified by taking time derivatives of stochastic processes ϕ_t and ξ_t defined in Eq. (37) and Eq. (36).

Applying Grönwall's inequality,

$$\mathcal{L}_t \leq e^{-C_o t}\mathcal{L}_0 + \int_0^t e^{-C_o(t-s)} 12L\beta ds \leq e^{-C_o t}\mathcal{L}_0 + \frac{12L\beta}{C_o}.$$

Using the definition of $\mathcal{L}_t$ in Eq. (38) we get,

$$f(\ell(z_t)) \leq e^{-C_o t} f(\ell(z_0)) + \xi_t + \phi_t.$$

Taking expectations with respect to the Brownian motion yields:

$$\mathbb{E}\left[f(\ell(z_t))\right] \leq e^{-C_o t}\mathbb{E}\left[f(\ell(z_0))\right] + \mathbb{E}\left[\xi_t\right] + \mathbb{E}\left[\phi_t\right]. \tag{40}$$

By the definition of ϕ_t in Eq. (37), we verify that $\mathbb{E}\left[\phi_t\right] = 0$, and by definition of ξ_t in Eq. (36),

$$\begin{aligned}
\mathbb{E}\left[\xi_t\right] &= \int_0^t e^{-C_o(t-s)}\mathbb{E}\left[\left\|x_t - x_{\lfloor \frac{t}{\delta}\rfloor\delta}\right\|_2\right] ds \\
&\leq \int_0^t e^{-C_o(t-s)}\mathbb{E}\left[\left\|\left(s - \left\lfloor\frac{s}{\delta}\right\rfloor\delta\right)\nabla U(x_{\lfloor \frac{s}{\delta}\rfloor\delta}) + \int_{\lfloor \frac{s}{\delta}\rfloor\delta}^s dB_r\right\|_2\right] ds \\
&\leq \int_0^t e^{-C_o(t-s)}\left(\mathbb{E}\left[\delta L\left\|x_{\lfloor \frac{s}{\delta}\rfloor\delta}\right\|_2\right] + \sqrt{\delta d}\right) ds \\
&\leq \int_0^t e^{-C_o(t-s)}\left(2\delta L\sqrt{R^2 + d/m} + \sqrt{\delta d}\right) ds \\
&\leq \frac{2\delta L\sqrt{R^2 + d/m} + \sqrt{\delta d}}{C_o}
\end{aligned}$$

We can also bound the initial value of $\mathbb{E}\left[f(\ell(z_0))\right]$ as follows:

$$\mathbb{E}\left[f(\ell(z_0))\right] \overset{(i)}{=} \mathbb{E}\left[f(\ell(y_0))\right] \overset{(ii)}{\leq} \mathbb{E}\left[\ell(y_0)\right] \overset{(iii)}{\leq} \mathbb{E}\left[\|y_0\|_2\right] + \beta/3 \overset{(iv)}{\leq} \sqrt{R^2 + \frac{d}{m}} + \beta/3,$$

where (i) is because $x^(0) = 0$ in Eq. (30), (ii) is by Lemma 31.(F3), (iii) is by Lemma 6.1, and finally (iv) is by Lemma 37.

Let n be the number of time steps, so that $t = n\delta$. Substituting into the inequality in Eq. (40), we get

$$\mathbb{E}\left[f(\ell(z_{n\delta}))\right] \leq e^{-C_o(n\delta)}\left(32\sqrt{R^2 + \frac{d}{m}} + \beta/3\right) + \frac{2\delta L\sqrt{R^2 + d/m} + \delta d}{C_o} + \frac{12L\beta}{C_o}.$$

Step 3: We translate our bound on $\mathbb{E}\left[f(\ell(z_{n\delta}))\right]$ to a bound on $\mathbb{E}\left[\|z_{n\delta}\|_2\right]$, which implies a bound in 1-Wasserstein distance. By Lemma 31(F3),

$$
\begin{aligned}
&\mathbb{E}\left[\|z_{n\delta}\|_2\right] \\
&\quad\le 2e^{L(R+\beta)^2/2}\left(e^{-C_o(n\delta)}\left(32\sqrt{R^2+\frac{d}{m}}+\beta/3\right)+\frac{2\delta L\sqrt{R^2+d/m}+\delta d}{C_o}+\frac{12L\beta}{C_o}\right)\\
&\quad\le 4e^{LR^2/2}\left(e^{-C_o(n\delta)}\left(32\sqrt{R^2+\frac{d}{m}}\right)+\frac{2\delta L\sqrt{R^2+d/m}+\delta d}{C_o}\right),
\end{aligned}
$$

where for the second inequality, it suffices to let $\beta = \delta d/6$

For a given ε, the first term is less than $\varepsilon/2$ if

$$
n\delta \ge 10\left(\log\left(\frac{R^2+d/m}{\varepsilon}\right)+LR^2\right)\cdot\frac{1}{C_o}.
$$

The second term is less than $\varepsilon/2$ if

$$
\delta \le \frac{1}{10}e^{-LR^2/2}\min\left\{\frac{\varepsilon}{\sqrt{R^2+d/m}},\frac{\varepsilon^2 C_o}{d}\right\}.
$$

By the definition of C_o in Eq. (35),

$$
C_o \le \frac{1}{8}\min\left\{\frac{\exp\left(-LR^2/2\right)}{R^2},m\right\} = \frac{\exp\left(LR^2/2\right)}{8R^2},
$$

where the equality is by our assumption on the strong convexity parameter m in the theorem statement. Recall that we also assume that $\varepsilon \le \frac{dR^2}{\sqrt{d/m+R^2}}$. Thus we can verify that

$$
\min\left\{\frac{\varepsilon}{\sqrt{R^2+d/m}},\frac{\varepsilon^2 C_o}{d}\right\} = \frac{\varepsilon^2 C_o}{d}.
$$

Putting everything together, we obtain a guarantee that $\mathbb{E}\left[\|z_t\|_2\right] \le \varepsilon$ if

$$
\delta = \frac{\varepsilon^2\exp\left(-LR^2\right)}{2^{10}R^2 d},
$$

and

$$
n \ge 2^{18}\log\left(\frac{R^2+d/m}{\varepsilon}\right)\cdot R^4\cdot\exp\left(\frac{3LR^2}{2}\right)\cdot\frac{d}{\varepsilon^2},
$$

as prescribed by the theorem statement.

Appendix D. Proofs for Underadmped Langevin Monte Carlo

D.1 Overview

The main idea behind the proof is to show that $\mathcal{L}_t$ contracts with probability one by a factor of $e^{-C_m\nu}$, going from $t = (k-1)\nu$ to $t = k\nu$. The result can be found in Lemma 26 in Section D.5. The proof considers four cases:

1. $\mu_{k-1} = 1, \mu_k = 1$. In Lemma 29 in Section D.5, we show that $\mathcal{L}_{k\nu} \le e^{-C_m\nu}\mathcal{L}_{(k-1)\nu}$. The proof of this result in turn uses Lemma 9 in Section D.2, which shows that $\mathcal{L}_t$ contracts at a rate of $-C_m$ over the interval $t \in [(k-1)\nu, k\nu]$.

2. $\mu_{k-1} = 1, \mu_k = 0$. In Lemma 30 in Section D.5, we show that $\mathcal{L}_{k\nu} \le e^{-C_m\nu}\mathcal{L}_{(k-1)\nu}$. The proof of this result is almost identical to the preceding case $\mu_{k-1} = 1, \mu_k = 1$. (In particular, $\mathcal{L}_t$ undergoes no jump in value at $t = k\nu$, in spite in the change in value from $\mu_{k-1} = 1$ to $\mu_k = 0$. See proof for details.)

3. $\mu_{k-1} = 0, \mu_k = 0$. In Lemma 28 in Section D.5, we show that $\mathcal{L}_{k\nu} \le e^{-C_m\nu}\mathcal{L}_{(k-1)\nu}$. The proof of this result is mainly based on the definition of $\mathcal{L}_t$.

4. $\mu_{k-1} = 0, \mu_k = 1$. In Lemma 27 in Section D.5, we show that $\mathcal{L}_{k\nu} \le e^{-C_m\nu}\mathcal{L}_{(k-1)\nu}$. This case is somewhat tricky, as $\mathcal{L}_t$ undergoes a jump in value at $t = k\nu$. Specifically, $\mathcal{L}_t$ jumps from $e^{-C_m T_{sync}}\big(f(r_{\tau_{k-1}}) - \xi_{\tau_{k-1}}\big) - (\sigma_{k\nu} + \phi_{k\nu})$ to $f(r_{k\nu}) - \xi_{k\nu} - (\sigma_{k\nu} + \phi_{k\nu})$. We prove that this jump is always negative (Lemma 10, Section D.3). The proof of Lemma 12 in turn relies on a contraction result in Lemma 13.

Having proven Lemma 26, we prove Theorem 3 by applying Lemma 26 recursively, and showing that $\mathbb{E}[\mathcal{L}_t]$ sandwiches the Wasserstein distance $W_1(p_t, p^*)$.

D.2 Contraction under Reflection Coupling

Our main result is stated as Lemma 9. It shows that $\mu_k f(r_t)$ contracts at a rate of $\exp(-C_m t)$, plus some discretization error terms.

Lemma 9 *For any positive integer k, with probability one we have,*

$$\begin{aligned}
&\mu_k \cdot \big(f(r_{(k+1)\nu}) - \xi_{(k+1)\nu}\big) - \big(\sigma_{(k+1)\nu} + \phi_{(k+1)\nu}\big) \\
&\qquad \le e^{-C_m\nu}(\mu_k \cdot (f(r_{k\nu}) - \xi_{k\nu}) - (\sigma_{k\nu} + \phi_{k\nu})) + 5\beta\nu.
\end{aligned}$$

Proof If $\mu_k = 0$, both sides of the inequality are identically zero. To simplify notation, we leave out the factor of μ_k in subsequent expressions and assume that $\mu_k = 1$ unless otherwise stated.

For the rest of this proof, we will consider time $s \in [k\nu, (k+1)\nu)$ for some k.

Let us first establish some useful derivatives of the function f:

$$\begin{aligned}
\nabla_z f(r_s) &= f'(r_s)\cdot(1+2c_\kappa)q'(\|z_s\|_2)\cdot\frac{z_s}{\|z_s\|_2} + f'(r_s)\cdot q'(z_s+w_s)\cdot\frac{z_s+w_s}{\|z_s+w_s\|_2},\\
\nabla_w f(r_s) &= f'(r_s)\cdot q'(\|z_s+w_s\|_2)\cdot\frac{z_s+w_s}{\|z_s+w_s\|_2},\\
\nabla_w^2 f(r_s) &= f'(r_s)\cdot q'(\|z_s+w_s\|_2)\cdot\frac{1}{\|z_s+w_s\|_2}\left(I-\frac{(z_s+w_s)(z_s+w_s)^T}{\|z_s+w_s\|_2^2}\right)\\
&\quad + f'(r_s)q''(\|z_s+w_s\|_2)\frac{(z_s+w_s)(z_s+w_s)^T}{\|z_s+w_s\|_2^2}\\
&\quad + f''(r_s)q'(\|z_s+w_s\|_2)^2\frac{(z_s+w_s)(z_s+w_s)^T}{\|z_s+w_s\|_2^2}. \qquad (41)
\end{aligned}$$

The derivatives follow from Lemma 6 and by the definition of r_t in Eq. (19). From Lemma 6.3, $\nabla_w^2 f((1+2c_\kappa)\ell(z)+\ell(z+w))$ exists everywhere and is continuous, with $\nabla_w^2 f((1+2c_\kappa)\ell(z)+\ell(z+w))\big|_{z+w=0}=0$. Note that, we use the convention $0/0=0$.

For any $s\in[k\nu,(k+1)\nu)$, we have:

$$\begin{aligned}
d\mu_k\cdot f(r_s) &\overset{(i)}{=} \mu_k\cdot\langle\nabla_z f(r_s), dz_s\rangle + \langle\nabla_w f(r_s), dw_s\rangle\\
&\quad + \mu_k\cdot\frac{8c_\kappa}{L}\gamma_s^T\nabla_w^2 f(r_s)\gamma_s ds + \frac{2c_\kappa}{L}\bar{\gamma}_s^T\nabla_w^2 f(r_s)\bar{\gamma}_s ds\\
&\overset{(ii)}{=} \mu_k\cdot\underbrace{\left(\langle\nabla_z f(r_s), w_s\rangle + \left\langle\nabla_w f(r_s), -2w_s-\frac{c_\kappa}{L}\nabla_s-\frac{c_\kappa}{L}\Delta_s\right\rangle\right)}_{=:\spadesuit} ds\\
&\quad + \mu_k\cdot\underbrace{\left(\left(\frac{8c_\kappa}{L}\gamma_s^T\nabla_w^2 f(r_s)\gamma_s + \frac{2c_\kappa}{L}\bar{\gamma}_s^T\nabla_w^2 f(r_s)\bar{\gamma}_s\right)\right)}_{=:\heartsuit} ds\\
&\quad + \mu_k\cdot\left\langle\nabla_w f(r_s), \left(4\sqrt{\frac{c_\kappa}{L}}\gamma_t\gamma_t^T dB_t + 2\sqrt{\frac{c_\kappa}{L}}\bar{\gamma}_t\bar{\gamma}_t^T dA_t\right)\right\rangle ds, \qquad (42)
\end{aligned}$$

where (i) follows from Itô's Lemma, and (ii) follows from Eqs. (11) - (14), and the definition of ∇_t and Δ_t in Eq. (20).

In the sequel, we upper bound the terms $\spadesuit, \heartsuit, \clubsuit$ separately. Before we proceed, we verify the following inequalities:

$$q'(\|z_s\|_2)\left\langle\frac{z_s}{\|z_s\|_2}, w_s\right\rangle = q'(\|z_s\|_2)\left\langle\frac{z_s}{\|z_s\|_2}, z_s+w_s-z_s\right\rangle \overset{(i)}{\le} q'(\|z_s\|_2)(\|z_s+w_s\|_2-\|z_s\|_2),$$

where (i) is by Cauchy-Schwarz, and:

$$\begin{aligned}
&q'(\|z_s + w_s\|_2) \left\langle \frac{z_s + w_s}{\|z_s + w_s\|_2}, -w_s - \frac{c_\kappa}{L}\nabla_s \right\rangle \\
&\qquad = q'(\|z_s + w_s\|_2) \left\langle \frac{z_s + w_s}{\|z_s + w_s\|_2}, -z_s - w_s + z_s - \frac{c_\kappa}{L}\nabla_s \right\rangle \\
&\qquad \overset{(i)}{\leq} q'(\|z_s + w_s\|_2)\left(-\|z_s + w_s\|_2 + \|z_s\|_2 + \left\langle \frac{z_s + w_s}{\|z_s + w_s\|_2}, -\frac{c_\kappa}{L}\nabla_s \right\rangle \right) \\
&\qquad \overset{(ii)}{\leq} q'(\|z_s + w_s\|_2)(-\|z_s + w_s\|_2 + (1 + c_\kappa)\|z_s\|_2),
\end{aligned}$$

where (i) is again by Cauchy-Schwarz and (ii) is by Cauchy-Schwarz combined with Assumption (A1). Finally:

$$q'(z_s + w_s) \left\langle \frac{z_s + w_s}{\|z_s + w_s\|_2}, -\frac{c_\kappa}{L}\Delta_s \right\rangle \leq q'(\|z_s + w_s\|_2)\frac{c_\kappa}{L}\|\Delta_s\|_2, \tag{43}$$

where the inequality above is by Cauchy-Schwarz along with the fact that $q'(r) \geq 0$ for all r from Lemma 7.

Bounding $\spadesuit$: From Eqs. (42) and (41):

$$\begin{aligned}
\spadesuit &= (1 + 2c_\kappa) f'(r_s) q'(\|z_s\|_2) \left\langle \frac{z_s}{\|z_s\|_2}, w_s \right\rangle \\
&\quad + f'(r_s) q'(\|z_s + w_s\|_2) \left\langle \frac{z_s + w_s}{\|z_s + w_s\|_2}, w_s \right\rangle \\
&\quad + f'(r_s) q'(\|z_s + w_s\|_2) \left\langle \frac{z_s + w_s}{\|z_s + w_s\|_2}, -2w_s - \frac{c_\kappa}{L}\nabla_s - \frac{c_\kappa}{L}\Delta_s \right\rangle \\
&= (1 + 2c_\kappa) f'(r_s) q'(\|z_s\|_2) \left\langle \frac{z_s}{\|z_s\|_2}, w_s \right\rangle \\
&\quad + f'(r_s) q'(\|z_s + w_s\|_2) \left\langle \frac{z_s + w_s}{\|z_s + w_s\|_2}, -w_s - \frac{c_\kappa}{L}\nabla_s - \frac{c_\kappa}{L}\Delta_s \right\rangle =: \spadesuit_1.
\end{aligned} \tag{44}$$

We again highlight the fact that $q'(\|z\|_2)\frac{z}{\|z\|_2}$ is defined for all z, particularly at $\|z\|_2 = 0$, as $q(r) = o(r^2)$ near zero (see Lemma 7).

Substituting the inequality in Eq. (43) into $\spadesuit_1$:

$$\begin{aligned}
\spadesuit_1 &= (1 + 2c_\kappa) f'(r_s) q'(\|z_s\|_2) \left\langle \frac{z_s}{\|z_s\|_2}, w_s \right\rangle \\
&\quad + f'(r_s) q'(\|z_s + w_s\|_2) \left\langle \frac{z_s + w_s}{\|z_s + w_s\|_2}, -w_s - \frac{c_\kappa}{L}\nabla_s - \frac{c_\kappa}{L}\Delta_s \right\rangle \\
&\leq (1 + 2c_\kappa) f'(r_s) q'(\|z_s\|_2)(\|z_s + w_s\|_2 - \|z_s\|_2) \\
&\quad + f'(r_s) q'(\|z_s + w_s\|_2)(-\|z_s + w_s\|_2 + (1 + c_\kappa)\|z_s\|_2) + \frac{c_\kappa}{L}\|\Delta_s\|_2,
\end{aligned}$$

where the inequality uses Cauchy-Schwarz and (F2) of Lemma 31.

Now consider a few cases. We will use the expression for $q'(r)$ from Eq. (7) a number of times:

1. If $\|z_s\|_2 \in [\beta, \infty), \|z_s + w_s\|_2 \in [\beta, \infty)$, then $q'(\|z_s\|_2) = q'(\|z_s + w_s\|_2) = 1$, so that

$$\begin{aligned} \spadesuit_1 &\leq f'(r_s)(\|z_s + w_s\|_2 - \|z_s\|_2 - \|z_s + w_s\|_2 + (1 + c_\kappa)\|z_s\|_2) + \frac{c_\kappa}{L}\|\Delta_s\|_2 \\ &= f'(r_s)(c_\kappa\|z_s\|_2) + \frac{c_\kappa}{L}\|\Delta_s\|_2 \\ &\leq 2c_\kappa f'(r_s)r_s + \beta + \frac{c_\kappa}{L}\|\Delta_s\|_2, \end{aligned}$$

   where we use the definition of r_t defined in Eq. (19) and Lemma 6.1.

2. If $\|z_s\|_2 \in [0, \beta), \|z_s + w_s\|_2 \in [\beta, \infty)$, then $q'(\|z_s\|_2) \in [0, 1]$ and $q'(\|z_s + w_s\|_2) = 1$, so that

$$\begin{aligned} \spadesuit_1 &\overset{(i)}{\leq} f'(r_s)\big((1 + 2c_\kappa)q'(\|z_s\|_2)\|w_s\|_2 - \|z_s + w_s\|_2 + (1 + c_\kappa)\|z_s\|_2\big) + \frac{c_\kappa}{L}\|\Delta_s\|_2 \\ &\overset{(ii)}{\leq} f'(r_s)(2c_\kappa\|w_s\|_2 + 3\|z_s\|_2) + \frac{c_\kappa}{L}\|\Delta_s\|_2 \\ &\overset{(iii)}{\leq} f'(r_s)(2c_\kappa\|w_s\|_2 + 3\beta) + \frac{c_\kappa}{L}\|\Delta_s\|_2 \\ &\overset{(iv)}{\leq} 2c_\kappa f'(r_s)r_s + 5\beta + \frac{c_\kappa}{L}\|\Delta_s\|_2, \end{aligned}$$

   where (i) uses $\|z_s + w_s\|_2 - \|z_s\|_2 \leq \|w_s\|_2$, (ii) uses $\|w_s\|_2 - \|z_s + w_s\|_2 \leq \|z)s\|_2$, (iii) uses our upper bound in $\|z_s\|_2$ and (iv) uses the definition of r_t in Eq. (19) and Lemma 6.1.

3. If $\|z_s\|_2 \in [\beta, \infty), \|z_s + w_s\|_2 \in [0, \beta)$, then $q'(\|z_s\|_2) = 1$ and $q'(\|z_s + w_s\|_2) \in [0, 1]$, so that

$$\begin{aligned} \spadesuit_1 &\overset{(i)}{\leq} f'(r_s)((1 + 2c_\kappa)(\|z_s + w_s\|_2 - \|z_s\|_2) - \|z_s + w_s\|_2 + (1 + c_\kappa)\|z_s\|_2) + \frac{c_\kappa}{L}\|\Delta_s\|_2 \\ &= f'(r_s)(2c_\kappa\|z_s + w_s\|_2 - c_\kappa\|z_s\|_2) + \frac{c_\kappa}{L}\|\Delta_s\|_2 \\ &\overset{(ii)}{\leq} f'(r_s)\Big(3c_\kappa\|z_s + w_s\|_2 - \frac{c_\kappa}{2}r_s + 2\beta\Big) + \frac{c_\kappa}{L}\|\Delta_s\|_2 \\ &\leq f'(r_s)\Big(-\frac{c_\kappa}{2}r_s\Big) + 5\beta + \frac{c_\kappa}{L}\|\Delta_s\|_2, \end{aligned}$$

   where (i) uses our expression for $q'(\cdot)$, and (ii) uses the expression for r_t in Eq. (19), the fact that $c_\kappa \leq 1/1000$ and Lemma 6.1.

4. Finally, if $\|z_s\|_2 \in [0, \beta), \|z_s + w_s\|_2 \in [0, \beta)$, then $q'(\|z_s\|_2) \in [0, 1]$ and $q'(\|z_s + w_s\|_2) \in [0, 1]$, so that

$$\spadesuit_1 \leq f'(r_s)(3\beta) + \frac{c_\kappa}{L}\|\Delta_s\|_2 \leq f'(r_s)\Big(-\frac{c_\kappa}{2}r_s\Big) + 5\beta + \frac{c_\kappa}{L}\|\Delta_s\|_2,$$

   where we again use the expression for r_s in Eq. (19) and Lemma 6.1.

Combining the four cases above we find that,

$$\begin{aligned}\spadesuit \le \spadesuit_1 \le & \mathbb{1}\left\{\|z_s+w_s\|_2 \in [0,\beta)\right\} \cdot \left(f'(r_s)\Big(-\frac{c_\kappa}{2}r_s\Big) + 4\beta + \frac{c_\kappa}{L}\|\Delta_s\|_2\right) \\ & + \mathbb{1}\left\{\|z_s+w_s\|_2 \in [\beta,\infty)\right\} \cdot \left(2c_\kappa f'(r_s)r_s + 5\beta + \frac{c_\kappa}{L}\|\Delta_s\|_2\right), \end{aligned} \tag{45}$$

where we use Lemma 31.(F2), Lemma 6.1 and Eq. (18).

Bounding $\heartsuit$:

$$\begin{aligned}\heartsuit &\overset{(i)}{=} \left(\frac{8c_\kappa}{L}\gamma_s^T\nabla_w^2 f(r_s)\gamma_s + \frac{2c_\kappa}{L}\bar{\gamma}_s^T\nabla_w^2 f(r_s)\bar{\gamma}_s\right) \\ &\overset{(ii)}{=} \frac{8c_\kappa}{L}\cdot\gamma_s^T\left(f'(r_s)\cdot q'(\|z_s+w_s\|_2)\cdot\frac{1}{\|z_s+w_s\|_2}\left(I - \frac{(z_s+w_s)(z_s+w_s)^T}{\|z_s+w_s\|_2^2}\right)\right)\gamma_s \\ &\quad + \frac{8c_\kappa}{L}\cdot\gamma_s^T\left(f'(r_s)q''(\|z_s+w_s\|_2)\frac{(z_s+w_s)(z_s+w_s)^T}{\|z_s+w_s\|_2^2}\right)\gamma_s \\ &\quad + \frac{8c_\kappa}{L}\cdot\gamma_s^T\left(f''(r_s)q'(\|z_s+w_s\|_2)^2\frac{(z_s+w_s)(z_s+w_s)^T}{\|z_s+w_s\|_2^2}\right)\gamma_s \\ &\quad + \frac{2c_\kappa}{L}\cdot\bar{\gamma}_s^T\left(f'(r_s)\cdot q'(\|z_s+w_s\|_2)\cdot\frac{1}{\|z_s+w_s\|_2}\left(I - \frac{(z_s+w_s)(z_s+w_s)^T}{\|z_s+w_s\|_2^2}\right)\right)\bar{\gamma}_s \\ &\quad + \frac{2c_\kappa}{L}\cdot\bar{\gamma}_s^T\left(f'(r_s)q''(\|z_s+w_s\|_2)\frac{(z_s+w_s)(z_s+w_s)^T}{\|z_s+w_s\|_2^2}\right)\bar{\gamma}_s \\ &\quad + \frac{2c_\kappa}{L}\cdot\bar{\gamma}_s^T\left(f''(r_s)q'(\|z_s+w_s\|_2)^2\frac{(z_s+w_s)(z_s+w_s)^T}{\|z_s+w_s\|_2^2}\right)\bar{\gamma}_s \\ &\overset{(iii)}{=} \frac{8c_\kappa}{L}\cdot\left(\left(f''(r_s)q'(\|z_s+w_s\|_2)^2 + f'(r_s)q''(\|z_s+w_s\|_2)\right)\right)\cdot\|\gamma_s\|_2^2 \\ &\quad + \frac{2c_\kappa}{L}\cdot\left(f''(r_s)q'(\|z_s+w_s\|_2)^2 + f'(r_s)q''(\|z_s+w_s\|_2)\right)\cdot\|\bar{\gamma}_s\|_2^2, \end{aligned}$$

where (i) is by Eq. (41), (ii) is by Lemma 41 and (iii) is because $\left\langle \gamma_s, \frac{z_s+w_s}{\|z_s+w_s\|_2}\right\rangle = \|\gamma_s\|_2$ and $\left\langle \bar{\gamma}_s, \frac{z_s+w_s}{\|z_s+w_s\|_2}\right\rangle = \|\bar{\gamma}_s\|_2$ (see Eq. (15)).

From Lemma 7.4, $q''(\|z_s+w_s\|_2) = 0$ for $\|z_s+w_s\|_2 \ge \beta/2$ and from Eq. (15), $\gamma_s = \bar{\gamma}_s = 0$ for $\|z_s+w_s\|_2 \le \beta/2$. Thus the above simplifies to

$$\begin{aligned}\heartsuit &\overset{(i)}{\le} \frac{8c_\kappa}{L}\cdot\left(f''(r_s)q'(\|z_s+w_s\|_2)^2\right)\cdot\|\gamma_s\|_2^2 + \frac{2c_\kappa}{L}\cdot\left(f''(r_s)q'(\|z_s+w_s\|_2)^2\right)\cdot\|\bar{\gamma}_s\|_2^2 \\ &\overset{(ii)}{\le} \frac{8c_\kappa}{L}\cdot\left(f''(r_s)q'(\|z_s+w_s\|_2)^2\right)\cdot\|\gamma_s\|_2^2 \\ &\le \mathbb{1}\left\{\|z_s+w_s\|_2 \ge \beta\right\}\cdot\frac{8c_\kappa}{L}\cdot f''(r_s), \end{aligned} \tag{46}$$

where (i) is by Lemma 31 (F5), which implies that $\frac{2c_\kappa}{L}\cdot\left(f''(r_s)q'(\|z_s+w_s\|_2)^2\right)\cdot\|\bar{\gamma}_s\|_2^2 \le 0$. The inequality in (ii) is because $f''(r) \le 0$ for all r (Lemma 31.(F5)), along with the facts that $\mathbb{1}\left\{\|z_s+w_s\|_2 \geq \beta\right\} \cdot q'(\|z_s+w_s\|_2) = \mathbb{1}\left\{\|z_s+w_s\|_2 \geq \beta\right\}$ (by Lemma 7.3), and $\mathbb{1}\left\{r \geq \beta\right\} q'(r)^2 = \mathbb{1}\left\{r \geq \beta\right\}$ (by Eq. (15)).

Combining our upper bounds on $\spadesuit$ and $\heartsuit$ from Eq. (45) and Eq. (46),

$$
\begin{aligned}
\spadesuit + \heartsuit \leq\ & \mathbb{1}\left\{\|z_s+w_s\|_2 < \beta\right\} \cdot \left(f'(r_s)\left(-\frac{c_\kappa}{2} r_s\right) + 5\beta + \frac{c_\kappa}{L}\|\Delta_s\|_2 \right) \\
& + \mathbb{1}\left\{\|z_s+w_s\|_2 \geq \beta\right\} \cdot \left(2c_\kappa f'(r_s) r_s + 5\beta + \frac{c_\kappa}{L}\|\Delta_s\|_2 \right) \\
& + \mathbb{1}\left\{\|z_s+w_s\|_2 \geq \beta\right\} \cdot \frac{8c_\kappa}{L} \cdot f''(r_s) \\
\stackrel{(i)}{=}\ & \mathbb{1}\left\{\|z_s+w_s\|_2 \geq \beta, r_s \leq \sqrt{12}R\right\} \cdot \left(\frac{8c_\kappa}{L} f''(r_s) + 2c_\kappa f'(r_s) \cdot r_s \right) \\
& + \mathbb{1}\left\{\|z_s+w_s\|_2 \geq \beta, r_s > \sqrt{12}R\right\} \cdot \left(2c_\kappa f'(r_s) r_s\right) \\
& + \mathbb{1}\left\{\|z_s+w_s\|_2 < \beta\right\} \cdot \left(f'(r_s)\left(-\frac{c_\kappa}{2} r_s\right)\right) \\
& + 5\beta + \frac{c_\kappa}{L}\|\Delta_s\|_2 \\
\stackrel{(ii)}{=}\ & \mathbb{1}\left\{\|z_s+w_s\|_2 \geq \beta, r_s \leq \sqrt{12}R\right\} \cdot \left(\frac{8c_\kappa}{L}\left(f''(r_s) + \frac{L}{4} f'(r_s) \cdot r_s \right)\right) \\
& + \mathbb{1}\left\{\|z_s+w_s\|_2 \geq \beta, r_s > \sqrt{12}R\right\} \cdot \left(2c_\kappa f'(r_s) r_s\right) \\
& + \mathbb{1}\left\{\|z_s+w_s\|_2 < \beta\right\} \cdot \left(f'(r_s)\left(-\frac{c_\kappa}{2} r_s\right)\right) \\
& + 5\beta + \frac{c_\kappa}{L}\|\Delta_s\|_2,
\end{aligned}
$$

where (i) and (ii) follow from algebraic manipulations. Continuing forward we find that,

$$
\begin{aligned}
\spadesuit + \heartsuit \stackrel{(i)}{\leq}\ & \mathbb{1}\left\{\|z_s+w_s\|_2 \geq \beta, r_s \leq \sqrt{12}R\right\} \cdot \left(-\frac{8c_\kappa}{L} \cdot \frac{e^{-6LR^2}}{48R^2} f(r_s) \right) \\
& + \mathbb{1}\left\{\|z_s+w_s\|_2 \geq \beta, r_s > \sqrt{12}R\right\} \cdot \left(2c_\kappa f'(r_s) r_s\right) \\
& + \mathbb{1}\left\{\|z_s+w_s\|_2 < \beta\right\} \cdot \left(-\frac{c_\kappa e^{-6LR^2}}{4} f(r_s) \right) \\
& + 5\beta + \frac{c_\kappa}{L}\|\Delta_s\|_2 \\
\stackrel{(ii)}{\leq}\ & \mathbb{1}\left\{\|z_s+w_s\|_2 \geq \beta, r_s \leq \sqrt{12}R\right\} \cdot (-C_m f(r_s)) \\
& + \mathbb{1}\left\{\|z_s+w_s\|_2 < \beta\right\} \cdot (-C_m f(r_s)) \\
& + \mathbb{1}\left\{r_s > \sqrt{12}R\right\} \cdot 2r_s + 5\beta + \frac{c_\kappa}{L}\|\Delta_s\|_2 \\
\stackrel{(iii)}{\leq}\ & -C_m f(r_s) + \mathbb{1}\left\{r_s > \sqrt{12}R\right\} \cdot (C_m f(r_s) + 2r_s) + 5\beta + \frac{c_\kappa}{L}\|\Delta_s\|_2 \\
\stackrel{(iv)}{\leq}\ & -C_m f(r_s) + \mathbb{1}\left\{r_s > \sqrt{12}R\right\} \cdot (4r_s) + 5\beta + \frac{c_\kappa}{L}\|\Delta_s\|_2, \qquad (47)
\end{aligned}
$$

where (i) is by Lemma 31 (F4) combined with the choice of α_f and $\mathcal{R}_f$, third line is by Lemma 31 (F2) and Lemma 31 (F3). (ii) follows immediately from the definition of C_m in (9). (iii) can be verified from algebra, and finally (iv) is from the fact that $C_m \leq 1$ and $f(r) \leq r$ for all r (Lemma 31 (F3)).

Thus, by combining the bounds on ♠ and ♡ in Eqs. (47) back into Eq. (42),

$$\begin{aligned}
d\mu_k f(r_s) \leq &- \mu_k C_m f(r_s) ds \\
&\quad + \mu_k \Big(\mathbb{1}\left\{ r_s > \sqrt{12}R \right\} \cdot 4r_s + 5\beta + \frac{c_\kappa}{L} \|\Delta_s\|_2 \Big) ds \\
&\quad + \mu_k \left\langle \nabla_w f(r_s), 4\sqrt{\frac{c_\kappa}{L}} \left(\gamma_s \gamma_s^T dB_s + \frac{1}{2} \bar{\gamma}_s \bar{\gamma}_s^T dA_s \right) \right\rangle \\
\leq &- \mu_k C_m f(r_s) ds \\
&\quad + \mu_k \Big(\mathbb{1}\left\{ r_s > \sqrt{12}R \right\} \cdot 4r_s ds + 5\beta + c_\kappa \left\| x_s - x_{\lfloor \frac{s}{\delta} \rfloor \delta} \right\|_2 \Big) ds \\
&\quad + \mu_k \left\langle \nabla_w f(r_s), 4\sqrt{\frac{c_\kappa}{L}} \left(\gamma_s \gamma_s^T dB_s + \frac{1}{2} \bar{\gamma}_s \bar{\gamma}_s^T dA_s \right) \right\rangle .
\end{aligned} \tag{48}$$

By taking the time derivative of Eq. (24)-(26), we can verify that for $s \in [k\nu, (k+1)\nu)$,

$$\begin{aligned}
d\mu_k \xi_s &= - \mu_k \cdot C_m \xi_s ds + \mu_k \cdot c_\kappa \left\| x_s - x_{\lfloor \frac{t}{\delta} \rfloor \delta} \right\|_2 ds, \\
d\sigma_s &= - \mu_k C_m \sigma_s ds + \mu_k \cdot \mathbb{1}\left\{ r_s \geq \sqrt{12}R \right\} \cdot 4r_s ds, \\
d\phi_s &= - \mu_k C_m \phi_s ds + \mu_k \cdot \left\langle \nabla_w f(r_s), 4\sqrt{\frac{c_\kappa}{L}} \left(\gamma_s \gamma_s^T dB_s + \frac{1}{2} \bar{\gamma}_s \bar{\gamma}_s^T dA_s \right) \right\rangle .
\end{aligned}$$

By combining with Eq. (48) we get

$$d(\mu_k \cdot (f(r_s) - \xi_s) - \sigma_s - \phi_s) \leq - C_m(\mu_k \cdot (f(r_s) - \xi_s) - (\sigma_s + \phi_s)) + 5\beta ds.$$

An application of Grönwall's Lemma over the interval $s \in [k\nu, (k+1)\nu)$ gives us the claimed result:

$$\begin{aligned}
&\mu_k \cdot \left(f(r_{(k+1)\nu}) - \xi_{(k+1)\nu} \right) - \left(\sigma_{(k+1)\nu} + \phi_{(k+1)\nu} \right) \\
&\qquad \leq e^{-C_m \nu} (\mu_k \cdot (f(r_{k\nu}) - \xi_{k\nu}) - (\sigma_{k\nu} + \phi_{k\nu})) + 5\beta\nu.
\end{aligned}$$

■

D.3 Main results for synchronous coupling

Our main result in this section is Lemma 10, which shows that over a period of T_{sync}, $f(r_s)$ contracts by an amount $\exp(-C_m T_{sync})$ with probability one. Note that this is weaker than showing a contraction rate of $\exp(-C_m t)$ for all t, but is sufficient for our purposes.

Lemma 10 *Assume that $e^{72LR^2} \geq 2$. With probability one, for all k,*

$$\begin{aligned}
&\mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot \left(f(r_{k\nu}) - \xi_{k\nu}\right) \\
&\qquad \leq \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot \exp\left(-C_m T_{sync}\right) \cdot \left(f\left(r_{\tau_{k-1}}\right) - \xi_{\tau_{k-1}}\right) + 5\beta.
\end{aligned}$$

Proof From our definition of c_κ in Eq. (6), r_t in Eq. (19), and from Lemma 6.1, it can be verified that

$$\begin{aligned}
r_{k\nu} \leq& 1.002(\|z_{k\nu}\|_2 + \|z_{k\nu} + w_{k\nu}\|_2) + 2\beta \\
\leq& \sqrt{2.002}\sqrt{\|z_{k\nu}\|_2^2 + \|z_{k\nu} + w_{k\nu}\|_2^2} + 2\beta.
\end{aligned}$$

On the other hand, by $\|\cdot\|_1 \geq \|\cdot\|_2$ and by Lemma 6,

$$r_{\tau_{k-1}} \geq \sqrt{\left\|z_{\tau_{k-1}}\right\|_2^2 + \left\|z_{\tau_{k-1}} + w_{\tau_{k-1}}\right\|_2^2} - 2\beta.$$

Combining the inequality in the display above with the statement of Lemma 13 gives:

$$\begin{aligned}
\mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot r_{k\nu} \leq& \sqrt{\frac{47}{50}} \cdot \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot r_{\tau_{k-1}} \\
&+ \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot c_\kappa \int_{\tau_{k-1}}^{k\nu} e^{-\frac{c_\kappa^2}{3}(k\nu - t)} \left\|x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta}\right\|_2 dt + 5\beta.
\end{aligned}$$

Combining the above with (F2), (F3) and (F6) of Lemma 31, and by using the definition of f in Eq. (21),

$$\begin{aligned}
&\mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot f(r_{k\nu}) \\
&\quad \leq \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot \exp\left(-\frac{1 - \sqrt{47/50}}{4} e^{-6LR^2}\right) f(r_{\tau_{k-1}}) \\
&\qquad + \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot c_\kappa \int_{\tau_{k-1}}^{k\nu} e^{-\frac{c_\kappa^2}{3}(k\nu - t)} \left\|x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta}\right\|_2 dt + 5\beta \\
&\quad \leq \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot \exp\left(-C_m T_{sync}\right) f\left(r_{\tau_{k-1}}\right) \\
&\qquad + \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot c_\kappa \int_{\tau_{k-1}}^{k\nu} e^{-\frac{c_\kappa^2}{3}(k\nu - t)} \left\|x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta}\right\|_2 dt + 5\beta \\
&\quad \overset{(i)}{\leq} \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot \exp\left(-C_m T_{sync}\right) f\left(r_{\tau_{k-1}}\right) \\
&\qquad + \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot c_\kappa \int_{\tau_{k-1}}^{k\nu} e^{-C_m(k\nu - t)} \left\|x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta}\right\|_2 dt + 5\beta, \qquad (49)
\end{aligned}$$

where the first line in (i) follows from the definition of T_{sync} and C_m in Eq. (8) and Eq. (9) along with the fact that $(1 - \sqrt{47/50})/4 \geq 1/200$. The second line in (i) is because $C_m \leq \frac{c_\kappa^2}{3}$ from Eq. (9).

By definition of ξ_t in Eq. (24),

$$\begin{aligned}
&\mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\}\xi_{k\nu} \\
&\quad= \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot \int_0^{k\nu} e^{-C_m(k\nu - t)} c_\kappa \left\| x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta} \right\|_2 dt \\
&\quad= \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot e^{-C_m(k\nu - \tau_{k-1})} \int_0^{\tau_{k-1}} e^{-C_m(\tau_{k-1} - t)} c_\kappa \left\| x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta} \right\|_2 dt \\
&\qquad+ \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot \int_{\tau_{k-1}}^{k\nu} e^{-C_m(k\nu - t)} c_\kappa \left\| x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta} \right\|_2 dt \\
&\quad= \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot \exp\left(-C_m(k\nu - \tau_{k-1})\right) \xi_{\tau_{k-1}} \\
&\qquad+ c_\kappa \int_{\tau_{k-1}}^{k\nu} \exp\left(-C_m(k\nu - t)\right) \left\| x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta} \right\|_2 dt \\
&\quad= \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot \exp\left(-C_m T_{sync}\right) \xi_{\tau_{k-1}} \\
&\qquad+ c_\kappa \int_{\tau_{k-1}}^{k\nu} \exp\left(-C_m(k\nu - t)\right) \left\| x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta} \right\|_2 dt. \qquad (50)
\end{aligned}$$

By subtracting the left and the right hand sides of Eq. (50) and Eq. (49) thus gives us that,

$$\begin{aligned}
&\mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot \left(f(r_{k\nu}) - \xi_{k\nu}\right) \\
&\qquad\le \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot \exp\left(-C_m T_{sync}\right) \cdot \left(f\left(r_{\tau_{k-1}}\right) - \xi_{\tau_{k-1}}\right) + 5\beta.
\end{aligned}$$

■

We now state and prove several auxillary lemmas which are required for the proof of Lemma 10.

Lemma 11 *If* $\|z_s\|_2^2 + \|z_s + w_s\|_2^2 \ge 2.2R^2$, *then*

$$\langle z_s, w_s \rangle + \left\langle z_s + w_s, -w_s - \frac{c_\kappa}{L}\nabla_s \right\rangle \le -\frac{c_\kappa^2}{3}\left(\|z_s\|_2^2 + \|z_s + w_s\|_2^2\right).$$

Proof We begin by expanding the differentials $d\|z_s\|_2^2 + d\|z_s + w_s\|_2^2$:

$$\begin{aligned}
d\|z_s\|_2^2 + d\|z_s + w_s\|_2^2 &= 2\left\langle z_s, w_s \right\rangle + 2\left\langle z_s + w_s, -w_s - \frac{c_\kappa}{L}\nabla_s \right\rangle \\
&= -2\|w_s\|_2^2 - 2\left\langle z_s, \frac{c_\kappa}{L}\nabla_s \right\rangle - 2\left\langle w_s, \frac{c_\kappa}{L}\nabla_s \right\rangle \\
&= -2\|w_s\|_2^2 - 2\left\langle z_s, \frac{c_\kappa}{L}\nabla_s \right\rangle + \|w_s\|_2^2 + \frac{c_\kappa^2}{L^2}\|\nabla_t\|_2^2 - \|w_t + \frac{c_\kappa}{L}\nabla_t\|_2^2 \\
&\le -\|w_s\|_2^2 - 2\left\langle z_s, \frac{c_\kappa}{L}\nabla_s \right\rangle + \frac{c_\kappa^2}{L^2}\|\nabla_s\|_2^2 \\
&\le -\|w_s\|_2^2 - 2\left\langle z_s, \frac{c_\kappa}{L}\nabla_s \right\rangle + c_\kappa^2\|z_s\|_2^2 =: \spadesuit. \qquad (51)
\end{aligned}$$

Now consider two cases.

Case 1: ($\|z_s\|_2 \le R$) By Young's inequality,

$$\|z_s + w_s\|_2^2 \le 11\|w_s\|_2^2 + 1.1\|z_s\|_2^2.$$

Furthermore, by our assumption that $\|z_s\|_2^2 + \|z_s + w_s\|_2^2 \ge 2.2R^2$,

$$\begin{aligned} 11\|w_s\|_2^2 &\ge \|z_s + w_s\|_2^2 - 1.1\|z_s\|_2^2 \\ &= \|z_s\|_2^2 + \|z_s + w_s\|_2^2 - 1.1\|z_s\|_2^2 - \|z_s\|_2^2 \\ &\ge 2.2R^2 - 2.1R^2 \\ &\ge 0.1R^2 \\ &\ge 0.1\|z_s\|_2^2, \\ \implies \|z_s\|_2^2 &\le \frac{1000}{9}\|w_s\|_2^2. \end{aligned} \tag{52}$$

With this implication ♠ can now be upper bounded by

$$\begin{aligned} \spadesuit &= -\|w_s\|_2^2 - 2\left\langle z_s, \frac{c_\kappa}{L}\nabla_s\right\rangle + c_\kappa^2\|z_s\|_2^2 \\ &\overset{(i)}{\le} -\|w_s\|_2^2 + 2c_\kappa\|z_s\|_2^2 + c_\kappa^2\|z_s\|_2^2 \\ &\overset{(ii)}{\le} -\|w_s\|_2^2 + 3c_\kappa\|z_s\|_2^2 \\ &\overset{(iii)}{\le} -\frac{2}{3}\|w_s\|_2^2 \\ &\overset{(iv)}{\le} -\frac{c_\kappa^2}{3}\Big(\|z_s\|_2^2 + \|z_s + w_s\|_2^2\Big), \end{aligned}$$

where (i) is by Assumption (A1) and Cauchy-Schwarz, and (ii) is because $c_\kappa := \frac{1}{1000\kappa} \le \frac{1}{1000}$. The inequality (iii) is by the implication in Eq. (52), which gives $3c_\kappa\|z_s\|_2^2 \le \frac{1000c_\kappa}{3}\|w_s\|_2^2 \le \frac{1}{3}\|w_s\|_2^2$. Finally, (iv) can be verified as follows:

$$\begin{aligned} \|z_s\|_2^2 + \|z_s + w_s\|_2^2 &\overset{(i)}{\le} 3\|z_s\|_2^2 + 2\|w_s\|_2^2 \\ &\overset{(ii)}{\le} \frac{1000}{3}\|w_t\|_2^2 + 2\|w_t\|_2^2 \\ &\le \frac{1006}{3}\|w_t\|_2^2. \\ \implies \qquad \Big(\|z_s\|_2^2 + \|z_s + w_s\|_2^2\Big) &\le \frac{1006}{3}\|w_s\|_2^2 \\ &\le \frac{1}{2c_\kappa}\|w_s\|_2^2. \\ \implies \qquad \frac{2}{3}\|w_s\|_2^2 &\ge \frac{4c_\kappa}{3}\Big(\|z_s\|_2^2 + \|z_s + w_s\|_2^2\Big) \\ &\overset{(iii)}{\ge} \frac{c_\kappa^2}{3}\Big(\|z_s\|_2^2 + \|z_s + w_s\|_2^2\Big), \end{aligned}$$

where (i) is by Young's inequality, (ii) is by Eq. (52), and (iii) is by $c_\kappa \le \frac{1}{1000}$.

Case 2: ($\|z_s\|_2 \geq R$) We have,

$$
\begin{aligned}
\spadesuit =& -\|w_s\|_2^2 - 2\left\langle z_s, \frac{c_\kappa}{L}\nabla_s \right\rangle + c_\kappa^2 \|z_s\|_2^2 \\
\overset{(i)}{\leq}& -\|w_s\|_2^2 - 2c_\kappa^2\|z_s\|_2^2 + c_\kappa^2\|z_s\|_2^2 \\
\leq& -\|w_s\|_2^2 - c_\kappa^2\|z_s\|_2^2 \\
\leq& -c_\kappa^2\Big(\|w_s\|_2^2 + \|z_s\|_2^2\Big) \\
\overset{(ii)}{\leq}& -\frac{c_\kappa^2}{3}\Big(\|z_s\|_2^2 + \|z_s + w_s\|_2^2\Big),
\end{aligned}
$$

where (i) is by Assumption (A3) and (ii) is because

$$
\begin{aligned}
\|z_s\|_2^2 + \|z_s + w_s\|_2^2 \leq& 3\|z_s\|_2^2 + 2\|w_s\|_2^2 \\
\leq& 3\Big(\|z_s\|_2^2 + \|w_s\|_2^2\Big).
\end{aligned}
$$

Hence, we have proved the result under both cases. ∎

Lemma 12 *With probability one,*

$$
\begin{aligned}
&(1-\mu_k)\cdot\left(\sqrt{\left\|z_{(k+1)\nu}\right\|_2^2 + \left\|z_{(k+1)\nu} + w_{(k+1)\nu}\right\|_2^2} - \sqrt{2.2}R\right)_+ \\
&\quad \leq (1-\mu_k)\cdot e^{-\frac{c_\kappa^2\nu}{3}}\left(\sqrt{\|z_{k\nu}\|_2^2 + \|z_{k\nu} + w_{k\nu}\|_2^2} - \sqrt{2.2}R\right)_+ \\
&\qquad + (1-\mu_k)\cdot c_\kappa \int_{k\nu}^{(k+1)\nu} e^{-\frac{c_\kappa^2}{3}((k+1)\nu - t)}\left\|x_t - x_{\lfloor \frac{t}{\delta}\rfloor\delta}\right\|_2 dt.
\end{aligned}
$$

Proof When $\mu_k = 1$, the inequality holds trivially $(0 = 0)$, so for the rest of this proof, we consider the case $\mu_k = 0$. To simplify notation, we leave out the multiplier $(1-\mu_k)$ in all subsequent expressions.

We can verify from Eqs. (11)-(14) and Eq. (18) that when $\mu_k = 0$, for any $s \in [k\nu, (k+1)\nu)$,

$$
\begin{aligned}
dz_s =& w_s ds \\
d(z_s + w_s) =& \Big(-w_s + \frac{c_\kappa}{L}(\nabla_s + \Delta_s)\Big)ds.
\end{aligned}
$$

Thus, for any $s \in [k\nu, (k+1)\nu)$,

$$
\begin{aligned}
&d\left(\left(\sqrt{\|z_s\|_2^2 + \|z_s + w_s\|_2^2} - \sqrt{2.2}R\right)_+^2\right) \\
&\quad \overset{(i)}{=} \frac{\left(\sqrt{\|z_s\|_2^2 + \|z_s + w_s\|_2^2} - \sqrt{2.2}R\right)_+}{\sqrt{\|z_s\|_2^2 + \|z_s + w_s\|_2^2}} \left\langle \begin{bmatrix} z_s \\ z_s + w_s \end{bmatrix}, \begin{bmatrix} w_s \\ -w_s - \frac{c_\kappa}{L}(\nabla_s + \Delta_s) \end{bmatrix} \right\rangle ds \\
&\quad \overset{(ii)}{\leq} -\frac{c_\kappa^2}{3} \frac{\left(\sqrt{\|z_s\|_2^2 + \|z_s + w_s\|_2^2} - \sqrt{2.2}R\right)_+}{\sqrt{\|z_s\|_2^2 + \|z_s + w_s\|_2^2}} \cdot \left(\|z_s\|_2^2 + \|z_s + w_s\|_2^2\right) ds \\
&\qquad + \frac{\left(\sqrt{\|z_s\|_2^2 + \|z_s + w_s\|_2^2} - \sqrt{2.2}R\right)_+}{\sqrt{\|z_s\|_2^2 + \|z_s + w_s\|_2^2}} \cdot (\|z_s + w_s\|_2 \|\Delta_s\|_2) ds \\
&\quad \leq -\frac{c_\kappa^2}{3} \left(\sqrt{\|z_s\|_2^2 + \|z_s + w_s\|_2^2} - \sqrt{2.2}R\right)_+ \cdot \sqrt{\|z_s\|_2^2 + \|z_s + w_s\|_2^2} ds \\
&\qquad + \frac{c_\kappa}{L} \left(\sqrt{\|z_s\|_2^2 + \|z_s + w_s\|_2^2} - \sqrt{2.2}R\right)_+ \cdot \|\Delta_s\|_2 ds \\
&\quad \leq -\frac{c_\kappa^2}{3} \cdot \left(\sqrt{\|z_s\|_2^2 + \|z_s + w_s\|_2^2} - \sqrt{2.2}R\right)_+^2 ds \\
&\qquad + c_\kappa \left(\sqrt{\|z_s\|_2^2 + \|z_s + w_s\|_2^2} - \sqrt{2.2}R\right)_+ \cdot \left\| x_s - x_{\lfloor \frac{s}{\delta} \rfloor \delta} \right\|_2 ds,
\end{aligned}
$$

where (i) is by the expression for dz_s and dw_s established above, and (ii) is by Lemma 11 and Cauchy-Schwarz, the last two inequalities follow by algebraic manipulations.

Dividing throughout by $\left(\sqrt{\|z_s\|_2^2 + \|z_s + w_s\|_2^2} - \sqrt{2.2}R\right)_+$ gives us that

$$
\begin{aligned}
&d\left(\sqrt{\|z_s\|_2^2 + \|z_s + w_s\|_2^2} - \sqrt{2.2}R\right)_+ \\
&\qquad \leq \left(-\frac{c_\kappa^2}{3}\left(\sqrt{\|z_s\|_2^2 + \|z_s + w_s\|_2^2} - \sqrt{2.2}R\right)_+ + c_\kappa \left\| x_s - x_{\lfloor \frac{s}{\delta} \rfloor \delta} \right\|_2\right) dt.
\end{aligned}
$$

We can verify that the inequality implies that

$$
\begin{aligned}
&d\left(\left(\sqrt{\|z_s\|_2^2 + \|z_s + w_s\|_2^2} - \sqrt{2.2}R\right)_+ - c_\kappa \int_{k\nu}^{s} e^{-\frac{c_\kappa^2}{3}(s-t)} \left\| x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta} \right\|_2 dt\right) \\
&\quad \leq -\frac{c_\kappa^2}{3}\left(\left(\sqrt{\|z_s\|_2^2 + \|z_s + w_s\|_2^2} - \sqrt{2.2}R\right)_+ - c_\kappa \int_{k\nu}^{s} e^{-\frac{c_\kappa^2}{3}(s-t)} \left\| x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta} \right\|_2 dt\right) dt.
\end{aligned}
$$

Thus by Grönwall's Lemma,

$$\left(\sqrt{\|z_{(k+1)\nu}\|_2^2 + \|z_{(k+1)\nu} + w_{(k+1)\nu}\|_2^2} - \sqrt{2.2}R\right)_+ - c_\kappa \int_{k\nu}^{(k+1)\nu} e^{-\frac{c_\kappa^2}{3}((k+1)\nu - t)} \left\|x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta}\right\|_2 dt$$
$$\leq e^{-\frac{c_\kappa^2}{3}((k+1)\nu - k\nu)} \left(\left(\sqrt{\|z_{k\nu}\|_2^2 + \|z_{k\nu} + w_{k\nu}\|_2^2} - \sqrt{2.2}R\right)_+ - c_\kappa \int_{k\nu}^{k\nu} e^{-\frac{c_\kappa^2}{3}(k\nu - t)} \left\|x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta}\right\|_2 dt\right)$$
$$= e^{-\frac{c_\kappa^2 \nu}{3}} \left(\sqrt{\|z_{k\nu}\|_2^2 + \|z_{k\nu} + w_{k\nu}\|_2^2} - \sqrt{2.2}R\right)_+ .$$

This proves the statement of the Lemma. ■

Lemma 13 *Assume that $e^{72LR^2} \geq 2$. With probability one, for all positive integers k,*

$$\begin{aligned}
&\mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot \sqrt{\|z_{k\nu}\|^2 + \|z_{k\nu} + w_{k\nu}\|^2} \\
&\qquad \leq \sqrt{\frac{23}{50}} \cdot \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot \sqrt{\left\|z_{\tau_{k-1}}\right\|^2 + \left\|z_{\tau_{k-1}} + w_{\tau_{k-1}}\right\|^2} \\
&\qquad\quad + \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot c_\kappa \int_{\tau_{k-1}}^{k\nu} e^{-\frac{c_\kappa^2}{3}(k\nu - t)} \left\|x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta}\right\|_2 dt + 3\beta.
\end{aligned}$$

Proof By our choice ν we know that T_{sync}/ν is an integer, thus we have,

$$k\nu = \tau_{k-1} + T_{sync} \Rightarrow (k-1)\nu < \tau_{k-1} + T_{sync}.$$

Thus,

$$\begin{aligned}
\mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} &\overset{(i)}{=} \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot \mathbb{1}\left\{(k-1)\nu < \tau_{k-1} + T_{sync}\right\} \\
&\overset{(ii)}{=} \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot (1 - \mu_{k-1}) \\
&\overset{(iii)}{=} \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot \prod_{i \in S_{k-1}} (1 - \mu_i), \qquad (53)
\end{aligned}$$

where $S_{k-1} := \left\{\frac{\tau_{k-1}}{\nu}, \frac{\tau_{k-1}}{\nu} + 1, ..., k-1\right\}$ (as defined in Lemma 14). Above, (i) is because $k\nu = \tau_{k-1} + T_{sync} \Rightarrow (k-1)\nu < \tau_{k-1} + T_{sync}$, (ii) is because $(k-1)\nu < \tau_{k-1} + T_{sync} \Rightarrow \mu_{k-1} = 0$ (see Eq. (18)) and (iii) is by Part 2 of Lemma 14.

We can now recursively apply Lemma 12 as follows: (to simplify notation, let $\alpha := \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\}$):

$$\begin{aligned}
&\alpha \cdot \prod_{i \in S_{k-1}} (1-\mu_i) \cdot \left(\sqrt{\|z_{k\nu}\|_2^2 + \|z_{k\nu} + w_{k\nu}\|_2^2} - \sqrt{2.2}R\right)_+ \\
&\quad \le \alpha \cdot \prod_{i \in S_{k-1}} (1-\mu_i) \cdot e^{-\frac{c_\kappa^2}{3}\nu} \left(\sqrt{\left\|z_{(k-1)\nu}\right\|_2^2 + \left\|z_{(k-1)\nu} + w_{(k-1)\nu}\right\|_2^2} - \sqrt{2.2}R\right)_+ \\
&\qquad + \alpha \cdot \prod_{i \in S_{k-1}} (1-\mu_i) \cdot c_\kappa \int_{(k-1)\nu}^{k\nu} e^{-\frac{c_\kappa^2}{3}(k\nu - t)} \left\|x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta}\right\|_2 dt \\
&\quad \le \alpha \cdot \prod_{i \in S_{k-1}} (1-\mu_i) \cdot e^{-\frac{c_\kappa^2}{3}T_{sync}} \left(\sqrt{\left\|z_{\tau_{k-1}}\right\|_2^2 + \left\|z_{\tau_{k-1}} + w_{\tau_{k-1}}\right\|_2^2} - \sqrt{2.2}R\right)_+ \\
&\qquad + \alpha \cdot \prod_{i \in S_{k-1}} (1-\mu_i) \cdot c_\kappa \int_{\tau_{k-1}}^{k\nu} e^{-\frac{c_\kappa^2}{3}(k\nu - t)} \left\|x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta}\right\|_2 dt,
\end{aligned} \tag{54}$$

where the last inequality uses the fact that $\nu \cdot (k - \tau_{k-1}) = T_{sync}$ in the definition of α.

Thus, we have,

$$\begin{aligned}
&\mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot \left(\sqrt{\|z_{k\nu}\|_2^2 + \|z_{k\nu} + w_{k\nu}\|_2^2} - \sqrt{2.2}R\right)_+ \\
&\quad \overset{(i)}{=} \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot \prod_{i \in S_{k-1}} (1-\mu_i) \cdot \left(\sqrt{\|z_{k\nu}\|_2^2 + \|z_{k\nu} + w_{k\nu}\|_2^2} - \sqrt{2.2}R\right)_+ \\
&\quad \overset{(ii)}{\le} \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot \prod_{i \in S_{k-1}} (1-\mu_i) \cdot \left(\sqrt{\|z_{k\nu}\|_2^2 + \|z_{k\nu} + w_{k\nu}\|_2^2} - \sqrt{2.2}R\right)_+ \\
&\qquad + \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot \prod_{i \in S_{k-1}} (1-\mu_i) \cdot c_\kappa \int_{\tau_{k-1}}^{k\nu} e^{-\frac{c_\kappa^2}{3}(k\nu - t)} \left\|x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta}\right\|_2 dt \\
&\quad \overset{(iii)}{=} \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot e^{-\frac{c_\kappa^2}{3}T_{sync}} \left(\sqrt{\left\|z_{\tau_{k-1}}\right\|_2^2 + \left\|z_{\tau_{k-1}} + w_{\tau_{k-1}}\right\|_2^2} - \sqrt{2.2}R\right)_+ \\
&\qquad + \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot c_\kappa \int_{\tau_{k-1}}^{k\nu} e^{-\frac{c_\kappa^2}{3}(k\nu - t)} \left\|x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta}\right\|_2 dt \\
&\quad \overset{(iv)}{\le} \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot \frac{1}{100} \left(\sqrt{\left\|z_{\tau_{k-1}}\right\|_2^2 + \left\|z_{\tau_{k-1}} + w_{\tau_{k-1}}\right\|_2^2} - \sqrt{2.2}R\right)_+ \\
&\qquad + \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot c_\kappa \int_{\tau_{k-1}}^{k\nu} e^{-\frac{c_\kappa^2}{3}(k\nu - t)} \left\|x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta}\right\|_2 dt,
\end{aligned} \tag{55}$$

where (i) is by Eq. (53), (ii) is by Eq. (54), (iii) is by Eq. (53) again, and (iv) is by the definition $T_{sync} = \frac{3}{c_\kappa^2}\log(100)$.

Let $j := \tau_{k-1}/\nu$. Then by the first part of Lemma 14, we know that $\tau_j = \tau_{k-1} = j\nu$. From the update rule for τ_k, Eq. (17), this must imply that

$$\sqrt{\left\|z_{\tau_{k-1}}\right\|_2^2 + \left\|z_{\tau_{k-1}} + w_{\tau_{k-1}}\right\|_2^2} = \sqrt{\|z_{j\nu}\|_2^2 + \|z_{j\nu} + w_{j\nu}\|_2^2} \geq \sqrt{5}R. \tag{56}$$

Thus finally,

$$\begin{aligned}
&\mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot \left(\sqrt{\|z_{k\nu}\|_2^2 + \|z_{k\nu} + w_{k\nu}\|_2^2}\right) \\
&\overset{(i)}{\leq} \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot \left(\sqrt{\|z_{k\nu}\|_2^2 + \|z_{k\nu} + w_{k\nu}\|_2^2} - \sqrt{2.2}R\right)_+ \\
&\quad + \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot \sqrt{2.2}R \\
&\overset{(ii)}{\leq} \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \frac{1}{100}\left(\sqrt{\left\|z_{\tau_{k-1}}\right\|_2^2 + \left\|z_{\tau_{k-1}} + w_{\tau_{k-1}}\right\|_2^2} - \sqrt{2.2}R\right)_+ \\
&\quad + \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot \sqrt{2.2}R \\
&\quad + \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot c_\kappa \int_{\tau_{k-1}}^{k\nu} e^{-\frac{c_\kappa^2}{3}(k\nu - t)} \left\|x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta}\right\|_2 dt \\
&\overset{(iii)}{\leq} \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \frac{1}{100}\left(\sqrt{\left\|z_{\tau_{k-1}}\right\|_2^2 + \left\|z_{\tau_{k-1}} + w_{\tau_{k-1}}\right\|_2^2} - \sqrt{2.2}R\right) \\
&\quad + \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot \sqrt{\frac{22}{50}} \sqrt{\left\|z_{\tau_{k-1}}\right\|_2^2 + \left\|z_{\tau_{k-1}} + w_{\tau_{k-1}}\right\|_2^2} \\
&\quad + \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot c_\kappa \int_{\tau_{k-1}}^{k\nu} e^{-\frac{c_\kappa^2}{3}(k\nu - t)} \left\|x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta}\right\|_2 dt \\
&\overset{(iv)}{\leq} \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \sqrt{\frac{23}{50}} \sqrt{\left\|z_{\tau_{k-1}}\right\|_2^2 + \left\|z_{\tau_{k-1}} + w_{\tau_{k-1}}\right\|_2^2} \\
&\quad + \mathbb{1}\left\{k\nu = \tau_{k-1} + T_{sync}\right\} \cdot c_\kappa \int_{\tau_{k-1}}^{k\nu} e^{-\frac{c_\kappa^2}{3}(k\nu - t)} \left\|x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta}\right\|_2 dt,
\end{aligned}$$

where (i) is by an algebraic manipulation, (ii) is by Eq. (55), (iii) is by Eq. (56) and (iv) is because $1/100 + \sqrt{22/50} \leq \sqrt{23/50}$. ■

Lemma 14 *Let k be a positive integer, then:*

1. *Let $j = \tau_k/\nu$. Then for all $i \in \{j, j+1, ..., k\}$, $\tau_i = \tau_k = j\nu$.*
2. *If $\mu_k = 0$, then $\mu_i = 0$ for all $i \in \{\tau_k/\nu ... k\}$, $\mu_i = 0$. Equivalently,*

$$\mathbb{1}\left\{\mu_k = 0\right\} = \prod_{i \in S_k} \mathbb{1}\left\{\mu_i = 0\right\},$$

where $S_k := \left\{\frac{\tau_k}{\nu}, ..., k\right\}$.

Proof For the first claim: By definition of the update for τ_k, if $j = \tau_k/\nu$ for any k, then $j\nu = \tau_j = \tau_k$. Note that τ_i is nondecreasing with i, so that $j = \tau_k \leq k$, which implies that $\tau_j \leq \tau_{j+1} \leq ... \leq \tau_j$. Since $\tau_j = \tau_k$, the inequalities must hold with equality.

For the second claim: By the definition of μ_k; $\mu_k = 0$ implies that $k\nu < \tau_k + T_{sync}$. From the first claim, we know that for all $i \in \{\tau_k/\nu ... k\}$, $\tau_i = \tau_k$. Thus
$i\nu \leq k\nu < \tau_k + T_{sync} = \tau_i + T_{sync}$. ∎

D.4 Discretization Error Bound

In this section, we bound the various *discretization errors*. First, in Section D.4.1, we establish a bound on $\mathbb{E}\left[\xi_t\right]$. Then in Lemma 18, we bound $\mathbb{E}\left[\sigma_t\right]$. Finally, in Lemma 23, we show that $\mathbb{E}\left[\phi_t\right] = 0$ as it is a martingale.

D.4.1 Bound on $\mathbb{E}\left[\xi_t\right]$

In this subsection, we establish a bound on $\mathbb{E}\left[\xi_t\right]$. This term represents the discretization error that arises because in the SDE in Eq. (12), the update to u_t uses the gradient $\nabla U\left(x_{\lfloor \frac{t}{\delta} \rfloor \delta}\right)$ instead of $\nabla U(x_t)$. Our main result is Lemma 15, which in turn relies on the uniform bound for all $t \geq 0$ on $\mathbb{E}\left[\left\|x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta}\right\|_2^8\right]$ established in Corollary 16 (based on the moment bounds established in Appendix F).

Lemma 15 *For all $t \geq 0$,*

$$\mathbb{E}\left[\xi_t\right] \leq \delta \cdot \frac{2^9 c_\kappa \left(R + \sqrt{d/m}\right)}{C_m}.$$

Proof By the bound in Corollary 16,

$$\mathbb{E}\left[\left\|x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta}\right\|_2^8\right] \leq \delta^8 2^{72} \left(R^2 + \frac{d}{m}\right)^4.$$

Further, by Jensen's inequality,

$$\mathbb{E}\left[\left\|x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta}\right\|_2\right] \leq \delta \cdot 2^9 \left(R + \sqrt{\frac{d}{m}}\right).$$

By integrating from up to time t,

$$\begin{aligned}
\mathbb{E}\left[\xi_t\right] &= \int_0^t e^{-C_m(t-s)} \mathbb{E}\left[c_\kappa \left\|x_t - x_{\lfloor \frac{s}{\delta} \rfloor \delta}\right\|_2\right] ds \\
&\leq \int_0^t e^{-C_m(t-s)} c_\kappa \delta 2^9 \left(R + \sqrt{\frac{d}{m}}\right) ds \\
&\leq \delta \cdot \frac{2^9 c_\kappa \left(R + \sqrt{d/m}\right)}{C_m}.
\end{aligned}$$

∎

Corollary 16 *For all $t \geq 0$,*

$$\mathbb{E}\left[\left\|x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta}\right\|_2^8\right] \leq \delta^8 2^{72}\left(R^2 + \frac{d}{m}\right)^4.$$

Proof This follows directly by combining the results of Lemma 32 and Lemma 17. ■

Lemma 17 *Suppose that the step size $\delta \leq \frac{1}{1000}$. Then for all $t \in [\lfloor \frac{t}{\delta} \rfloor \delta, (\lfloor \frac{t}{\delta} \rfloor + 1)\delta]$,*

$$\mathbb{E}\left[\left\|x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta}\right\|_2^8\right] \leq \delta^8\left(1.1\mathbb{E}\left[\left(\left\|x_{\lfloor \frac{t}{\delta} \rfloor \delta}\right\|_2^8 + \left\|u_{\lfloor \frac{t}{\delta} \rfloor \delta}\right\|_2^8\right)\right] + 2^{12}\left(R^2 + \frac{d}{m}\right)^4\right).$$

Proof

$$\begin{aligned}
\mathbb{E}\left[\left\|x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta}\right\|_2^8\right] &= \mathbb{E}\left[\left\|\int_{\lfloor \frac{t}{\delta} \rfloor \delta}^t w_s ds\right\|_2^8\right] \\
&\leq \delta^7 \int_{\lfloor \frac{t}{\delta} \rfloor \delta}^t \mathbb{E}\left[\|w_s\|_2^8\right] ds \\
&= \delta^8\left(1.1\mathbb{E}\left[\left(\left\|x_{\lfloor \frac{t}{\delta} \rfloor \delta}\right\|_2^8 + \left\|u_{\lfloor \frac{t}{\delta} \rfloor \delta}\right\|_2^8\right)\right] + 2^{12}\left(R^2 + \frac{d}{m}\right)^4\right),
\end{aligned}$$

where for the last inequality, we use Lemma 34.

■

D.4.2 Bounds on $\mathbb{E}[\sigma_t]$ and $\mathbb{E}[\phi_t]$

In this subsection, we bound $\mathbb{E}[\sigma_t]$ (Lemma 18). This term represents the discretization error that arises because τ_k (and hence μ_k) is updated at discrete time intervals of ν. We highlight the fact that $\mathbb{E}[\sigma_t]$ is bounded by a term that depends on ν, which can be made arbitrarily small. The main ingredient of this proof is a bound on $\mathbb{E}\left[\mu_k \cdot \mathbb{1}\left\{r_s \geq \sqrt{12}R\right\}\right]$ in Lemma 20.

Lemma 18 *For $\beta \leq 0.0001R$. There exists a $C_5 = poly(L, 1/m, d, R, \frac{1}{C_m})$ and $C_3 = 1/poly(L, 1/m, d, R)$, such that for all $\nu \leq C_3$, for all positive integers k, and for all $t \geq 0$,*

$$\mathbb{E}[\sigma_t] \leq C_5\nu^2.$$

Proof By the definition of σ_t in Eq. (25),

$$\begin{aligned}
\mathbb{E}\left[\sigma_t\right] &= \mathbb{E}\left[\int_0^t \mu_{\lfloor \frac{s}{\nu} \rfloor} \cdot e^{-C_m(t-s)} \mathbb{1}\left\{r_s \geq \sqrt{12}R\right\} \cdot 4r_s ds\right]\\
&= 4\int_0^t e^{-C_m(t-s)} \mathbb{E}\left[\mu_{\lfloor \frac{s}{\nu} \rfloor} \mathbb{1}\left\{r_s \geq \sqrt{12}R\right\} r_s\right] ds\\
&\overset{(i)}{\leq} 4\int_0^t e^{-C_m(t-s)} \nu^2 \cdot C_4\\
&\leq \frac{4\nu^2 C_4}{C_m}\\
&= \nu^2 \cdot C_5,
\end{aligned}$$

where (i) is by Corollary 21. ■

Lemma 19 *For all $s \geq 0$,*

$$\mathbb{E}\left[r_s^2\right] \leq 2^{32}\left(R^2 + \frac{d}{m}\right).$$

Proof Recall that,

$$\begin{aligned}
r_s^2 &= \left((1+2c_\kappa)\|z_s\|_2 + \|z_s + w_s\|_2\right)^2\\
&\leq \left((2+2c_\kappa)\|z_s\|_2 + \|w_s\|_2\right)^2\\
&\leq \left(2.1\|x_s\|_2 + 2.1\|y_s\|_2 + \|u_s\|_2 + \|v_s\|_2\right)^2\\
&\overset{(i)}{\leq} 16\left(\|x_s\|_2^2 + \|u_s\|_2^2 + \|y_s\|_2^2 + \|v_s\|_2^2\right)\\
&\leq 2^{16}\left(2^{72}\left(R^2 + \frac{d}{m}\right)^4\right)^{1/4}\\
&= 2^{32}\left(R^2 + \frac{d}{m}\right),
\end{aligned}$$

where (i) is by Lemma 32 and Lemma 33. ■

Lemma 20 *For every $\beta \leq 0.0001R$, there exists a $C_2 = poly(L, 1/m, d, R)$, $C_3 = 1/poly(L, 1/m, d, R)$, such that for all $\nu \leq C_3$, for all positive integers k, and for all $s \in [k\nu, (k+1)\nu]$,*

$$\mathbb{E}\left[\mu_k \cdot \mathbb{1}\left\{r_s \geq \sqrt{12}R\right\}\right] \leq C_2\nu^4.$$

Proof By definition of μ_k in Eq. (18), we know that $\mu_k = 1$ implies that $k\nu - \tau_k \geq T_{sync}$ which further implies that $\tau_k = \tau_{k-1}$ (otherwise τ_k must equal $k\nu$ by the definition of τ_t, in which case $k\nu - \tau_k = 0 < T_{sync}$). This then implies that $k\nu - \tau_{k-1} \geq T_{sync}$. It must thus be the case that $\sqrt{\|z_{k\nu}\|_2^2 + \|z_{k\nu} + w_{k\nu}\|_2^2} < \sqrt{5}R$, because otherwise $\tau_k = k\nu$, which contradicts $\mu_k = 1$. Thus,

$$\mu_k \leq \mathbb{1}\left\{\sqrt{\|z_{k\nu}\|_2^2 + \|z_{k\nu} + w_{k\nu}\|_2^2} < \sqrt{5}R\right\}. \tag{57}$$

By a standard inequality between $\|\cdot\|_1$ and $\|\cdot\|_2$,

$$\begin{aligned}\sqrt{\|z_{k\nu}\|_2^2 + \|z_{k\nu} + w_{k\nu}\|_2^2} &\geq \frac{1}{\sqrt{2}}(\|z_{k\nu}\|_2 + \|z_{k\nu} + w_{k\nu}\|_2) \\ &\overset{(i)}{\geq} \frac{1}{\sqrt{2}}(\ell(z_{k\nu}) + \ell(z_{k\nu} + w_{k\nu})) - \beta \\ &\overset{(ii)}{\geq} \frac{1}{1.002\sqrt{2}} r_{k\nu} - \beta,\end{aligned}$$

where (i) is by Lemma 6.1, and (ii) is by definition of r_t in Eq. (18) and by definition of c_κ. Combining with the inequality (57),

$$\begin{aligned}\mu_k &\leq \mathbb{1}\left\{\frac{1}{1.002\sqrt{2}} r_{k\nu} - \beta < \sqrt{5}R\right\} \\ &= \mathbb{1}\left\{r_{k\nu} < 1.002\sqrt{10}R + \beta\right\} \\ &\leq \mathbb{1}\left\{r_{k\nu} < \sqrt{11}R\right\},\end{aligned} \tag{58}$$

where the final inequality uses our assumption that $\beta \leq 0.0001R$. Thus,

$$\begin{aligned}\mu_k \cdot \mathbb{1}\left\{r_s \geq \sqrt{12}R\right\} \leq& \mathbb{1}\left\{r_{k\nu} < \sqrt{11}R\right\} \cdot \mathbb{1}\left\{r_s \geq \sqrt{12}R\right\} \\ \leq& \mathbb{1}\left\{|r_s - r_{k\nu}| \geq 0.14R\right\}.\end{aligned}$$

Taking expectations,

$$\begin{aligned}\mathbb{E}\left[\mu_k \cdot \mathbb{1}\left\{r_s \geq \sqrt{12}R\right\}\right] &\leq \mathbb{E}\left[\mathbb{1}\left\{|r_s - r_{k\nu}| \geq 0.14R\right\}\right] \\ &\overset{(i)}{\leq} \frac{\mathbb{E}\left[(r_s - r_{k\nu})^8\right]}{(0.14R)^8} \\ &\overset{(ii)}{\leq} \frac{2^{10}\mathbb{E}\left[\|z_s - z_{k\nu}\|_2^8 + \|w_s - w_{k\nu}\|_2^8\right] + 2^{10}\beta^4}{(0.14R)^8},\end{aligned} \tag{59}$$

where (i) by Markov's inequality, (ii) can be verified by using Lemma 6.1 and some algebra. Next, by the dynamics of z_t we have that

$$\begin{aligned}\|z_s - z_{k\nu}\|_2^8 &= \left\|\int_{k\nu}^s w_s dt\right\|_2^8 \\ &\leq (s - k\nu)^7 \int_{k\nu}^s \|w_s\|_2^8 dt \\ &\leq 2^3 (s - k\nu)^7 \int_{k\nu}^s \|u_s\|_2^8 + \|v_s\|_2^8 dt.\end{aligned} \tag{60}$$

Further by the definition of the dynamics of w_t we get,

$$
\begin{aligned}
&\|w_s - w_{k\nu}\|_2^8 \\
&= \left\| \int_{k\nu}^s -2w_t - \frac{c_\kappa}{L}\nabla U(x_{\lfloor \frac{t}{\delta} \rfloor}) + \frac{c_\kappa}{L}\nabla U(y_t)dt + 4\sqrt{\frac{c_\kappa}{L}}\int_{k\nu}^s \gamma_t\gamma_t^T dB_t + 2\sqrt{\frac{c_\kappa}{L}}\int_{k\nu}^s \bar{\gamma}_t\bar{\gamma}_t^T dA_t \right\|_2^8 \\
&\overset{(i)}{\leq} 2^{20}(s-k\nu)^7\left(\int_{k\nu}^s \|w_t\|_2^8 + \frac{c_\kappa^8}{L^8}\|\nabla U(y_t)\|_2^8 + \frac{c_\kappa^8}{L^8}\left\|\nabla U(x_{\lfloor \frac{t}{\nu} \rfloor})\right\|_2^8 dt\right) \\
&\qquad + 2^{12}\frac{c_\kappa^4}{L^4}\left\|\int_{k\nu}^s \gamma_t\gamma_t^T dB_t\right\|_2^8 + 2^{12}\frac{c_\kappa^4}{L^4}\left\|\int_{k\nu}^s \bar{\gamma}_t\bar{\gamma}_t^T dA_t\right\|_2^8 \\
&\overset{(ii)}{\leq} 2^{30}(s-k\nu)^7\left(\int_{k\nu}^s \|u_t\|_2^8 + \|v_t\|_2^8 + c_\kappa^8\|y_t\|_2^8 + c_\kappa^8\left\|x_{\lfloor \frac{t}{\nu} \rfloor}\right\|_2^8 dt\right) \\
&\qquad + 2^{12}\frac{c_\kappa^4}{L^4}\left\|\int_{k\nu}^s \gamma_t\gamma_t^T dB_t\right\|_2^8 + 2^{12}\frac{c_\kappa^4}{L^4}\left\|\int_{k\nu}^s \bar{\gamma}_t\bar{\gamma}_t^T dA_t\right\|_2^8 \\
&\overset{(iii)}{\leq} 2^{30}(s-k\nu)^7\left(\int_{k\nu}^s \|u_t\|_2^8 + \|v_t\|_2^8 + \|y_t\|_2^8 + \left\|x_{\lfloor \frac{t}{\nu} \rfloor}\right\|_2^8 dt\right) \\
&\qquad + 2^{12}\frac{1}{L^4}\left\|\int_{k\nu}^s \gamma_t\gamma_t^T dB_t\right\|_2^8 + 2^{12}\frac{1}{L^4}\left\|\int_{k\nu}^s \bar{\gamma}_t\bar{\gamma}_t^T dA_t\right\|_2^8, \qquad (61)
\end{aligned}
$$

where (i) is by the triangle inequality and Young's inequality, (ii) uses Assumption (A1), and (iii) uses the fact that $c_\kappa \leq 1$.

Therefore, summing the two inequalities above and taking expectations,

$$
\begin{aligned}
&\mathbb{E}\left[\|z_s - z_{k\nu}\|_2^8 + \|w_s - w_{k\nu}\|_2^8\right] \\
&\quad \leq \mathbb{E}\left[2^{30}(s-k\nu)^7\left(\int_{k\nu}^s \left\|x_{\lfloor \frac{t}{\delta} \rfloor \delta}\right\|_2^8 + \|u_t\|_2^8 + \|y_t\|_2^8 + \|v_t\|_2^8 dt\right)\right] \\
&\qquad + \mathbb{E}\left[2^{12}\frac{1}{L^4}\left\|\int_{k\nu}^s \gamma_t\gamma_t^T dB_t\right\|_2^8 + 2^{12}\frac{1}{L^4}\left\|\int_{k\nu}^s \bar{\gamma}_t\bar{\gamma}_t^T dA_t\right\|_2^8\right] \\
&\quad \leq 2^{32}(s-k\nu)^8\left(R^2 + \frac{d}{m}\right)^4 + 2^{52}\cdot(s-k\nu)^4\cdot\frac{1}{L^4},
\end{aligned}
$$

where the last inequlaity is by combining Lemma 32, Lemma 33 and Lemma 22 and by noting that by their definition in Eq. (15), $\|\gamma_t\|_2 \leq 1$ and $\|\bar{\gamma}_t\|_2 \leq 1$ for all t, with probability one.

There exists $C_1 = poly(R, d, \frac{1}{m})$ and $C_3 = 1/poly(R, d, \frac{1}{m})$, such that for all $\nu < C_3$ and for all $s \in [k\nu, (k+1)\nu]$, the right-hand side of the inequality above is upper bounded by

$$
\mathbb{E}\left[\|z_s - z_{k\nu}\|_2^8 + \|w_s - w_{k\nu}\|_2^8\right] \leq \nu^4 C_1.
$$

Combining the above with inequality (59), we find that there exists $C_2 = poly(R, d, \frac{1}{m})$ and $C_3 = 1/poly(R, d, \frac{1}{m})$, such that for all $\nu < C_3$ and for all $s \in [k\nu, (k+1)\nu]$

$$\begin{aligned}\mathbb{E}\left[\mu_k \cdot \mathbb{1}\left\{r_s \geq \sqrt{12}R\right\}\right] \leq& \frac{\mathbb{E}\left[(r_s - r_{k\nu})^8\right]}{(0.14R)^8}\\ \leq& \frac{2^{10}\mathbb{E}\left[\|z_s - z_{k\nu}\|_2^8 + \|w_s - w_{k\nu}\|_2^8\right] + 2^{10}\beta^4}{(0.14R)^8}\\ \leq& \nu^4 C_2,\end{aligned}$$

where β is absorbed into C_2 due to our assumption that $\beta \leq 0.0001R$.

■

Corollary 21 *For $\beta \leq 0.0001R$. There exists constants, $C_3 = 1/poly(L, 1/m, d, R)$ and $C_4 = poly(L, 1/m, d, R)$, such that for all $\nu \leq C_3$, for all positive integers k, and for all $s \in [k\nu, (k+1)\nu]$,*

$$\mathbb{E}\left[\mu_k \mathbb{1}\left\{r_s \geq \sqrt{12}R\right\} r_s\right] \leq \sqrt{\mathbb{E}\left[\mu_k \mathbb{1}\left\{r_s \geq \sqrt{12}R\right\}\right]}\sqrt{\mathbb{E}\left[r_s^2\right]} \leq C_4\nu^2.$$

Proof Proof follows by combining the results of Lemma 19 and Lemma 20. ■

Lemma 22 *Let γ_t be a d-dimensional adapted process satisfying $\|\gamma_t\|_2 \leq 1$ for all $t > 0$ with probability one. Then*

$$\mathbb{E}\left[\left\|\int_0^t \gamma_s\gamma_s^T dB_s\right\|_2^8\right] \leq 2^{20}t^4.$$

Proof Let us define $\beta_t := \int_0^t \gamma_s\gamma_s^T dB_s$. Define the function $l(\beta) := \|\beta\|_2^8$ for this proof. The derivates of this function are,

$$\begin{aligned}\nabla l(\beta) &= 8l(\beta)^{3/4}\beta\\ \nabla^2 l(\beta) &= 8l(\beta)^{3/4}I + 48l(\beta)^{2/4}\beta\beta^T.\end{aligned}$$

By Itô's Lemma,

$$\begin{aligned}dl(\beta_t) =& \left\langle 8l(\beta_t)^{3/4}\beta_t, \beta_t\beta_t^T dB_t\right\rangle + 4l(\beta_t)^{3/4}\|\gamma_t\|_2^2 dt + 24l(\beta_t)^{2/4}(\langle\beta_t, \gamma_t\rangle)^2\|\gamma_t\|_2^2 dt\\ \leq& \left\langle 8l(\beta_t)^{3/4}\beta_t, \beta_t\beta_t^T dB_t\right\rangle + 4l(\beta_t)^{3/4}dt + 24l(\beta_t)^{2/4}\|\beta_t\|_2^2 dt\\ =& \left\langle 4l(\beta_t)^{3/4}\beta_t, \beta_t\beta_t^T dB_t\right\rangle + 28l(\beta_t)^{3/4}dt.\end{aligned}$$

Taking expectations,

$$\frac{d}{dt}\mathbb{E}\left[l(\beta_t)\right] \leq 28\mathbb{E}\left[l(\beta_t)^{3/4}\right] \leq 28\mathbb{E}\left[l(\beta_t)\right]^{3/4}.$$

Thus,

$$\begin{aligned}
& \frac{d}{dt}\mathbb{E}\left[l(\beta_t)\right]^{1/4} \leq 28 \\
\Rightarrow \quad & \mathbb{E}\left[l(\beta_t)\right]^{1/4} \leq 28t \\
\Rightarrow \quad & \mathbb{E}\left[l(\beta_t)\right] \leq 2^{20}t^4,
\end{aligned}$$

as claimed. ■

Lemma 23 *For all $t \geq 0$, $\mathbb{E}\left[\phi_t\right] = 0$.*

Proof By the definition of ϕ_t it is a martingale. Hence, $\mathbb{E}\left[\phi_t\right] = 0$. ■

D.5 Putting it all together

In this section, we combine the results from Appendices D.2, D.3 and D.4 to prove Theorem 3. The heart of the proof is Lemma 26, which shows that $\mathcal{L}_t$ contracts with probability one at a rate of $-C_m$. This lemma essentially combines the results of Lemmas 27, 28 (proved in Appendix D.2) and Lemmas 29, 30 (proved in Appendix D.3).

Proof [Proof of Theorem 3] From Lemma 26 we have,

$$\mathcal{L}_{k\nu} \leq e^{-C_m k\nu}\mathcal{L}_0 + \frac{(3\nu+5)\beta}{C_m\nu}. \tag{62}$$

while from Lemma 25,

$$f(r_{k\nu}) \leq 200\mathcal{L}_{k\nu} + 400\xi_{k\nu} + \sigma_{k\nu} + \phi_{k\nu} + 400\beta.$$

Taking expectations,

$$\begin{aligned}
&\mathbb{E}\left[f(r_{k\nu})\right] \\
&\quad \overset{(i)}{\leq} 200\mathbb{E}\left[\mathcal{L}_{k\nu}\right] + 400\mathbb{E}\left[\xi_{k\nu}\right] + \mathbb{E}\left[\sigma_{k\nu}\right] + \mathbb{E}\left[\phi_{k\nu}\right] + 400\beta \\
&\quad \overset{(ii)}{\leq} 200e^{-C_m k\nu}\mathbb{E}\left[\mathcal{L}_0\right] + 400\mathbb{E}\left[\xi_{k\nu}\right] + \mathbb{E}\left[\sigma_{k\nu}\right] + \mathbb{E}\left[\phi_{k\nu}\right] + \frac{2000(\nu+1)}{C_m\nu}\beta \\
&\quad = 200e^{-C_m k\nu}\mathbb{E}\left[f(r_0)\right] + 400\mathbb{E}\left[\xi_{k\nu}\right] + \mathbb{E}\left[\sigma_{k\nu}\right] + \mathbb{E}\left[\phi_{k\nu}\right] + \frac{2000(\nu+1)}{C_m\nu}\beta \\
&\quad \leq 200e^{-C_m k\nu}\mathbb{E}\left[r_0\right] + 400\mathbb{E}\left[\xi_{k\nu}\right] + \mathbb{E}\left[\sigma_{k\nu}\right] + \mathbb{E}\left[\phi_{k\nu}\right] + \frac{3000(\nu+1)}{C_m\nu}\beta,
\end{aligned}$$

where (i) is by Eq. (62) and (ii) can be verified from the initialization in Eq. (10) and the definition of the Lyapunov function $\mathcal{L}_t$ in Eq. (27).

From Lemmas 15, 18 and 23,

$$400\mathbb{E}\left[\xi_{k\nu}\right] + \mathbb{E}\left[\sigma_{k\nu}\right] + \mathbb{E}\left[\phi_{k\nu}\right] \leq \delta \cdot \frac{2^{18}c_\kappa\left(R + \sqrt{d/m}\right)}{C_m} + C_5\nu^2,$$

where $C_5 = poly(L, 1/m, d, R, 1/C_m)$ as defined in Lemma 18.

From Lemma 33, our choice of $x_0 = u_0 = 0$ in Eq. (10) and our definition of r_t in Eq. (18),

$$\mathbb{E}\left[r_0\right] \leq 3\mathbb{E}\left[\|y_0\|_2 + \|v_0\|_2\right] \leq 2^{10}\left(R + \sqrt{\frac{d}{m}}\right) + 3\beta.$$

By plugging the bound on $\mathbb{E}\left[r_0\right]$ and $\mathbb{E}\left[\xi_{k\nu}\right]$ into the bound on $\mathbb{E}\left[f(r_{k\nu})\right]$ above gives us that

$$\mathbb{E}\left[f(r_{k\nu})\right] \leq e^{-C_m k\nu} 2^{18}\left(R + \sqrt{\frac{d}{m}}\right) + \delta \cdot \frac{2^{18} c_\kappa \left(R + \sqrt{d/m}\right)}{C_m} + C_5 \nu^2 + \frac{3000(\nu+1)}{C_m \nu}\beta.$$

This inequality along with (F3) of Lemma 31, and Lemma 6.1 also implies that,

$$\begin{aligned}\mathbb{E}\left[\|z_{k\nu}\|_2\right] &\leq \mathbb{E}\left[r_{k\nu}\right] + \beta \\ &\leq 2e^{6LR^2} \cdot \mathbb{E}\left[f(r_{ku})\right] + \beta \\ &\leq e^{6LR^2} \cdot e^{-C_m k\nu} 2^{19}\left(R + \sqrt{\frac{d}{m}}\right) + e^{6LR^2} \cdot \delta \cdot \frac{2^{19} c_\kappa \left(R + \sqrt{d/m}\right)}{C_m} \\ &\quad + 2e^{6LR^2} \cdot \left(C_5 \nu^2 + \frac{3000(\nu+1)}{C_m \nu}\beta + \beta\right).\end{aligned}$$

We can take ν and β to be arbitrarily small without any additional computation cost, so let $\nu = \left(2^{20}\delta\left(R + \sqrt{d/m}\right)/(C_m C_5)\right)^{-1/2}$ and
$\beta = \min\left\{2^{20}\delta\left(R + \sqrt{d/m}\right)/(C_m), 2^9\delta\left(R + \sqrt{d/m}\right), 2^9\delta\nu\left(R + \sqrt{d/m}\right)\right\}$, so that the terms containing β and ν are less than the other terms.

We can ensure that the second term $\left(e^{6LR^2} \cdot \delta \cdot \frac{2^{19} c_\kappa \left(R + \sqrt{d/m}\right)}{C_m}\right)$ is less than $\varepsilon/2$ by setting

$$\delta = \varepsilon 2^{-20} e^{-6LR^2} \frac{C_m}{R + \sqrt{d/m}} \frac{1}{c_\kappa}.$$

We can ensure that the first term $\left(e^{6LR^2} \cdot e^{-C_m k\nu} 2^{19}\left(R + \sqrt{\frac{d}{m}}\right)\right)$ is less than $\varepsilon/2$ by setting

$$k\nu \geq \frac{\log\frac{1}{\varepsilon} + 6LR^2 + \log\left(2^{20}\left(R^2 + \frac{d}{m}\right)\right)}{C_m}.$$

Recalling the definition of $C_m := \min\left\{\frac{e^{-6LR^2}}{6000\kappa LR^2}, \frac{e^{-6LR^2}}{21\cdot 10^7 \cdot \log(100)\cdot\kappa^2}, \frac{1}{3\cdot 10^6 \kappa^2}\right\}$ in Eq. (9), and $c_\kappa := 1/(1000\kappa)$, some algebra shows that it suffices to let

$$\delta = \frac{\varepsilon}{R + \sqrt{d/m}} \cdot e^{-12LR^2} \cdot 2^{-35} \min\left(\frac{1}{LR^2}, \frac{1}{\kappa}\right).$$

The number of steps of the algorithm is thus

$$
\begin{aligned}
n = \frac{k\nu}{\delta} &\geq 2^{60} \cdot \frac{R + \sqrt{d/m}}{\varepsilon} \cdot e^{18LR^2} \cdot \kappa \cdot \max\left\{LR^2, \kappa\right\}^2 \cdot \left(\log\frac{1}{\varepsilon} + LR^2 + \log\left(R^2 + \frac{d}{m}\right)\right) \\
&= \widetilde{\mathcal{O}}\left(\frac{\sqrt{d}}{\varepsilon} e^{18LR^2}\right).
\end{aligned}
$$

This completes the proof. ∎

Lemma 24 *With probability one, for all positive integers k,*

$$
(1-\mu_k) \cdot f(r_{k\nu}) \leq (1-\mu_k) \cdot 2\left(f(r_{\tau_k}) + c_\kappa \int_{\tau_k}^{k\nu} e^{-\frac{c_\kappa^2}{3}(k\nu - t)} \left\|x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta}\right\|_2 dt\right) + 6\beta.
$$

Proof First, by Eq. (19) and Lemma 6.1,

$$
\begin{aligned}
(1-\mu_k) \cdot r_{k\nu} =& (1-\mu_k) \cdot ((1+2c_\kappa)\ell(z_{k\nu})_2 + \ell(z_{k\nu} + w_{k\nu})) \\
\leq& (1-\mu_k) \cdot ((1+2c_\kappa)\|z_{k\nu}\|_2 + \|z_{k\nu} + w_{k\nu}\|) + 3\beta.
\end{aligned}
$$

Note that by Lemma 14 we have,

$$
1 - \mu_k = \mathbb{1}\left\{\mu_k = 0\right\} = \prod_{i \in S_k} \mathbb{1}\left\{\mu_i = 0\right\} = \prod_{i \in S_k} (1 - \mu_i), \tag{63}
$$

where $S_k := \left\{\frac{\tau_k}{\nu}, ..., k\right\}$. Thus using this characterization of $1 - \mu_k$ we get,

$$
\begin{aligned}
&(1-\mu_k) \cdot ((1+2c_\kappa)\|z_{k\nu}\|_2 + \|z_{k\nu} + w_{k\nu}\|_2) \\
&\quad \overset{(i)}{\leq} (1-\mu_k) \cdot 2\sqrt{\|z_{k\nu}\|_2^2 + \|z_{k\nu} + w_{k\nu}\|_2^2} \\
&\quad \overset{(ii)}{\leq} (1-\mu_k) \cdot 2\left(\left(\sqrt{\|z_{k\nu}\|_2^2 + \|z_{k\nu} + w_{k\nu}\|_2^2} - \sqrt{2.2}R\right)_+ + \sqrt{2.2}R\right),
\end{aligned}
$$

where (i) is by defintion of c_κ in Eq. (6) and (ii) inequality is by algebra. Unpacking this further we get that:

$$
\begin{aligned}
&(1-\mu_k)\cdot\left((1+2c_\kappa)\|z_{k\nu}\|_2+\|z_{k\nu}+w_{k\nu}\|_2\right)\\
&\overset{(i)}{\le}(1-\mu_k)\cdot 2\left(\left(\prod_{i\in S_k}(1-\mu_i)\right)\cdot\left(\sqrt{\|z_{k\nu}\|_2^2+\|z_{k\nu}+w_{k\nu}\|_2^2}-\sqrt{2.2}R\right)_+ +\sqrt{2.2}R\right)\\
&\overset{(ii)}{\le}(1-\mu_k)\cdot 2\left(\left(\prod_{i\in S_k}(1-\mu_i)\right)\cdot e^{-\frac{c_\kappa^2}{3}(k\nu-\tau_k)}\left(\sqrt{\|z_{\tau_k}\|_2^2+\|z_{\tau_k}+w_{\tau_k}\|_2^2}-\sqrt{2.2}R\right)_+\right)\\
&\quad+(1-\mu_k)\cdot 2\left(\left(\prod_{i\in S_k}(1-\mu_i)\right)\cdot c_\kappa\int_{\tau_k}^{k\nu}e^{-\frac{c_\kappa^2}{3}(k\nu-t)}\left\|x_t-x_{\lfloor\frac{t}{\delta}\rfloor\delta}\right\|_2 dt\right)\\
&\quad+(1-\mu_k)\cdot 2\left(\sqrt{2.2}R\right)\\
&\overset{(iii)}{\le}(1-\mu_k)\cdot 2\left(\left(\sqrt{\|z_{\tau_k}\|_2^2+\|z_{\tau_k}+w_{\tau_k}\|_2^2}-\sqrt{2.2}R\right)_+ +\sqrt{2.2}R\right)\\
&\quad+(1-\mu_k)\cdot 2\left(c_\kappa\int_{\tau_k}^{k\nu}e^{-\frac{c_\kappa^2}{3}(k\nu-t)}\left\|x_t-x_{\lfloor\frac{t}{\delta}\rfloor\delta}\right\|_2 dt\right)\\
&\overset{(iv)}{=}(1-\mu_k)\cdot 2\left(\sqrt{\|z_{\tau_k}\|_2^2+\|z_{\tau_k}+w_{\tau_k}\|_2^2}\right)\\
&\quad+(1-\mu_k)\cdot 2\left(c_\kappa\int_{\tau_k}^{k\nu}e^{-\frac{c_\kappa^2}{3}(k\nu-t)}\left\|x_t-x_{\lfloor\frac{t}{\delta}\rfloor\delta}\right\|_2 dt\right)\\
&\overset{(v)}{\le}(1-\mu_k)\cdot 2\left(r_{\tau_k}+\left(c_\kappa\int_{\tau_k}^{k\nu}e^{-\frac{c_\kappa^2}{3}(k\nu-t)}\left\|x_t-x_{\lfloor\frac{t}{\delta}\rfloor\delta}\right\|_2 dt\right)\right)+3\beta,
\end{aligned}
$$

where (i) is by Eq. (63), (ii) follows by Lemma 12, applied recursively for $i\in\left\{\frac{\tau_k}{\nu}...k\right\}$, while (iii) is again by Eq. (63). The equality in (iv) can be verified as follows: By Lemma 14 we know that $\tau_{\tau_k/\nu}=\tau_k$, which implies that $\sqrt{\|z_{\tau_k}\|_2^2+\|z_{\tau_k}+w_{\tau_k}\|_2^2}\ge\sqrt{5}R$ based on the dynamics of τ_k in Eq. (17). Finally (v) is by definition of r_t in Eq. (19).

Our conclusion thus follows from the concavity of f and the fact that $f(0)=0$, so that for all $a,b,c\in\mathbb{R}^+$, $f(4b)\le 4f(b)$ and $a\le b+c$ implies that $f(a)\le f(b)+c$:

$$
(1-\mu_k)\cdot f(r_{k\nu})\le(1-\mu_k)\cdot 2\left(f(r_{\tau_k})+c_\kappa\int_{\tau_k}^{k\nu}e^{-\frac{c_\kappa^2}{3}(k\nu-t)}\left\|x_t-x_{\lfloor\frac{t}{\delta}\rfloor\delta}\right\|_2 dt\right)+6\beta.
$$

∎

Lemma 25 *For all positive integer k, with probability one,*

$$
f(r_{k\nu})\le 200\mathcal{L}_{k\nu}+400\xi_{k\nu}+\sigma_{k\nu}+\phi_{k\nu}+400\beta.
$$

Proof From Lemma 24,

$$(1-\mu_k)\cdot f(r_{k\nu}) \le 2(1-\mu_k)\cdot f(r_{\tau_k}) + 2(1-\mu_k)\cdot c_\kappa \int_{\tau_k}^{k\nu} e^{-\frac{c_\kappa^2}{3}(k\nu - t)} \left\| x_t - x_{\lfloor \frac{t}{\delta}\rfloor \delta}\right\|_2 dt + 6\beta$$
$$\le 2(1-\mu_k)\cdot f(r_{\tau_k}) + 2(1-\mu_k)\xi_{k\nu} + 6\beta, \tag{64}$$

where the last inequality is by Eq. (24).

We can also verify from the definition of μ_t in Eq. (18) that $\mu_k = 0 \Leftrightarrow k\nu \le \tau_k + T_{sync}$. Thus,

$$\begin{aligned}(1-\mu_k)\cdot e^{-C_m(k\nu-\tau_k)} &\overset{(i)}{\ge} (1-\mu_k)\cdot e^{-C_m T_{sync}}\\ &\overset{(ii)}{\ge} (1-\mu_k)\cdot \exp\left(-\frac{c_\kappa^2}{3}\cdot T_{sync}\right)\\ &= (1-\mu_k)\cdot \frac{1}{100},\end{aligned} \tag{65}$$

where (i) is by Eq. (9) and (ii) line is by Eq. (8).

Combining the above with the definition of $\xi_{k\nu}$ in Eq. (24) we get,

$$\begin{aligned}(1-\mu_k)\xi_{k\nu} &= (1-\mu_k)e^{-C_m(k\nu-\tau_k)}\xi_{\tau_k} + \int_{\tau_k}^{k\nu} e^{-C_m(k\nu-s)} c_\kappa \left\| x_s - x_{\lfloor \frac{s}{\delta}\rfloor\delta}\right\|_2 ds\\ &\ge (1-\mu_k)e^{-C_m(k\nu-\tau_k)}\xi_{\tau_k}.\end{aligned} \tag{66}$$

Thus,

$$\begin{aligned}\mathcal{L}_{k\nu} &\overset{(i)}{=} \mu_k(f(r_{k\nu}) - \xi_{k\nu}) + (1-\mu_k)\cdot e^{-C_m(k\nu-\tau_k)}\cdot (f(r_{\tau_k}) - \xi_{\tau_k}) - (\sigma_{k\nu} + \phi_{k\nu})\\ &\overset{(ii)}{\ge} \mu_k(f(r_{k\nu}) - \xi_{k\nu}) + (1-\mu_k)\cdot e^{-C_m(k\nu-\tau_k)}\cdot \left(\frac{1}{2}f(r_{k\nu}) - \xi_{k\nu} - \xi_{\tau_k} - 2\beta\right) - (\sigma_{k\nu} + \phi_{k\nu})\\ &\overset{(iii)}{\ge} \mu_k(f(r_{k\nu}) - \xi_{k\nu}) + (1-\mu_k)\cdot\cdot \left(\frac{e^{-C_m(k\nu-\tau_k)}}{2} f(r_{k\nu}) - 2\xi_{k\nu} - 2\beta\right) - (\sigma_{k\nu} + \phi_{k\nu})\\ &\overset{(iv)}{\ge} \mu_k(f(r_{k\nu}) - \xi_{k\nu}) + (1-\mu_k)\cdot \left(\frac{1}{200}f(r_{k\nu}) - 2\xi_{k\nu} - 2\beta\right) - (\sigma_{k\nu} + \phi_{k\nu})\\ &\overset{(v)}{\ge} \frac{1}{200}(\mu_k\cdot f(r_{k\nu}) + (1-\mu_k)\cdot f(r_{k\nu})) - (2\xi_{k\nu} + \sigma_{k\nu} + \phi_{k\nu}) - 2\beta\\ &\overset{(vi)}{=} \frac{1}{200}(f(r_{k\nu})) - (2\xi_{k\nu} + \sigma_{k\nu} + \phi_{k\nu}) - 2\beta,\end{aligned}$$

where (i) is by definition of $\mathcal{L}$ in Eq. (27). (ii) is by Eq. (64). (iii) is by Eq. (66) and the positivity of f, ξ, β. (iv) is by Eq. (65) and the fact that $f(r_t) \ge 0$ and $\xi_t \ge 0$ for all t. The inequalities (v) and (vi) are by algebraic manipulations.

Rearranging terms gives

$$f(r_{k\nu}) \le 200\mathcal{L}_{k\nu} + 400\xi_{k\nu} + \sigma_{k\nu} + \phi_{k\nu} + 400\beta.$$

■

Lemma 26 *Assume that $e^{72LR^2} \geq 2$. With probability one, for all positive integers k,*

$$\mathcal{L}_{k\nu} \leq e^{-C_m \nu} \mathcal{L}_{(k-1)\nu} + (3\nu + 5)\beta.$$

Applying this recursively,

$$\mathcal{L}_{k\nu} \leq e^{-C_m k\nu} \mathcal{L}_0 + \frac{(3\nu + 5)\beta}{C_m \nu}.$$

Proof We get the conclusion by summing the results of Lemmas 27, 28, 29 and 30. ■

Below, we state the lemmas which are needed to prove Lemma 26.

Lemma 27 *Assume that $e^{72LR^2} \geq 2$. For all positive integers k, with probability 1,*

$$\mathbb{1}\{\mu_k = 1, \mu_{k-1} = 0\} \cdot \mathcal{L}_{k\nu} \leq \mathbb{1}\{\mu_k = 1, \mu_{k-1} = 0\} \cdot e^{-C_m \nu} \mathcal{L}_{(k-1)\nu} + 5\beta.$$

Proof Given the definition of $\mathcal{L}_t$ in Eq. (27) we find that
$\mathbb{1}\{\mu_k = 1\}\mathcal{L}_{k\nu} = \mathbb{1}\{\mu_k = 1\} f(r_{k\nu})$ and
$\mathbb{1}\{\mu_{k-1} = 0\}\mathcal{L}_{(k-1)\nu} = \mathbb{1}\{\mu_{k-1} = 0\}\big(e^{-C_m((k-1)\nu - \tau_{k-1})} f(r_{\tau_{k-1}}) - \big(\sigma_{(k-1)\nu} + \phi_{(k-1)\nu}\big)\big)$.

By the dynamics of μ_k, we can verify that

$$\begin{aligned}
\mu_k = 1 \Leftrightarrow \quad & k\nu \geq \tau_k + T_{sync} \\
\Rightarrow \quad & k\nu \neq \tau_k \\
\Rightarrow \quad & \tau_k = \tau_{k-1} \\
\Rightarrow \quad & k\nu \geq \tau_{k-1} + T_{sync}.
\end{aligned}$$

We can also verify that

$$\mu_{k-1} = 0 \Rightarrow (k-1)\nu < \tau_{k-1} + T_{sync}.$$

By our choice of ν, T_{sync}/ν is an integer (see comment following Eq. (8)), and the inequalities above imply that $k\nu = \tau_{k-1} + T_{sync}$. Thus,

$$\mathbb{1}\{\mu_k = 1, \mu_{k-1} = 0\} = \mathbb{1}\{\mu_k = 1, \mu_{k-1} = 0\} \cdot \mathbb{1}\{k\nu = \tau_{k-1} + T_{sync}\}. \tag{67}$$

To reduce clutter, let us define $\alpha := \mathbb{1}\{\mu_k = 1, \mu_{k-1} = 0\}$ and $\alpha' := \mathbb{1}\{k\nu = \tau_{k-1} + T_{sync}\}$. Hence we have,

$$\begin{aligned}
\alpha \cdot \mathcal{L}_{k\nu} &\overset{(i)}{=} \alpha \cdot (f(r_{k\nu}) - \xi_{k\nu}) - \alpha \cdot (\sigma_{k\nu} + \phi_{k\nu}) \\
&\overset{(ii)}{=} \alpha \cdot \alpha'(f(r_{k\nu}) - \xi_{k\nu}) - \alpha \cdot (\sigma_{k\nu} + \phi_{k\nu}) \\
&\overset{(iii)}{\leq} \alpha \cdot \alpha' \cdot e^{-C_m T_{sync}}\big(f(r_{\tau_{k-1}}) - \xi_{\tau_{k-1}}\big) - \alpha \cdot (\sigma_{k\nu} + \phi_{k\nu}) + 5\beta \\
&\overset{(iv)}{=} \alpha \cdot \alpha' \cdot e^{-C_m(k\nu - \tau_{k-1})}\big(f(r_{\tau_{k-1}}) - \xi_{\tau_{k-1}}\big) - \alpha \cdot (\sigma_{k\nu} + \phi_{k\nu}) + 5\beta \\
&\overset{(v)}{=} \alpha \cdot \alpha' \cdot e^{-C_m \nu} e^{-C_m((k-1)\nu - \tau_{k-1})}\big(f(r_{\tau_{k-1}}) - \xi_{\tau_{k-1}}\big) - \alpha \cdot (\sigma_{k\nu} + \phi_{k\nu}) + 5\beta \\
&\overset{(vi)}{=} \alpha \cdot e^{-C_m \nu} e^{-C_m((k-1)\nu - \tau_{k-1})}\big(f(r_{\tau_{k-1}}) - \xi_{\tau_{k-1}}\big) - \alpha \cdot (\sigma_{k\nu} + \phi_{k\nu}) + 5\beta,
\end{aligned} \tag{68}$$

where (i) is by definition of $\mathcal{L}_{k\nu}$, (ii) by Eq. (67), (iii) is by Lemma 10, (iv) is by the fact that $\alpha' = \mathbb{1}\{T_{sync} = k\nu - \tau_{k-1}\}$, (v) is by algebra and finally (vi) is again by Eq. (67).

By definition of σ_t in Eq. (25),

$$\begin{aligned}\alpha \cdot \sigma_{k\nu} &= \alpha \int_0^{k\nu} \mu_{\lfloor \frac{s}{\nu} \rfloor} \cdot e^{-C_m(k\nu - s)} \cdot 4r_s ds \\ &\overset{(i)}{=} \alpha \int_0^{(k-1)\nu} \mu_{\lfloor \frac{s}{\nu} \rfloor} \cdot e^{-C_m(k\nu - s)} \cdot 4r_s ds \\ &= \alpha e^{-C_m \nu} \sigma_{(k-1)\nu}, \end{aligned} \tag{69}$$

where (i) is because $\alpha = 1$ implies that $\mu_{\lfloor \frac{s}{\nu} \rfloor} = \mu_{k-1} = 0$ for all $s \in [(k-1)\nu, k\nu)$.

Similarly, by the definition of ϕ_t in Eq. (26),

$$\begin{aligned}&\alpha \cdot \phi_{k\nu} \\ &= \alpha \int_0^{k\nu} \mu_{\lfloor \frac{s}{\nu} \rfloor} \cdot e^{-C_m(k\nu - s)} f'(r_s) q'(\|z_s + w_s\|_2) \left\langle \frac{z_s + w_s}{\|z_s + w_s\|_2}, 4\sqrt{\frac{c_\kappa}{L}} \left(\gamma_s \gamma_s^T dB_s + \frac{1}{2} \bar{\gamma}_s \bar{\gamma}_s^T dA_s \right) \right\rangle \\ &\overset{(i)}{=} \alpha \int_0^{(k-1)\nu} \mu_{\lfloor \frac{s}{\nu} \rfloor} \cdot e^{-C_m(k\nu - s)} f'(r_s) q'(\|z_s + w_s\|_2) \left\langle \frac{z_s + w_s}{\|z_s + w_s\|_2}, 4\sqrt{\frac{c_\kappa}{L}} \left(\gamma_s \gamma_s^T dB_s + \frac{1}{2} \bar{\gamma}_s \bar{\gamma}_s^T dA_s \right) \right\rangle \\ &= \alpha e^{-C_m \nu} \phi_{(k-1)\nu}, \end{aligned} \tag{70}$$

where (i) is again because $\alpha = 1$ implies that $\mu_{\lfloor \frac{s}{\nu} \rfloor} = \mu_{k-1} = 0$ for all $s \in [(k-1)\nu, k\nu)$.

Combining these results,

$$\begin{aligned}\alpha \mathcal{L}_{k\nu} &\overset{(i)}{\le} \alpha \cdot e^{-C_m \nu} e^{-C_m((k-1)\nu - \tau_{k-1})} \big(f(r_{\tau_{k-1}}) - \xi_{\tau_{k-1}} \big) \\ &\quad - \alpha \cdot (\sigma_{k\nu} + \phi_{k\nu}) + 5\beta \\ &\overset{(ii)}{=} \alpha \cdot e^{-C_m \nu} e^{-C_m((k-1)\nu - \tau_{k-1})} \big(f(r_{\tau_{k-1}}) - \xi_{\tau_{k-1}} \big) \\ &\quad - \alpha e^{-C_m \nu} \cdot \big(\sigma_{(k-1)\nu} + \phi_{(k-1)\nu} \big) + 5\beta \\ &\overset{(iii)}{=} \alpha \cdot e^{-C_m \nu} \cdot \mathcal{L}_{(k-1)\nu} + 5\beta, \end{aligned}$$

where (i) is by Eq. (68) and (ii) is by Eq. (69) and Eq. (70). Inequality (iii) is by the definition of $\mathcal{L}_t$ in Eq. (27), and because $\mathbb{1}\{\mu_{k-1} = 0\} \mathcal{L}_{(k-1)\nu} = \mathbb{1}\{\mu_{k-1} = 0\} \big(e^{-C_m((k-1)\nu - \tau_{k-1})} f(r_{\tau_{k-1}}) - \big(\sigma_{(k-1)\nu} + \phi_{(k-1)\nu} \big) \big)$, as noted in the beginning of the proof. ■

Lemma 28 *For all positive integers k, with probability one,*

$$\mathbb{1}\{\mu_k = 0, \mu_{k-1} = 0\} \cdot \mathcal{L}_{k\nu} \le \mathbb{1}\{\mu_k = 0, \mu_{k-1} = 0\} \cdot e^{-C_m \nu} \mathcal{L}_{(k-1)\nu} + 5\beta.$$

Proof Define α_1, α_2 and α_3 to be indicators for the following events:

$$\alpha_1 := \mathbb{1}\{\mu_k = 0, \mu_{k-1} = 0\}, \alpha_2 := \mathbb{1}\{k\nu = \tau_k\} \text{ and } \alpha_3 := \mathbb{1}\{k\nu = \tau_{k-1} + T_{sync}\}.$$

By the definition of the Lyapunov function in Eq. (27) we find that

$$\begin{aligned}
&\alpha_1 \cdot \mathcal{L}_{k\nu} = \alpha_1 \cdot \Big(e^{-C_m(k\nu-\tau_k)}(f(r_{\tau_k}) - \xi_{\tau_k}) - (\sigma_{k\nu} + \phi_{k\nu})\Big), \quad \text{and,} \\
&\alpha_1 \cdot \mathcal{L}_{(k-1)\nu} = \alpha_1 \cdot \Big(e^{-C_m((k-1)\nu-\tau_{k-1})}\big(f(r_{\tau_{k-1}}) - \xi_{\tau_{k-1}}\big) - \big(\sigma_{(k-1)\nu} + \phi_{(k-1)\nu}\big)\Big).
\end{aligned} \tag{71}$$

We now consider two cases: when $k\nu = \tau_k$ and when $k\nu \neq \tau_k$ and prove the result in both of these cases.

Case 1: $k\nu = \tau_k$
From the definition of τ_t in Eq. (17), we know that $k\nu = \tau_k \Rightarrow k\nu - \tau_{k-1} \geq T_{sync}$. Additionally, $\mu_{k-1} = 0 \Rightarrow (k-1)\nu - \tau_{k-1} < T_{sync}$. By our choice of ν; T_{sync}/ν is an integer (immediately below (8)). Thus it must be that $k\nu = \tau_{k-1} + T_{sync}$. Hence we have shown that

$$\alpha_1 \cdot \alpha_2 = \alpha_1 \cdot \alpha_2 \cdot \alpha_3. \tag{72}$$

Thus,

$$\begin{aligned}
&\alpha_1 \cdot \alpha_2 \cdot \mathcal{L}_{k\nu} \\
&\overset{(i)}{=} \alpha_1 \cdot \alpha_2 \cdot \alpha_3 \cdot \Big(e^{-C_m(k\nu-\tau_k)}(f(r_{\tau_k}) - \xi_{\tau_k}) - (\sigma_{\tau_k} + \phi_{\tau_k})\Big) \\
&\overset{(ii)}{=} \alpha_1 \cdot \alpha_2 \cdot \alpha_3 \cdot ((f(r_{k\nu}) - \xi_{k\nu}) - (\sigma_{k\nu} + \phi_{k\nu})) \\
&\overset{(iii)}{\leq} \alpha_1 \cdot \alpha_2 \cdot \alpha_3 \cdot \Big(e^{-C_m(k\nu-T_{sync})} \cdot \big(f(r_{\tau_{k-1}}) - \xi_{\tau_{k-1}}\big) - (\sigma_{k\nu} + \phi_{k\nu})\Big) + 5\beta \\
&\overset{(iv)}{=} \alpha_1 \cdot \alpha_2 \cdot \alpha_3 \cdot \Big(e^{-C_m(k\nu-T_{sync})} \cdot \big(f(r_{\tau_{k-1}}) - \xi_{\tau_{k-1}}\big) - e^{-C_m\nu}\big(\sigma_{(k-1)\nu} + \phi_{(k-1)\nu}\big)\Big) + 5\beta \\
&\overset{(v)}{=} \alpha_1 \cdot \alpha_2 \cdot \alpha_3 \cdot \big(e^{-C_m\nu}\mathcal{L}\big(\theta_{(k-1)\nu}\big)\big) + 5\beta \\
&\overset{(vi)}{=} \alpha_1 \cdot \alpha_2 \cdot \big(e^{-C_m\nu}\mathcal{L}\big(\theta_{(k-1)\nu}\big)\big) + 5\beta,
\end{aligned}$$

where (i) is by Eq. (72), (ii) is because $\alpha_2 = 1$ implies $\tau_k = k\nu$, (iii) is by Lemma 10. Inequality (iv) is because $\alpha_1 = 1$ implies $\mu_{k-1} = 0$, we can thus verify from Eq. (25) and Eq. (26) that $\alpha_1 \cdot (\sigma_{k\nu} + \phi_{k\nu}) = \alpha_1 \cdot e^{-C_m\nu}\big(\sigma_{(k-1)\nu} + \phi_{(k-1)\nu}\big)$ (the detailed proof is identical to proof of Eq. (69) and (70), and is not repeated here). (v) follows by our expression for $\mathcal{L}_{(k-1)\nu}$ in Eq. (71) and (vi) is again by Eq. (72).

Case 2: $k\nu \neq \tau_k$
In this case, by the definition of τ_t (in Eq. (17)) that $\tau_k = \tau_{k-1}$. Thus,

$$\begin{aligned}
&\alpha_1 \cdot (1-\alpha_2) \cdot \mathcal{L}_{k\nu} \\
&\overset{(i)}{=} \alpha_1 \cdot (1-\alpha_2) \cdot e^{-C_m(k\nu-\tau_k)}(f(r_{\tau_k}) - \xi_{\tau_k}) - \alpha_1 \cdot (1-\alpha_2) \cdot (\sigma_{k\nu} + \phi_{k\nu}) \\
&\overset{(ii)}{=} \alpha_1 \cdot (1-\alpha_2) \cdot e^{-C_m(k\nu-\tau_{k-1})}\big(f(r_{\tau_{k-1}}) - \xi_{\tau_{k-1}}\big) - \alpha_1 \cdot (1-\alpha_2) \cdot (\sigma_{k\nu} + \phi_{k\nu}) \\
&\overset{(iii)}{=} \alpha_1 \cdot (1-\alpha_2) \cdot e^{-C_m(k\nu-\tau_{k-1})}\big(f(r_{\tau_{k-1}}) - \xi_{\tau_{k-1}}\big) - \alpha_1 \cdot (1-\alpha_2) \cdot e^{-C_m\nu}\big(\sigma_{(k-1)\nu} + \phi_{(k-1)\nu}\big) \\
&\overset{(iv)}{=} \alpha_1 \cdot (1-\alpha_2) \cdot e^{-C_m\nu}\mathcal{L}_{(k-1)\nu},
\end{aligned}$$

where (i) is by the expression for $\mathcal{L}_{k\nu}$ in Eq. (71), (ii) is because $\tau_k = \tau_{k-1}$. Inequality (iii) is because $\alpha_1 \cdot (\sigma_{k\nu} + \phi_{k\nu}) = \alpha_1 \cdot e^{-C_m\nu}(\sigma_{(k-1)\nu} + \phi_{(k-1)\nu})$. The proof of this fact is identical to proof of inequalities Eqs. (69) and (70), and is not repeated here. Finally (iv) is by pulling out a factor of $e^{-C_m\nu}$, and then using the equality in Eq. (71).

Therefore, summing the two cases, we get our conclusion that

$$\mathbb{1}\left\{\mu_k = 0, \mu_{k-1} = 0\right\} \cdot \mathcal{L}_{k\nu} \leq \mathbb{1}\left\{\mu_k = 0, \mu_{k-1} = 0\right\} \cdot e^{-C_m\nu}\mathcal{L}_{(k-1)\nu} + 5\beta.$$

■

Lemma 29 *For all positive integers k, with probability 1,*

$$\mathbb{1}\left\{\mu_k = 1, \mu_{k-1} = 1\right\} \cdot \mathcal{L}_{k\nu} \leq \mathbb{1}\left\{\mu_k = 1, \mu_{k-1} = 1\right\} \cdot e^{-C_m\nu}\mathcal{L}_{(k-1)\nu} + 5\beta\nu.$$

Proof Let α denote the indicator of the following event, $\alpha := \mathbb{1}\left\{\mu_k = 1, \mu_{k-1} = 1\right\}$. By the definition of our Lyapunov function (see Eq. (27)) that

$$\begin{aligned}
&\alpha \cdot \mathcal{L}_{k\nu} = \alpha \cdot ((f(r_{k\nu}) - \xi_{k\nu}) - (\sigma_{k\nu} + \phi_{k\nu})), \quad \text{and,} \\
&\alpha \cdot \mathcal{L}_{(k-1)\nu} = \alpha \cdot \left(\left(f(r_{(k-1)\nu}) - \xi_{(k-1)\nu}\right) - \left(\sigma_{(k-1)\nu} + \phi_{(k-1)\nu}\right)\right).
\end{aligned} \tag{73}$$

Thus we have,

$$\begin{aligned}
\alpha \cdot \mathcal{L}_{k\nu} &\overset{(i)}{=} \alpha \cdot ((f(r_{k\nu}) - \xi_{k\nu}) - (\sigma_{k\nu} + \phi_{k\nu})) \\
&\overset{(ii)}{=} \alpha \cdot (\mu_k(f(r_{k\nu}) - \xi_{k\nu}) - (\sigma_{k\nu} + \phi_{k\nu})) \\
&\overset{(iii)}{\leq} \alpha \cdot \left(e^{-C_m\nu}\mu_k \cdot \left(f(r_{(k-1)\nu}) - \xi_{(k-1)\nu}\right) - \left(\sigma_{(k-1)\nu} + \phi_{(k-1)\nu}\right)\right) + 5\beta\nu \\
&\overset{(iv)}{=} \alpha \cdot \left(e^{-C_m\nu} \cdot \left(f(r_{(k-1)\nu}) - \xi_{(k-1)\nu}\right) - \left(\sigma_{(k-1)\nu} + \phi_{(k-1)\nu}\right)\right) + 5\beta\nu \\
&\overset{(v)}{=} \alpha \cdot e^{-C_m\nu}\mathcal{L}_{(k-1)\nu},
\end{aligned}$$

where (i) is by Eq. (73), (ii) is because $\alpha = \alpha \cdot \mu_k$, (iii) is by Lemma 9, (iv) is again because $\alpha = \alpha \cdot \mu_k$ and (v) is again by Eq. (73). ■

Lemma 30 *For all positive integers k, with probability 1,*

$$\mathbb{1}\left\{\mu_k = 0, \mu_{k-1} = 1\right\} \cdot \mathcal{L}_{k\nu} \leq \mathbb{1}\left\{\mu_k = 0, \mu_{k-1} = 1\right\} \cdot e^{-C_m\nu}\mathcal{L}_{(k-1)\nu} + 5\beta\nu.$$

Proof Let $\alpha := \mathbb{1}\left\{\mu_k = 0, \mu_{k-1} = 1\right\}$. We can verify using the definition of the Lyapunov function in Eq. (27) that:

$$\begin{aligned}
\alpha \cdot \mathcal{L}_{k\nu} &= \alpha \cdot \left(e^{-C_m(k\nu - \tau_k)}(f(r_{\tau_k}) - \xi_{\tau_k}) - (\sigma_{k\nu} + \phi_{k\nu})\right) \quad \text{and,} \\
\alpha \cdot \mathcal{L}_{(k-1)\nu} &= \alpha \cdot \left(\left(f(r_{(k-1)\nu}) - \xi_{(k-1)\nu}\right) - \left(\sigma_{(k-1)\nu} + \phi_{(k-1)\nu}\right)\right).
\end{aligned} \tag{74}$$

Additionally, we can verify from Eq. (18) that $\mu_k = 0$ implies that $k\nu - T_{sync} < \tau_k$ and that $\mu_{k-1} = 1$ implies thta $(k-1)\nu - T_{sync} \geq \tau_{k-1}$. Putting this together, we get

$$\tau_k > k\nu - T_{sync} > (k-1)\nu - T_{sync} \geq \tau_{k-1}.$$

Thus $\tau_k > \tau_{k-1}$. From the definition of μ_t (in Eq. (18)), we see that τ_k is either equal to τ_{k-1} or is equal to $k\nu$, so that it must be that

$$\tau_k = k\nu,$$

when $\alpha = 1$. In particular, this implies that

$$\begin{aligned}
\alpha \cdot \mathcal{L}_{k\nu} &\overset{(i)}{=} \alpha \cdot \left(e^{-C_m(k\nu-\tau_k)}(f(r_{\tau_k}) - \xi_{\tau_k}) - (\sigma_{k\nu} + \phi_{k\nu})\right)\\
&\overset{(ii)}{=} \alpha \cdot (\mu_k \cdot (f(r_{k\nu}) - \xi_{k\nu}) - (\sigma_{k\nu} + \phi_{k\nu}))\\
&\overset{(iii)}{\leq} \alpha \cdot \left(e^{-C_m\nu}\mu_k \cdot \left(f(r_{(k-1)\nu}) - \xi_{(k-1)\nu}\right) - \left(\sigma_{(k-1)\nu} + \phi_{(k-1)\nu}\right)\right) + 5\beta\nu\\
&\overset{(iv)}{\leq} \alpha \cdot \left(e^{-C_m\nu} \cdot \left(f(r_{(k-1)\nu}) - \xi_{(k-1)\nu}\right) - \left(\sigma_{(k-1)\nu} + \phi_{(k-1)\nu}\right)\right) + 5\beta\nu\\
&\overset{(v)}{=} \alpha \cdot e^{-C_m\nu}\mathcal{L}_{(k-1)\nu} + 5\beta\nu,
\end{aligned}$$

where (i) is by Eq. (74), (ii) is by $\alpha \cdot \mu_k = \alpha$ and because $\alpha = \alpha \cdot \mathbb{1}\{\tau_k = k\nu\}$, (iii) is by Lemma 9, (iv) is again by $\alpha \cdot \mu_k = \alpha$ and finally (v) is by Eq. (74). ■

Appendix E. Properties of f

Lemma 31 *Assume that $e^{72LR^2} \geq 2$. The function f defined in Eq. (23) has the following properties.*

(F1) $f(0) = 0$, $f'(0) = 1$.

(F2) $\frac{1}{2}e^{-2\alpha_f\mathcal{R}_f^2} \leq \frac{1}{2}\psi(r) \leq f'(r) \leq 1$.

(F3) $\frac{1}{2}e^{-2\alpha_f\mathcal{R}_f^2}r \leq \frac{1}{2}\Psi(r) \leq f(r) \leq \Psi(r) \leq r$.

(F4) For all $0 < r \leq \mathcal{R}_f$, $f''(r) + \alpha_f r f'(r) \leq -\frac{e^{-2\alpha_f\mathcal{R}_f^2}}{4\mathcal{R}_f^2}f(r)$

(F5) For all $r > 0$, f'' *is defined,* $f''(r) \leq 0$, *and* $f''(r) = 0$ *when* $r > 2\mathcal{R}_f$.

(F6) If $2\alpha_f\mathcal{R}_f^2 \geq \ln 2$, *for any* $0.5 < s < 1$, $f(sr) \leq \exp\left(-\frac{1-s}{4}e^{-2\alpha_f\mathcal{R}_f^2}\right)f(r)$.

(F7) For $r > 0$, $|f''(r)| \leq 4\alpha_f\mathcal{R}_f + \frac{4}{\mathcal{R}_f}$

Proof We refer to definitions of the functions ψ, Ψ, g in Eq. (22) and the definition of f in Eq. (23).

(F1) $f(0) = 0$ and $f'(0) = 1$ by the definition of f and ψ.

(F2),(F3) are verified from the definitions, noting that $\frac{1}{2} \leq g(r) \leq 1$ and $e^{-2\alpha_f\mathcal{R}_f^2} \leq \psi(2\mathcal{R}_f) \leq \psi(r) \leq \psi(0)$.

(F4) To prove this property first we observe that $f'(r) = \psi(r)g(r)$ so

$$f''(r) = \psi'(r)g(r) + \psi(r)g'(r).$$

By the definition of ψ, $\psi'(r) = -2\alpha_f r\psi(r)$ if $r < \mathcal{R}_f$, thus

$$\begin{aligned}
f''(r) + 2\alpha_f r f'(r) &= -2\alpha_f r\psi(r)g(r) + \psi(r)g'(r) + 2\alpha_f r f'(r)\\
&= \psi(r)g'(r)\\
&= -\frac{1}{2}\frac{h(r)\Psi(r)}{\int_0^\infty h(s)\frac{\Psi(s)}{\psi(s)}ds}\\
&\overset{(i)}{\le} -\frac{1}{2}\frac{f(r)}{\int_0^\infty h(s)\frac{\Psi(s)}{\psi(s)}ds}\\
&\overset{(ii)}{\le} -\frac{e^{-2\alpha_f\mathcal{R}_f^2}}{4\mathcal{R}_f^2}f(r),
\end{aligned}$$

where (i) is because $f(r) \le \Psi(r)$ and $h(r) = 1$ for $r \le \mathcal{R}_f$.

(ii) is because $f(r) \ge 0$ and

$$\int_0^\infty h(s)\frac{\Psi(s)}{\psi(s)}ds = \int_0^{2\mathcal{R}_f} h(s)\frac{\Psi(s)}{\psi(s)}ds \le \int_0^{2\mathcal{R}_f} \frac{2s}{e^{-2\alpha_f\mathcal{R}_f^2}}ds \le 4\mathcal{R}_f^2 e^{2\alpha_f\mathcal{R}_f^2}.$$

The first inequality above is by (F2), (F3) and the definition of $h(s)$.

(F5) $f''(r) \le 0$ follows from its expression $f''(r) = \psi'(r)g(r) + \psi(r)g'(r)$, and the fact that $\psi(r) \ge 0$ from (F2), $g(r) \ge 1/2$, $g'(r) \le 0$ and $\psi'(r) \le 0$ for all r. For $r > 2\mathcal{R}_f$, $\psi'(r) = g'(r) = 0$, so in that case $f''(r) = \psi'(r)g(r) + \psi(r)g'(r) = 0$.

(F6) For any $0 < c < 1$,

$$f((1+c)r) = f(r) + \int_r^{(1+c)r} f'(s)ds \ge f(r) + cr\cdot\frac{1}{2}e^{-2\alpha_f\mathcal{R}_f^2} \ge \left(1 + \frac{c}{2}e^{-2\alpha_f\mathcal{R}_f^2}\right)f(r),$$

where the first inequality follows from (F2), and the second inequality follows from (F3). Under the assumption that $e^{-2\alpha_f\mathcal{R}_f^2} \le \frac{1}{2}$, and using the inequality $1 + x \ge e^{x/2}$ for all $x \in [0, 1/2]$, we get $1 + (c/2)e^{-2\alpha_f\mathcal{R}_f^2} \ge e^{(c/4)e^{-2\alpha_f\mathcal{R}_f^2}}$.

Thus, for any $s \in (1/2, 1)$, let $r' := sr$, so that $r = \frac{1}{s}r' = \left(1 + \left(\frac{1}{s} - 1\right)\right)r'$ Applying the above with $c = \frac{1}{s} - 1$, we get

$$\begin{aligned}
f(sr) = f(r') &\le \frac{1}{1 + \frac{c}{2}\exp\left(-2\alpha_f\mathcal{R}_f^2\right)}f((1+c)r')\\
&= \frac{1}{1 + \frac{c}{2}\exp\left(-2\alpha_f\mathcal{R}_f^2\right)}f(r)\\
&\le \exp\left(-\frac{c}{4}e^{-2\alpha_f\mathcal{R}_f^2}\right)f(r)\\
&= \exp\left(-\frac{1/s - 1}{4}e^{-2\alpha_f\mathcal{R}_f^2}\right)f(r) \le \exp\left(-\frac{1-s}{4}e^{-2\alpha_f\mathcal{R}_f^2}\right)f(r).
\end{aligned}$$

where we use the fact that $-\frac{1-s}{s} \le -(1-s)$.

(F7) Recall that

$$f''(r) = \psi'(r)g(r) + \psi(r)g'(r)$$

Thus

$$\begin{aligned}\left|f''(r)\right| &\le \left|\psi'(r)g(r)\right| + \left|\psi(r)g'(r)\right| \\ &\le 2\alpha_f r h(r) + \left|\psi(r)g'(r)\right|\end{aligned}$$

From our definition of $h(r)$, we know that $rh(r) \le 2\mathcal{R}_f$. In addition, since $\psi(r)$ is monotonically decreasing, $\Psi(r) = \int_0^r \psi(s)ds \ge r\psi(r)$, so that

$$\frac{\Psi(r)}{\psi(r)} \ge r. \tag{75}$$

Thus $\Psi(r)/r \ge r$ for all r. On the other hand, using the fact that $\psi(s) \le 1$,

$$\Psi(r) = \int_0^r \psi(s)ds \le r. \tag{76}$$

Combining the previous expressions,

$$\begin{aligned}\left|\psi(r)\nu'(r)\right| &= \left|\frac{1}{2}\frac{h(r)\Psi(r)}{\int_0^{4\mathcal{R}_f}\frac{\mu(s)\Psi(s)}{\psi(s)}ds}\right| \\ &\le \left|\frac{1}{2}\frac{2\mathcal{R}_f}{\int_0^{\mathcal{R}_f}\frac{\Psi(s)}{\psi(s)}ds}\right| \\ &\le \left|\frac{1}{2}\frac{2\mathcal{R}_f}{\int_0^{\mathcal{R}_f} sds}\right| \\ &\le \frac{4}{\mathcal{R}_f},\end{aligned}$$

where the first inequality is by the definition of $h(r) = 1$ for $r \le \mathcal{R}_f$ and $h(r) = 0$ for $r \ge 2\mathcal{R}_f$, and the second-to-last inequality is by (75).

Put together, we get

$$\left|f''(r)\right| \le 4\alpha_f\mathcal{R}_f + \frac{4}{\mathcal{R}_f}.$$

■

Appendix F. Bounding moments

To bound the discretization error it is necessary to bound the moments of the random variables x_t, u_t and y_t, v_t. The main results of this section are Lemma 32 (which bounds the moments of x_t and u_t) and Lemma 33 (which bounds the moments of y_t and v_t).

Lemma 32 *For $\delta \leq 2^{-10} c_\kappa$, and for all $t \geq 0$,*

$$\mathbb{E}\left[\|x_t\|_2^8 + \|x_t + u_t\|_2^8\right] \leq 2^{70}\left(R^2 + \frac{d}{m}\right)^4.$$

Lemma 33 *For all $t \geq 0$,*

$$\mathbb{E}\left[\|y_t\|_2^8 + \|y_t + v_t\|_2^8\right] \leq 2^{66}\left(R^2 + \frac{d}{m}\right)^4.$$

F.1 Proof of Lemma 32

Let us consider the Lyapunov function $l(x_t, u_t) := \left(\|x_t\|_2^2 + \|x_t + u_t\|_2^2 - 4R^2\right)_+^4$.

By calculating the derivaties of l we can verify that:

$$\begin{aligned}
\nabla_x l(x_t, u_t) &= 8 l(x_t, u_t)^{3/4}(x_t) \\
\nabla_u l(x_t, u_t) &= 8 l(x_t, u_t)^{3/4}(x_t + u_t) \\
\nabla_u^2 l(x_t, u_t) &= 8 l(x_t, u_t)^{3/4} I + 24 l(x_t, u_t)^{2/4}(x_t + u_t)(x_t + u_t)^T.
\end{aligned}$$

The following are two useful inequalities which we will use in this proof:

$$\begin{aligned}
\|x\|_2^2 + \|x + u\|_2^2 &\leq l(x, u)^{1/4} + 4R^2 \\
\|x\|_2^2 + \|x + u\|_2^2 &\geq l(x, u)^{1/4}.
\end{aligned} \tag{77}$$

Recall from the dynamics defined in Eq. (11) and Eq. (12) that

$$\begin{aligned}
dx_t =& u_t dt \\
du_t =& -2u_t - \frac{c_\kappa}{L}\nabla U(x_{\lfloor \frac{t}{\delta} \rfloor \delta})dt + 2\sqrt{\frac{c_\kappa}{L}} dB_t.
\end{aligned}$$

Thus by studying the evolution of the Lyapunov function $l(x_t, u_t)$ we have:

$$\begin{aligned}
\frac{d}{dt}\mathbb{E}\left[l(x_t,u_t)\right] &= \mathbb{E}\left[8l(x_t,u_t)^{3/4}\Big(\langle x_t,u_t\rangle + \Big\langle x_t+u_t, -u_t - \frac{c_\kappa}{L}\nabla U(x_{\lfloor \frac{t}{\delta}\rfloor\delta})\Big\rangle\Big)\right] \\
&\quad + \mathbb{E}\left[\frac{16c_\kappa}{L}\Big(l(x_t,u_t)^{3/4}d + 3l(x_t,u_t)^{2/4}\|x_t+u_t\|_2^2\Big)\right] \\
&= \mathbb{E}\left[\underbrace{8l(x_t,u_t)^{3/4}\Big(\langle x_t,u_t\rangle + \Big\langle x_t+u_t, -u_t - \frac{c_\kappa}{L}\nabla U(x_t)\Big\rangle\Big)}_{=:\spadesuit}\right] \\
&\qquad + \mathbb{E}\left[\underbrace{8\cdot\frac{c_\kappa}{L}\cdot l(x_t,u_t)^{3/4}\Big(\Big\langle x_t+u_t, \nabla U(x_t) - \nabla U(x_{\lfloor \frac{t}{\delta}\rfloor\delta})\Big\rangle\Big)}_{=:\heartsuit}\right] \\
&\qquad + \mathbb{E}\left[\underbrace{\frac{16c_\kappa}{L}\Big(l(x_t,u_t)^{3/4}d + 3l(x_t,u_t)^{2/4}\|x_t+u_t\|_2^2\Big)}_{=:\clubsuit}\right].
\end{aligned}$$

We will bound the three terms separately. We begin by bounding $\spadesuit$:

$$\begin{aligned}
\spadesuit &= 8l(x_t,u_t)^{3/4}\Big(\langle x_t,u_t\rangle + \Big\langle x_t+u_t, -u_t - \frac{c_\kappa}{L}\nabla U(x_t)\Big\rangle\Big) \\
&\overset{(i)}{\le} -c_\kappa^2 l(x_t,u_t)^{3/4}\Big(\|x_t\|_2^2 + \|x_t+u_t\|_2^2\Big) \\
&\overset{(ii)}{\le} -c_\kappa^2 l(x_t,u_t),
\end{aligned}$$

where (i) is by invoking Lemma 35, and (ii) is by Eq. (77). Next consider the term $\heartsuit$:

$$\begin{aligned}
\heartsuit &= 8 \cdot \frac{c_\kappa}{L} \cdot l(x_t, u_t)^{3/4} \left(\left\langle x_t + u_t, \nabla U(x_t) - \nabla U(x_{\lfloor \frac{t}{\delta} \rfloor \delta}) \right\rangle \right) \\
&\overset{(i)}{\leq} 8 c_\kappa l(x_t, u_t)^{3/4} \|x_t + u_t\|_2 \left\| x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta} \right\|_2 \\
&\overset{(ii)}{\leq} 8 c_\kappa l(x_t, u_t)^{3/4} \left(l(x_t, u_t)^{1/8} + 2R \right) \left\| x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta} \right\|_2 \\
&\overset{(iii)}{\leq} 8 c_\kappa l(x_t, u_t)^{7/8} \left\| x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta} \right\|_2 + 16 c_\kappa l(x_t, u_t)^{3/4} R \left\| x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta} \right\|_2 \\
&\overset{(iv)}{\leq} \frac{c_\kappa^2}{8} l(x_t, u_t) + \frac{2^{32}}{c_\kappa^6} \left(\left\| x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta} \right\|_2^8 \right) + \frac{c_\kappa^2}{8} l(x_t, u_t) + \frac{2^{28}}{c_\kappa^2} \left(R^4 \left\| x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta} \right\|_2^4 \right) \\
&\overset{(v)}{\leq} \frac{c_\kappa^2}{4} l(x_t, u_t) + \frac{2^{32}}{c_\kappa^6} \left(\left\| x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta} \right\|_2^8 \right) + 2^{28} c_\kappa^2 R^8 + \frac{2^{28}}{c_\kappa^6} \left\| x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta} \right\|_2^8 \\
&\overset{(vi)}{\leq} \frac{c_\kappa^2}{4} l(x_t, u_t) + 2^{28} c_\kappa^2 R^8 + \frac{2^{33}}{c_\kappa^6} \left\| x_t - x_{\lfloor \frac{t}{\delta} \rfloor \delta} \right\|_2^8 \\
&\overset{(vii)}{\leq} \frac{c_\kappa^2}{4} l(x_t, u_t) + 2^{28} c_\kappa^2 R^8 + \frac{2^{33}}{c_\kappa^6} \left\| \int_{\lfloor \frac{t}{\delta} \rfloor \delta}^t u_s ds \right\|_2^8 \\
&\overset{(viii)}{\leq} \frac{c_\kappa^2}{4} l(x_t, u_t) + 2^{28} c_\kappa^2 R^8 + \frac{2^{33}}{c_\kappa^6} \left(\left(t - \left\lfloor \frac{t}{\delta} \right\rfloor \delta \right)^7 \int_{\lfloor \frac{t}{\delta} \rfloor \delta}^t \|u_s\|_2^8 ds \right) \\
&\overset{(ix)}{\leq} \frac{c_\kappa^2}{4} l(x_t, u_t) + 2^{28} c_\kappa^2 R^8 + \frac{2^{33}}{c_\kappa^6} \left(\delta^7 \int_{\lfloor \frac{t}{\delta} \rfloor \delta}^t \|u_s\|_2^8 ds \right),
\end{aligned}$$

where (i) is by Cauchy-Schwarz and Assumption (A1), (ii) is by Eq. (77), (iii) is again by Eq. (77), (iv) is by Young's inequality, (v) is again by Young's inequality, (vi) follows by an algebraic manipulation, (vii) is by the dynamics defined in Eq. (11), $(viii)$ is by Jensen's inequality and finally (ix) is because $t - \lfloor \frac{t}{\delta} \rfloor \delta \leq \delta$. Also:

$$\begin{aligned}
\clubsuit &= \frac{16 c_\kappa}{L} \left(l(x_t, u_t)^{3/4} d + 3 l(x_t, u_t)^{2/4} \|x_t + u_t\|_2^2 \right) \\
&\overset{(i)}{\leq} \frac{16 c_\kappa}{L} \left(l(x_t, u_t)^{3/4} d + 3 l(x_t, u_t)^{3/4} + 12 l(x_t, u_t)^{2/4} R^2 \right) \\
&\overset{(ii)}{\leq} \frac{c_\kappa^2}{16} l(x_t, u_t) + \frac{2^{28}}{c_\kappa^2 L^4} d^4 + \frac{c_\kappa^2}{16} l(x_t, u_t) + \frac{2^{36}}{c_\kappa^2 L^4} + \frac{c_\kappa^2}{16} l(x_t, u_t) + \frac{2^{16} c_\kappa^2 R^4}{L^2} \\
&\overset{(iii)}{\leq} \frac{c_\kappa^2}{4} l(x_t, u_t) + 2^{29} c_\kappa^2 \left(\frac{d^4}{m^4} + \frac{R^4}{L^2} \right) \\
&\overset{(iv)}{\leq} \frac{c_\kappa^2}{4} l(x_t, u_t) + 2^{30} c_\kappa^2 \left(\frac{d^4}{m^4} + R^8 \right),
\end{aligned}$$

where (i) is by Eq. (77), (ii) is by Young's inequality, (iii) follows by definition of c_κ in Eq. (6) and (iv) is by Young's inequality, and because $m \leq L$.

Putting together the upper bounds on $\spadesuit, \heartsuit, \clubsuit$:

$$
\begin{aligned}
&\frac{d}{dt}\mathbb{E}\left[l(x_t, u_t)\right] = \spadesuit + \heartsuit + \clubsuit \\
&\leq \mathbb{E}\left[-c_\kappa^2 l(x_t, u_t) + 2^{28} c_\kappa^2 R^8 + \frac{2^{33}}{c_\kappa^6}\left(\delta^7 \int_{\lfloor \frac{t}{\delta} \rfloor \delta}^t \|u_s\|_2^8 ds\right) + \frac{c_\kappa^2}{4} l(x_t, u_t) + 2^{30} c_\kappa^2 \left(\frac{d^4}{m^4} + R^8\right)\right] \\
&\leq -\frac{c_\kappa^2}{2}\mathbb{E}\left[l(x_t, u_t)\right] + 2^{33} c_\kappa^2 \left(\frac{d^4}{m^4} + R^8\right) + \frac{2^{33}}{c_\kappa^6}\delta^7 \int_{\lfloor \frac{t}{\delta} \rfloor \delta}^t \mathbb{E}\left[\|u_s\|_2^8\right] ds \\
&\overset{(i)}{\leq} -\frac{c_\kappa^2}{2}\mathbb{E}\left[l(x_t, u_t)\right] + 2^{33} c_\kappa^2 \left(\frac{d^4}{m^4} + R^8\right) + \frac{2^{33}}{c_\kappa^6}\delta^8 \left(1.1\mathbb{E}\left[\left(\left\|x_{\lfloor \frac{t}{\delta} \rfloor \delta}\right\|_2^8 + \left\|u_{\lfloor \frac{t}{\delta} \rfloor \delta}\right\|_2^8\right)\right] + 2\left(\frac{d}{m}\right)^4\right) \\
&\overset{(ii)}{\leq} -\frac{c_\kappa^2}{2}\mathbb{E}\left[l(x_t, u_t)\right] + 2^{33} c_\kappa^2 \left(\frac{d^4}{m^4} + R^8\right) + \frac{c_\kappa^2}{8}\left(\mathbb{E}\left[l\left(x_{\lfloor \frac{t}{\delta} \rfloor \delta}, u_{\lfloor \frac{t}{\delta} \rfloor \delta}\right)\right] + R^8 + \left(\frac{d}{m}\right)^4\right) \\
&\leq -\frac{c_\kappa^2}{2}\mathbb{E}\left[l(x_t, u_t)\right] + 2^{34} c_\kappa^2 \left(\frac{d^4}{m^4} + R^8\right) + \frac{c_\kappa^2}{8}\mathbb{E}\left[l\left(x_{\lfloor \frac{t}{\delta} \rfloor \delta}, u_{\lfloor \frac{t}{\delta} \rfloor \delta}\right)\right],
\end{aligned} \tag{78}
$$

where (i) is by Lemma 34, and (ii) is by Eq. (77) and Eq. (6) along with some algebra.

Consider an arbitrary positive interger k. By Grönwall's Lemma applied over $s \in [k\delta, (k+1)\delta)$,

$$
\begin{aligned}
&\mathbb{E}\left[l(x_{(k+1)\delta}, u_{(k+1)\delta}\right] \\
&\quad \leq e^{-\frac{c_\kappa^2}{2}\delta}\mathbb{E}\left[l(x_{k\delta}, u_{k\delta})\right] + \delta \cdot \left(2^{34} c_\kappa^2 \left(\frac{d^4}{m^4} + R^8\right) + \frac{c_\kappa^2}{8}\mathbb{E}\left[l\left(x_{\lfloor \frac{t}{\delta} \rfloor \delta}, u_{\lfloor \frac{t}{\delta} \rfloor \delta}\right)\right]\right) \\
&\quad \overset{(i)}{\leq} \left(1 - \frac{c_\kappa^2 \delta}{4}\right)\mathbb{E}\left[l(x_{k\delta}, u_{k\delta})\right] + \delta \cdot \left(2^{34} c_\kappa^2 \left(\frac{d^4}{m^4} + R^8\right) + \frac{c_\kappa^2}{8}\mathbb{E}\left[l\left(x_{\lfloor \frac{t}{\delta} \rfloor \delta}, u_{\lfloor \frac{t}{\delta} \rfloor \delta}\right)\right]\right) \\
&\quad \overset{(ii)}{\leq} e^{-\frac{c_\kappa^2 \delta}{8}}\mathbb{E}\left[l(x_{k\delta}, u_{k\delta})\right] + 2^{34} c_\kappa^2 \delta \left(2^{34}\left(\frac{d^4}{m^4} + R^8\right)\right),
\end{aligned}
$$

where (i) and (ii) use the fact that $c_\kappa^2 \delta \leq \frac{1}{10}$, along with $1 - a \leq e^{-a} \leq 1 - \frac{a}{2}$ for $|a| \leq \frac{1}{10}$.

Applying the above recursively, using the geometric sum, and Eq. (10), we show that for all positive integers k,

$$
\mathbb{E}\left[l(x_{k\delta}, u_{k\delta})\right] \leq 2^{38}\left(\frac{d^4}{m^4} + R^8\right).
$$

For an arbitrary $t \geq 0$, we can similarly verify using the above result, Eq. (78), and Grönwall's Lemma that

$$
\mathbb{E}\left[l(x_t, u_t)\right] \leq 2^{39}\left(\frac{d^4}{m^4} + R^8\right).
$$

This completes the proof of the lemma.

We now state and prove some auxillary lemmas that were useful in the proof above.

Lemma 34 *Assume that $\delta \le \frac{1}{1000}$. Then for all $t \ge 0$,*

$$\mathbb{E}\left[\|x_t\|_2^8 + \|u_t\|_2^8\right] \le 1.1\mathbb{E}\left[\left(\left\|x_{\lfloor\frac{t}{\delta}\rfloor\delta}\right\|_2^8 + \left\|u_{\lfloor\frac{t}{\delta}\rfloor\delta}\right\|_2^8\right)\right] + 2\left(\frac{d}{m}\right)^4.$$

Proof From the stochastic dynamics defined in Eq. (11), Eq. (12), Eq. (13) and Eq. (14), we can verify that

$$\begin{aligned}
\frac{d}{dt}\mathbb{E}\left[\left(\|x_t\|_2^8 + \|u_t\|_2^8\right)\right] &\overset{(i)}{=} \mathbb{E}\left[8\|x_t\|_2^6\left\langle x_t, u_t\right\rangle + 8\|u_t\|_2^6\left\langle u_t, -2u_t - \frac{c_\kappa}{L}\nabla U(x_{\lfloor\frac{t}{\delta}\rfloor\delta})\right\rangle\right] \\
&\quad + \mathbb{E}\left[\frac{8c_\kappa d}{L}\|u_t\|_2^6 + \frac{48c_\kappa}{L}\|u_t\|_2^6\right] \\
&\overset{(ii)}{\le} 8\mathbb{E}\left[\|x_t\|_2^8 + \|u_t\|_2^8 + \|u_t\|_2^8 + c_\kappa^8\left\|x_{\lfloor\frac{t}{\delta}\rfloor\delta}\right\|_2^8\right] + \mathbb{E}\left[\frac{d}{m}\|u_t\|_2^6\right] \\
&\overset{(iii)}{\le} 64\mathbb{E}\left[\|x_t\|_2^8 + \|u_t\|_2^8\right] + \mathbb{E}\left[\left\|x_{\lfloor\frac{t}{\delta}\rfloor\delta}\right\|_2^8\right] + \left(\frac{d}{m}\right)^4,
\end{aligned}$$

where (i) is by Itô's Lemma, (ii) is by Assumption (A1), Young's inequality and by the definition of c_κ in Eq. (6), and (iii) is again by Young's inequality and definition of c_κ.

Consider an arbitrary $t \ge 0$, and let $k := \lfloor\frac{t}{\delta}\rfloor$. Then for all $s \in [k\delta, (k+1)\delta)$, we have:

$$\begin{aligned}
&\mathbb{E}\left[\left(\|x_t\|_2^8 + \|u_t\|_2^8\right)\right] \\
&\quad \le e^{64(s-k\delta)}\mathbb{E}\left[\left(\left\|x_{\lfloor\frac{t}{\delta}\rfloor\delta}\right\|_2^8 + \left\|u_{\lfloor\frac{t}{\delta}\rfloor\delta}\right\|_2^8\right)\right] + \left(e^{64(s-k\delta)} - 1\right)\left(\mathbb{E}\left[\left\|x_{\lfloor\frac{t}{\delta}\rfloor\delta}\right\|_2^8\right] + \left(\frac{d}{m}\right)^4\right) \\
&\quad \le (1 + 128\delta)\mathbb{E}\left[\left(\left\|x_{\lfloor\frac{t}{\delta}\rfloor\delta}\right\|_2^8 + \left\|u_{\lfloor\frac{t}{\delta}\rfloor\delta}\right\|_2^8\right)\right] + 128\delta\left(\frac{d}{m}\right)^4 \\
&\quad \le 1.1\mathbb{E}\left[\left(\left\|x_{\lfloor\frac{t}{\delta}\rfloor\delta}\right\|_2^8 + \left\|u_{\lfloor\frac{t}{\delta}\rfloor\delta}\right\|_2^8\right)\right] + 2\frac{d}{m},
\end{aligned}$$

where the final two inequalities are both by our assumption that $\delta \le \frac{1}{1000}$. ■

Lemma 35 *For (x_t, u_t) satisfying $\|x_t\|_2^2 + \|x_t + u_t\|_2^2 \ge 4R^2$,*

$$\left\langle x_t, u_t\right\rangle + \left\langle x_t + u_t, -u_t - \frac{c_\kappa}{L}\nabla U(x_t)\right\rangle \le -\frac{c_\kappa^2}{6}\left(\|x_t\|_2^2 + \|x_t + u_t\|_2^2\right).$$

Proof We first verify that

$$\begin{aligned}
&\left\langle x_t, u_t\right\rangle + \left\langle x_t + u_t, -u_t - \frac{c_\kappa}{L}\nabla U(x_{i\nu})\right\rangle \\
&\quad = -\|u_t\|_2^2 - \frac{c_\kappa}{L}\left\langle x_t, \nabla U(x_t)\right\rangle - \left\langle u_t, \frac{c_\kappa}{L}\nabla U(x_t)\right\rangle \\
&\quad = -\|u_t\|_2^2 - \frac{c_\kappa}{L}\left\langle x_t, \nabla U(x_t)\right\rangle - \frac{c_\kappa}{L}\left\langle u_t, \nabla U(x_t)\right\rangle \\
&\quad = -\|u_t\|_2^2 - \frac{c_\kappa}{L}\left\langle x_t, \nabla U(x_t)\right\rangle + \frac{1}{2}\|u_t\|_2^2 + \frac{c_\kappa^2}{2L^2}\|\nabla U(x_t)\|_2^2 - \frac{1}{2}\left\|u_t + \frac{c_\kappa}{L}\nabla U(x_t)\right\|_2^2 \\
&\quad \le -\frac{1}{2}\|u_t\|_2^2 - \frac{c_\kappa}{L}\left\langle x_t, \nabla U(x_t)\right\rangle + \frac{c_\kappa^2}{2}\|x_t\|_2^2 =: \spadesuit \qquad (79)
\end{aligned}$$

Now consider two cases:

Case 1: ($\|x_t\|_2 \leq R$) By Young's inequality we get that,

$$\|x_t + u_t\|_2^2 \leq 11\|u_t\|_2^2 + 1.1\|x_t\|_2^2.$$

Furthermore, by our assumption that $\|x_t\|_2^2 + \|x_t + u_t\|_2^2 \geq 4R^2$,

$$\begin{aligned} 11\|u_t\|_2^2 \geq& \|x_t + u_t\|_2^2 - 1.1\|x_t\|_2^2 \\ \geq& 4R^2 - 2.1R^2 \\ \geq& 1.9R^2 \\ \geq& 1.9\|x_t\|_2^2. \end{aligned} \tag{80}$$

Thus in this case $\|u_t\|_2^2 \geq \frac{1}{10}R^2$, and ♠ can be upper bounded by

$$\begin{aligned} ♠ =& -\frac{1}{2}\|u_t\|_2^2 - \left\langle x_t, \frac{c_\kappa}{L}\nabla U(x_t)\right\rangle + \frac{c^2}{2}\|x_t\|_2^2 \\ \overset{(i)}{\leq}& -\frac{1}{2}\|u_t\|_2^2 + c_\kappa\|x_t\|_2^2 + \frac{c_\kappa^2}{2}\|x_t\|_2^2 \\ \overset{(ii)}{\leq}& -\frac{1}{2}\|u_t\|_2^2 + 2c_\kappa\|x_t\|_2^2 \\ \overset{(iii)}{\leq}& -\frac{1}{4}\|u_t\|_2^2 \\ \overset{(iv)}{\leq}& -\frac{1}{160}\Big(\|x_t\|_2^2 + \|x_t + u_t\|_2^2\Big), \end{aligned}$$

where (i) is by L-Lipschitz of ∇U and Cauchy-Schwarz, (ii) and (iii) are because $c_\kappa := \frac{1}{1000\kappa} \leq \frac{1}{1000}$ and by Eq. (80), the (iv) is because

$$\begin{aligned} \|x_t\|_2^2 + \|x_t + u_t\|_2^2 \leq& 3\|x_t\|_2^2 + 2\|u_t\|_2^2 \\ \leq& 30\|u_t\|_2 + 2\|u_t\|_2 \\ \leq& 40\|u_t\|_2^2, \end{aligned}$$

where the second inequality is by again by Eq. (80).

Case 2: ($\|x_t\|_2 \geq R$)

By Assumption (A3), $-\frac{c_\kappa}{L}\langle x_t, \nabla_t\rangle \leq -\frac{c_\kappa}{\kappa}\|x_t\|_2^2$. Thus we can upper bound ♠ as follows:

$$\begin{aligned} ♠ =& -\frac{1}{2}\|u_t\|_2^2 - \frac{c_\kappa}{L}\langle x_t, \nabla_t\rangle + \frac{c^2}{2}\|x_t\|_2^2 \\ \leq& -\frac{1}{2}\|u_t\|_2^2 - c_\kappa^2\|x_t\|_2^2 + \frac{c^2}{2}\|x_t\|_2^2 \\ \leq& -\|u_t\|_2^2 - \frac{c_\kappa^2}{2}\|x_t\|_2^2 \\ \leq& -\frac{c_\kappa^2}{3}\Big(\|x_t\|_2^2 + \|x_t + u_t\|_2^2\Big). \end{aligned}$$

Putting the previous two results together, and using Young's inequality:

$$\begin{aligned}
\spadesuit \leq & -\frac{c_\kappa^2}{3}\Big(\|x_t\|_2^2 + \|x_t + u_t\|_2^2\Big) \\
\leq & -\frac{c_\kappa^2}{3}\Big(\|x_t\|_2^2 + \|x_t + u_t\|_2^2\Big) \\
\leq & -\frac{c_\kappa^2}{6}\Big(\|x_t\|_2^2 + \|x_t + u_t\|_2^2\Big).
\end{aligned}$$

■

F.2 Proof of Lemma 33

Let us consider the Lyapunov function $l(y_t, v_t) := \Big(\|y_t\|_2^2 + \|y_t + v_t\|_2^2 - 4R^2\Big)_+^4$.

By calculating its derivatives we can verify that

$$\begin{aligned}
\nabla_x l(y_t, v_t) &= 8l(y_t, v_t)^{3/4}(y_t) \\
\nabla_u l(y_t, v_t) &= 8l(y_t, v_t)^{3/4}(y_t + v_t) \\
\nabla_u^2 l(y_t, v_t) &= 8l(y_t, v_t)^{3/4} I + 24l(y_t, v_t)^{2/4}(y_t + v_t)(y_t + v_t)^T.
\end{aligned}$$

Recall the dynamics of the variables y_t and v_t,

$$\begin{aligned}
dy_t =& v_t dt \\
dv_t =& -2v_t - \frac{c_\kappa}{L}\nabla U(x_{y_t})dt + 2\sqrt{\frac{c_\kappa}{L}} dB_t.
\end{aligned}$$

By Itô's lemma we can study the time evolution of this Lyapunov function:

$$\begin{aligned}
dl(y_t, v_t) &= 8l(y_t, v_t)^{3/4}\Big(\langle y_t, v_t\rangle + \Big\langle y_t + v_t, -v_t - \frac{c_\kappa}{L}\nabla U(y_t)\Big\rangle\Big)dt \\
&\quad + \frac{16c_\kappa}{L}\Big(l(y_t, v_t)^{3/4}d + l(y_t, v_t)^{2/4}\|y_t + v_t\|_2^2\Big)dt \\
&\qquad + 16\sqrt{\frac{c_\kappa}{L}}l(y_t, v_t)^{3/4}(\langle y_t, v_t\rangle + \langle y_t + v_t, dB_t\rangle) \\
&\overset{(i)}{\leq} 8l(y_t, v_t)^{3/4}\Big(-\frac{c_\kappa^2}{6}\Big(\|y_t\|_2^2 + \|y_t + v_t\|_2^2\Big)\Big)dt \\
&\quad + \frac{16c_\kappa}{L}\Big(l(y_t, v_t)^{3/4}d + l(y_t, v_t)^{2/4}\|y_t + v_t\|_2^2\Big)dt \\
&\qquad + 16\sqrt{\frac{c_\kappa}{L}}l(y_t, v_t)^{3/4}(\langle y_t, v_t\rangle + \langle y_t + v_t, dB_t\rangle) \\
&\overset{(ii)}{\leq} -c_\kappa^2 l(y_t, v_t)dt \\
&\quad + \frac{32c_\kappa}{L}\Big(l(y_t, v_t)^{3/4}d\Big)dt + \frac{64c_\kappa}{L}\Big(l(y_t, v_t)^{2/4}R^2\Big)dt \\
&\qquad + 16\sqrt{\frac{c_\kappa}{L}}l(y_t, v_t)^{3/4}(\langle y_t, v_t\rangle + \langle y_t + v_t, dB_t\rangle) \\
&\leq -c_\kappa^2 l(y_t, v_t)dt \\
&\quad + \frac{c_\kappa^2}{8}l(y_t, v_t)dt + \frac{2^{25}d^4}{c_\kappa^2 L^4}dt + \frac{c_\kappa^2}{8}l(y_t, v_t)dt + \frac{2^{16}R^4}{L^2} \\
&\qquad + 16\sqrt{\frac{c_\kappa}{L}}l(y_t, v_t)^{3/4}(\langle y_t, v_t\rangle + \langle y_t + v_t, dB_t\rangle) \\
&\leq -\frac{c_\kappa^2}{2}l(y_t, v_t)dt + 2^{26}\left(\frac{d^4}{c_\kappa^2 L^4} + c^2R^8\right)dt \\
&\quad + 16\sqrt{\frac{c_\kappa}{L}}l(y_t, v_t)^{3/4}(\langle y_t, v_t\rangle + \langle y_t + v_t, dB_t\rangle),
\end{aligned}$$

where (i) can be proved by an argument similar to the proof of Lemma 35, and is omitted, while (ii) follows because

$$\|y + v\|_2^2 \leq l(y, v)^{1/4} + 4R^2, \quad \text{and,} \quad \|y\|_2^2 + \|x + u\|_2^2 \geq l(y, v)^{1/4}$$

by the definition of $l(x, u)$. Taking expectations on both sides, the term involving the Brownian motion, dB_t, goes to zero. Note also that (y_t, v_t) is distributed according to the invariant distribution p^* for all $t \geq 0$, therefore,

$$\begin{aligned}
0 = \frac{d}{dt}\mathbb{E}\left[l(y_t, v_t)\right] &\leq -\frac{c_\kappa^2}{2}\mathbb{E}\left[l(y_t, v_t)\right] + 2^{26}\left(\frac{d^4}{c_\kappa^2 L^4} + c^2R^8\right) \\
&\leq -\frac{c_\kappa^2}{2}\mathbb{E}\left[l(y_t, v_t)\right] + 2^{26}c_\kappa^2\left(10^{12}\frac{d^4}{m^4} + R^8\right)
\end{aligned}$$

Thus

$$\mathbb{E}\left[l(y_t, v_t)\right] \leq 2^{27}\left(10^{12}\frac{d^4}{m^4} + R^8\right) \leq 2^{66}\left(\frac{d}{m} + R^2\right)^4.$$

This completes the proof of the lemma.

We now state and prove some auxillary lemmas that were useful in the proof above.

Lemma 36 *Let x_t be evolved according to the dynamics in Eq.* (30)*. Then for all $t \geq 0$,*

$$\mathbb{E}\left[\|x_t\|_2^2\right] \leq 2\left(R^2 + \frac{d}{m}\right).$$

Proof Let $\theta_k \sim \mathcal{N}(0, I)$ then we have,

$$\begin{aligned}
\mathbb{1}\left\{\|x_{k\delta}\|_2 \leq R\right\} \cdot \left\|x_{(k+1)\delta}\right\|_2^2 &= \mathbb{1}\left\{\|x_{k\delta}\|_2 \leq R\right\} \cdot \left\|x_{k\delta} - \delta\nabla U(x_k) + \sqrt{2\delta}\theta_k\right\|_2^2 \\
&\leq \mathbb{1}\left\{\|x_{k\delta}\|_2 \leq R\right\} \cdot \|x_{k\delta} - \delta\nabla U(x_k)\|_2^2 \\
&\quad + \mathbb{1}\left\{\|x_{k\delta}\|_2 \leq R\right\} \cdot \left\|\sqrt{2\delta}\theta_k\right\|_2^2 \\
&\quad + \mathbb{1}\left\{\|x_{k\delta}\|_2 \leq R\right\} \cdot 2\left\langle x_{k\delta} - \delta\nabla U(x_k), \sqrt{2\delta}\theta_k\right\rangle.
\end{aligned}$$

Consider two cases:

If $\|x_{k\delta}\|_2 \geq R$, then

$$\begin{aligned}
\|x_{k\delta} - \delta\nabla U(x_k)\|_2^2 &= \|x_{k\delta}\|_2^2 - \langle x_{k\delta}, 2\delta\nabla U(x_k)\rangle + \delta^2\|\nabla U(x_k)\|_2^2 \\
&\overset{(i)}{\leq} (1 - 2\delta m)\|x_{k\delta}\|_2^2 + \delta^2\|\nabla U(x_k)\|_2^2 \\
&\overset{(ii)}{\leq} \left(1 - 2\delta m + \delta^2 L^2\right)\|x_{k\delta}\|_2^2 \\
&\overset{(iii)}{\leq} (1 - \delta m)\|x_{k\delta}\|_2^2,
\end{aligned}$$

where (i) is by Assumption (A3), (ii) is by Assumption (A1), and (iii) is by our assumption that $\delta \leq \frac{1}{\kappa L}$.

While I=if $\|x_{k\delta}\|_2 \leq R$, then

$$\begin{aligned}
\|x_{k\delta} - \delta\nabla U(x_k)\|_2^2 &= \|x_{k\delta}\|_2^2 - \langle x_{k\delta}, 2\delta\nabla U(x_k)\rangle + \delta^2\|\nabla U(x_k)\|_2^2 \\
&\overset{(i)}{\leq} \left(1 + 2\delta L + \delta^2 L^2\right)\|x_k\|_2^2 \\
&\overset{(ii)}{\leq} (1 + 3\delta L)\|x_k\|_2^2,
\end{aligned}$$

where (i) is by Assumption (A1), and (ii) is by our assumption that $\delta \leq \frac{1}{\kappa L}$.

Thus for both cases we have,

$$\begin{aligned}
\|x_{k\delta} - \delta\nabla U(x_k)\|_2^2 \leq& \mathbb{1}\left\{\|x_{k\delta}\|_2 \geq R\right\}(1 - \delta m)\|x_{k\delta}\|_2^2 + \mathbb{1}\left\{\|x_{k\delta}\|_2 \leq R\right\}(1 + 3\delta L)\|x_k\|_2^2 \\
\leq& \|x_{k\delta}\|_2^2 - \delta m\|x_{k\delta}\|_2^2 + \mathbb{1}\left\{\|x_{k\delta}\|_2 \leq R\right\} \cdot (3\delta L + \delta m)\|x_{k\delta}\|_2^2.
\end{aligned}$$

Thus we have:

$$\mathbb{1}\left\{\|x_{k\delta}\|_2 \leq R\right\}\|x_{k\delta} - \delta\nabla U(x_k)\|_2^2 \leq \mathbb{1}\left\{\|x_{k\delta}\|_2 \leq R\right\}(1 - \delta m)\|x_{k\delta}\|_2^2.$$

By taking expectations with respect to the Brownian motion we get,

$$\begin{aligned}
\mathbb{E}\left[\mathbb{1}\left\{\|x_{k\delta}\|_2 \leq R\right\}\left\|x_{(k+1)\delta}\right\|_2^2\right] &\leq (1-\delta m)\mathbb{E}\left[\mathbb{1}\left\{\|x_{k\delta}\|_2 \leq R\right\}\|x_{k\delta}\|_2^2\right] \\
&\quad + \mathbb{E}\left[\mathbb{1}\left\{\|x_{k\delta}\|_2 \leq R\right\}\left\|\sqrt{2\delta}\theta_k\right\|_2^2\right] \\
&\leq (1-\delta m)\mathbb{E}\left[\mathbb{1}\left\{\|x_{k\delta}\|_2 \leq R\right\}\|x_{k\delta}\|_2^2\right] + 2\delta d.
\end{aligned}$$

Applying this inequality recursively over k steps we arrive at,

$$\begin{aligned}
\mathbb{E}\left[\mathbb{1}\left\{\|x_{k\delta}\|_2 \leq R\right\}\left\|x_{(k+1)\delta}\right\|_2^2\right] \leq& e^{-\delta m}\mathbb{E}\left[\mathbb{1}\left\{\|x_{k\delta}\|_2 \leq R\right\}\left\|x_{(k+1)\delta}\right\|_2^2\right] + \frac{2\delta d}{\delta m} \\
\leq& \frac{2d}{m}.
\end{aligned}$$

Thus we get that,

$$\mathbb{E}\left[\left\|x_{(k+1)\delta}\right\|_2^2\right] \leq 2\left(R^2 + \frac{d}{m}\right).$$

■

Lemma 37 *Let $y \sim p^*(y) \propto e^{-U(y)}$. Then*

$$\mathbb{E}\left[\|y\|_2^8\right] \leq 2^{20}\left(\frac{d^4}{m^4} + R^8\right).$$

Proof Let $l(y) := \left(\|y\|_2^2 - R^2\right)_+^4$. We calculate derivatives and verify that

$$\begin{aligned}
\nabla l(y) &= 8l(y)^{3/4} \cdot y \\
\nabla^2 l(y) &= 48 l(y)^{2/4} \cdot yy^T + 8l(y)^{3/4} I,
\end{aligned}$$

where I is the identity matrix. By Itô's Lemma:

$$dl(y_t) = \langle \nabla l(y_t), -\nabla U(y_t)\rangle\, dt + \left\langle \nabla l(y_t), \sqrt{2} dB_t\right\rangle + \frac{1}{2}\mathrm{tr}\left(\nabla^2 l(y_t)\right). \tag{81}$$

We start by analyzing the first term,

$$\begin{aligned}
\langle \nabla l(y_t), -\nabla U(y_t)\rangle &= \left\langle 8l(y_t)^{3/4} \cdot y_t, -\nabla U(y_t)\right\rangle \\
&\overset{(i)}{\leq} \mathbb{1}\left\{\|y_t\|_2 \geq R\right\}\left(8l(y_t)^{3/4}\right)\left(-m\|y_t\|_2^2\right) \\
&\qquad + \mathbb{1}\left\{\|y_t\|_2 < R\right\}\left\langle 8l(y_t)^{3/4} y, -\nabla U(y_t)\right\rangle \\
&\overset{(ii)}{=} \left(8l(y_t)^{3/4}\right)\left(-m\|y_t\|_2^2\right) \\
&\leq -8ml(y_t),
\end{aligned}$$

where (i) is by Assumption (A3), and, (ii) is because $\mathbb{1}\left\{\|y_t\|_2 < R\right\} \cdot l(y_t) = 0$ and $\mathbb{1}\left\{\|y_t\|_2 \geq R\right\} \cdot l(y_t) = l(y_t)$ by definition of $l(y_t)$.

Consider the other term on the right-hand side of Eq. (81):

$$
\begin{aligned}
\operatorname{tr}\left(\nabla^2 l(y_t)\right) &= 48 l(y_t)^{2/4}\|y\|_2 + 8 l(y_t)^{3/4} d \\
&\overset{(i)}{\leq} 48 l(y_t)^{3/4} + 8 l(y_t)^{3/4} d + 48 l(y_t)^{2/4} R^2 \\
&\leq 64 l(y_t)^{3/4} d + 48 l(y_t)^{2/4} R^2 \\
&\overset{(ii)}{\leq} 2ml(y_t) + 2^{21}\frac{d^4}{m^3} + 2ml(y_t) + 2^{11}\frac{R^4}{m} \\
&\overset{(iii)}{\leq} 4ml(y_t) + 2^{22}\left(\frac{d^4}{m^3} + mR^8\right),
\end{aligned}
$$

where (i) is by definition of $l(y)$, while (ii) and (iii) are by Young's inequality.

Put together into Eq. (81) and taking expectations,

$$
\begin{aligned}
\frac{d}{dt}\mathbb{E}\left[l(y_t)\right] &\leq -8m\mathbb{E}\left[l(y_t)\right] + 4ml(y_t) + 2^{22}\left(\frac{d^4}{m^3} + mR^8\right) \\
&\leq -4m\mathbb{E}\left[l(y_t)\right] + 2^{22}\left(\frac{d^4}{m^3} + mR^8\right).
\end{aligned}
$$

Since $y_t \sim p^*$ for all t, $\frac{d}{dt}\mathbb{E}\left[l(y_t)\right] = 0$, thus we get,

$$
\mathbb{E}\left[l(y_t)\right] \leq 2^{20}\left(\frac{d^4}{m^4} + R^8\right).
$$

■

Appendix G. Existence of Coupling

Proof [Proof of Lemma 5] We prove the existence of a unique strong solution for $(x_t, u_t, y_t, v_t, \tau_{\lfloor \frac{t}{\nu} \rfloor})$ inductively: Let k be an arbitrary nonnegative integer, and suppose that the lemma statement holds for all $s \in [0, k\nu]$. We show that the lemma statement holds for all $s \in [0, (k+1)\nu]$.

First, we can verify that for $t \in [k\nu, (k+1)\nu)$,

$$
\tau_{\lfloor \frac{t}{\nu} \rfloor} = \tau_k,
$$

that is, $\tau_{\lfloor \frac{t}{\nu} \rfloor}$ is a constant, and so $\mu_{\lfloor \frac{t}{\nu} \rfloor} = \mu_k$ is also a constant.

Next, we find that for $t \in [k\nu, (k+1)\nu)$, the following is algebraically equivalent to dynamics described by Eqs.(11)–(14):

$$
\begin{aligned}
dx_t &= u_t dt,\\
du_t &= -2u_t dt - \frac{c_\kappa}{L}\nabla U\Big(x_{\lfloor \frac{t}{\delta}\rfloor \delta}\Big)dt + 2\sqrt{\frac{c_\kappa}{L}}dB_t,\\
dy_t &= v_t dt,\\
dv_t &= -2v_t - \frac{c_\kappa}{L}\nabla U(y_t)dt + 2\sqrt{\frac{c_\kappa}{L}}dB_t - \mu_k \cdot \left(4\sqrt{\frac{c_\kappa}{L}}\gamma_t\gamma_t^T dB_t + 2\sqrt{\frac{c_\kappa}{L}}\bar{\gamma}_t\bar{\gamma}_t^T dA_t\right),
\end{aligned}
$$

where we use the fact that μ_t takes on a constant value over $t \in [k\nu, (k+1)\nu)$.

We proceed by applying Theorem 5.2.1 of Øksendal (2013), which states that if the following holds:

1. $\mathbb{E}\left[\|x_{k\nu}\|_2^2 + \|y_{k\nu}\|_2^2 + \|u_{k\nu}\|_2^2 + \|v_{k\nu}\|_2^2\right] \leq \infty$.
2. For all $x, y \in \mathbb{R}^d$, $\|\nabla U(x) - \nabla U(y)\|_2 \leq D\|x-y\|_2$ for some constant $D > 0$.
3. For all $(x, y, u, v), (x', y', u', v')$,
$$\left\|\gamma\gamma^T - \gamma'\gamma'^T\right\|_2 + \left\|\bar{\gamma}\bar{\gamma}_t'^T - \bar{\gamma}_t'\bar{\gamma}'^T\right\|_2 \leq D\big(\left\|x-x'\right\|_2 + \left\|y-y'\right\|_2 + \left\|u-u'\right\|_2 + \left\|v-v'\right\|_2\big),$$
for some constant D (where γ and $\bar{\gamma}$ are functions of (x, y, u, v), as defined in Eq. (15), similarly for γ', $\bar{\gamma}'$ and (x', y', u', v')),

then there is a solution (x_t, y_t, u_t, v_t) for $t \in [k\nu, (k+1)\nu]$ with the properties:

(a) (x_t, y_t, u_t, v_t) is unique and t-continuous with probability one.

(b) (x_t, y_t, u_t, v_t) is adapted to the filtration $\mathcal{F}_t$ generated by $(x_{k\nu}, y_{k\nu}, u_{k\nu}, v_{k\nu})$ and dB_t and dA_t for $t \in [k\nu, (k+1)\nu)$.

(c) $\int_0^T \mathbb{E}\left[\|x_t\|_2^2 + \|y_t\|_2^2 + \|u_t\|_2^2 + \|v_t\|_2^2\right] dt < \infty$.

We can verify the first condition holds by using Lemma 32 and Lemma 33. Condition 2 holds due to our smoothness assumption, Assumption (A1).

We can verify that Condition 3 also holds using the argument below:

From the definition of $\mathcal{M}$ in Eq. (15), we know that $|\mathcal{M}(r)'| \leq \frac{1}{2}|\sin(r\cdot 2\pi/\beta)| \cdot \frac{2\pi}{\beta} \leq \frac{\pi}{\beta}$.

By definition of γ_t in Eq. (15),

$$\gamma_t\gamma_t^T := \mathcal{M}(\|z_t + w_t\|_2) \cdot \frac{(z_t+w_t)(z_t+w_t)^T}{\|z_t+w_t\|_2^2}.$$

To simplify notation, consider an arbitrary $x, y \in \mathbb{R}^d$, and assume wlog that $\|x\|_2 \leq \|y\|_2$. We will bound

$$\left\|\mathcal{M}(\|x\|_2)\frac{xx^T}{\|x\|_2^2} - \mathcal{M}(\|y\|_2)\frac{yy^T}{\|y\|_2^2}\right\|_2 \leq D\|x-y\|_2,$$

for some D, which implies condition 3.

By the triangle inequality,

$$\begin{aligned}
&\left\|\mathcal{M}(\|x\|_2)\frac{xx^T}{\|x\|_2^2} - \mathcal{M}(\|y\|_2)\frac{yy^T}{\|y\|_2^2}\right\|_2 \\
&\leq \mathcal{M}(\|x\|_2)\left\|\frac{xx^T}{\|x\|_2^2} - \frac{yy^T}{\|y\|_2^2}\right\|_2 + \left\|\frac{yy^T}{\|y\|_2^2}\right\|_2 |\mathcal{M}(\|x\|_2) - \mathcal{M}(\|y\|_2)| \\
&\leq \mathcal{M}(\|x\|_2)\left\|\frac{xx^T}{\|x\|_2^2} - \frac{yy^T}{\|y\|_2^2}\right\|_2 + |\mathcal{M}(\|x\|_2) - \mathcal{M}(\|y\|_2)|. \qquad (82)
\end{aligned}$$

The second term can be bounded as

$$|\mathcal{M}(\|x\|_2) - \mathcal{M}(\|y\|_2)| \leq \frac{\pi}{\beta}|\|x\|_2 - \|y\|_2|_2 \leq \frac{\pi}{\beta}\|x-y\|_2,$$

where we use the upper bound we established on $|\mathcal{M}'(r)|$.

To bound the first term, we consider two cases:

If $\|x\|_2 \leq \beta/2$, $\mathcal{M}(\|x\|_2) = 0$ and we are done.

If $\|x\|_2 \geq \beta/2$, we verify that the transformation $T(x) = \frac{x}{\|x\|_2}$ has Jacobian $\nabla T(x) = \frac{1}{\|x\|_2}\left(I - \frac{xx^T}{\|x\|_2}\right)$, so that $\|\nabla T(x)\|_2 \leq \frac{1}{\|x\|_2}$. By our earlier assumption that $\|x\|_2 \leq \|y\|_2$, we know that $\|ax + (1-a)y\|_2 \geq \beta/2$ for all $a \in [0,1]$. Therefore,

$$\left\|\frac{x}{\|x\|_2} - \frac{y}{\|y\|_2}\right\|_2 = \|T(x) - T(y)\|_2 \leq \frac{1}{\|x\|_2}\|x-y\|_2 \leq \frac{2}{\beta}.$$

By the triangle inequality and some algebra, we obtain:

$$\begin{aligned}
&\left\|\frac{xx^T}{\|x\|_2^2} - \frac{yy^T}{\|y\|_2^2}\right\|_2 \\
&\leq \left\|\frac{x}{\|x\|_2} + \frac{y}{\|y\|_2}\right\|_2\left\|\frac{x}{\|x\|_2} - \frac{y}{\|y\|_2}\right\|_2 \\
&\leq 2\left\|\frac{x}{\|x\|_2} - \frac{y}{\|y\|_2}\right\|_2 \\
&\leq \frac{4}{\beta}\|x-y\|_2,
\end{aligned}$$

where the first two inequalities are due to the triangle inequality. Combined with the fact that $\mathcal{M}(r) \leq 1$ for all r, we can bound Eq. (82) by $\frac{8}{\beta}\|x-y\|_2$.

A similar argument can be used to show that $\bar{\gamma}_t$ is Lipschitz. Let $\mathcal{N}(x) := \left(1 - (1 - 2\mathcal{M}(\|x\|_2))^2\right)^{1/2}$. Then we verify that

$$\mathcal{N}(r) := \begin{cases} 1, & \text{for } r \in [\beta, \infty) \\ \sin\left(r \cdot \frac{2\pi}{\beta}\right), & \text{for } r \in [\beta/2, \beta] \\ 0, & \text{for } r \in [0, \beta/2] \end{cases}$$

$$\bar{\gamma}_t := (\mathcal{N}(\|z_t + w_t\|_2))^{1/2} \frac{z_t + w_t}{\|z_t + w_t\|_2}.$$

The proof is almost identical to the proof of (82), so we omit it, but highlight two crucial facts:

$$1.\ \mathcal{N}(r) \in [0, 1] \qquad \text{for all } r$$
$$2.\ \left|\mathcal{N}'(r)\right| \leq \frac{2\pi}{\beta} \qquad \text{for all } r.$$

Thus we find that Condition 3 is satisfied with $D = \frac{16}{\beta}$, and in turn show that (a)-(c) hold for $t \in [k\nu, (k+1)\nu]$. From Eq. (17) we know that $\tau_{(k+1)\nu}$ is a function of $(x_{(k+1)\nu}, u_{(k+1)\nu}, y_{(k+1)\nu}, v_{(k+1)\nu}, \tau_k)$. Thus we have shown the existence of a unique solution $(x_t, y_t, u_t, v_t, \tau_{\lfloor \frac{t}{\nu} \rfloor})$ for $t \in [k\nu, (k+1)\nu]$, where (x_t, y_t, u_t, v_t) is t-continuous.

The proof of the lemma now follows by induction over k. ■

Lemma 38 *Let B_t and A_t be two independent Brownian motions, and let $\mathcal{F}_t$ be the σ-algebra generated by B_s, A_s; $s \leq t$, and (x_0, u_0, y_0, v_0).*

For all $t \geq 0$, the stochastic process ϕ_t defined in Eqs. (24) has a unique solution such that ϕ_t is t-continuous with probability one, and satisfies the following, for all $s \geq 0$:

1. *ϕ_t is adapted to the filtration $\mathcal{F}_s$.*
2. *$\mathbb{E}\left[\|\phi_t\|_2^2\right] \leq \infty$.*

Proof The proof is almost identical to that of Lemma 5. The main additional requirement is showing that there exists a constant D such that for any (x, y, u, v) and (x', y', u', v'),

$$\left\|\nabla_w f(r)\gamma\gamma^T - \nabla_{w'} f(r')\gamma'\gamma'^T\right\|_2, \tag{83}$$

with γ (resp γ') being a function of (x, y, u, v) (resp γ') as defined in (15). and r being a function of (x, y, u, v) as defined in (18). In the proof of Lemma 5, we already showed that $\gamma\gamma^T$ and $\gamma'\gamma'^T$ are uniformly bounded and lipschitz, thus it is sufficient to show that

$$\begin{aligned} &1.\ \left\|\nabla_w f(r) - \nabla_{w'} f(r')\right\|_2 \leq D\left\|x - x'\right\|_2 + \left\|y - y'\right\|_2 + \left\|u - u'\right\|_2 + \left\|v - v'\right\|_2 \\ &2.\ \|\nabla_w f(r)\|_2 \leq D. \end{aligned} \tag{84}$$

The second point is easy to verify:

$$\nabla_w f(r) = f'(r)\nabla_w r = f'(r)(\nabla_w \ell(z + w))$$

Thus $\|\nabla_w f(r)\|_2 \leq 1$ using item (F2) of Lemma 31 and item 2 of Lemma 6.

To prove the first point, we verify that

$$\nabla_w^2 f(r) = f''(r)(\nabla_w \ell(z+w)) + f'(r)\big(\nabla_w^2 \ell(z+w)\big).$$

Using item (F7) of Lemma 31 and item

$$\left\|\nabla_w^2 f(r)\right\|_2 \leq 4\alpha_f \mathcal{R}_f + \frac{4}{\mathcal{R}_f} + \frac{8}{\beta};$$

this implies 84 which in turn implies (83). Note that $\|w-w'\|_2 \leq \|u-u'\|_2 + \|v-v'\|$.

■

Appendix H. Coupling and Discretization

Proof [Proof of Lemma 4] Let us define

$$\bar{B}_t := \int_0^t dB_t - \mathbb{1}\left\{k\nu \geq \tau_{\lfloor \frac{t}{\nu} \rfloor} + T_{sync}\right\} \cdot \big(2\gamma_t\gamma_t^T dB_t + \bar{\gamma}_t\bar{\gamma}_t^T dA_t\big).$$

We will show that $\bar{B}_t$ is a Brownian motion by using Levy's characterization. The conclusion then follows immediately from the dynamics defined in Eq. (5).

Since B_t and A_t are Brownian motions, $\bar{B}_t$ is also a continuous martingale with respect to the filtration $\mathcal{F}_t$. Further the quadratic variation of $\bar{B}_t$ over an interval $[s,s']$ is

$$\int_s^{s'} \underbrace{\left(I - 2\mathbb{1}\left\{k\nu \geq \tau_{\lfloor \frac{t}{\nu} \rfloor} + T_{sync}\right\}\gamma_t\gamma_t^T\right)^2 + \mathbb{1}\left\{k\nu \geq \tau_{\lfloor \frac{t}{\nu} \rfloor} + T_{sync}\right\}\left(\bar{\gamma}_t\bar{\gamma}_t^T\right)^2}_{=:\spadesuit} dt.$$

If $\mathbb{1}\left\{k\nu \geq \tau_{\lfloor \frac{t}{\nu} \rfloor} + T_{sync}\right\} = 0$, then the above is clearly the identity matrix – I.

If, on the other hand, $\mathbb{1}\left\{k\nu \geq \tau_{\lfloor \frac{t}{\nu} \rfloor} + T_{sync}\right\} = 1$, define $c_t := z_t + w_t$; then by the definition of γ_t and $\bar{\gamma}_t$ in Eq. (15), we find that

$$\begin{aligned}
\spadesuit &= \left(I - 2\mathcal{M}(\|c_t\|_2)\frac{c_t c_t^T}{\|c_t\|_2^2}\right)^2 + \left(1 - (1-2\mathcal{M}(\|c_t\|_2))^2\right)\frac{c_t c_t^T}{\|c_t\|_2^2} \\
&\overset{(i)}{=} I - \frac{c_t c_t^T}{\|c_t\|_2^2} + (1-2\mathcal{M}(\|c_t\|_2))^2\frac{c_t c_t^T}{\|c_t\|_2^2} + \left(1 - (1-2\mathcal{M}(\|c_t\|_2))^2\right)\frac{c_t c_t^T}{\|c_t\|_2^2} \\
&= I,
\end{aligned}$$

where (i) follows by the eigenvalue decomposition of the matrix $\left(I - 2\mathcal{M}(\|c_t\|_2)\frac{c_t c_t^T}{\|c_t\|_2^2}\right)^2$.

Thus the quadratic variation of $\bar{B}_t$ over the interval $[s,s]$ is $(s'-s)I$, thus satisfying Levy's characterization of a Brownian motion.

■

Proof [Proof of Lemma 8] Using similar steps as Lemma 4, we can verify that

$$\bar{B}_t := \int_0^t \sqrt{2} dB_t - 2\sqrt{2}\gamma_t \gamma_t^T dB_t + \sqrt{2}\bar{\gamma}_t \bar{\gamma}_t^T dA_t.$$

is a Brownian motion. The proof follows immediately. ■

Lemma 39 *Given $(x_{k\delta}, u_{k\delta})$, the solution (x_t, u_t), for $t \in (k\delta, (k+1)\delta]$, of the discrete underdamped Langevin diffusion defined by the dynamics in Eq. (7) is*

$$u_t = u_{k\delta} e^{-2(t-k\delta)} - \frac{c_\kappa}{L}\left(\int_{k\delta}^t e^{-2(t-s)} \nabla U(x_{k\delta}) ds\right) + \sqrt{\frac{4c_\kappa}{L}} \int_{k\delta}^t e^{-2(t-s)} dB_s \tag{85}$$

$$x_t = x_{k\delta} + \int_{k\delta}^t u_s ds.$$

Proof It can be easily verified that the above expressions have the correct initial values $(x_{k\delta}, u_{k\delta})$. By taking derivatives, one can also verify that they satisfy the stochastic differential equations in Eq. (7). ■

Lemma 40 *Conditioned on $(x_{k\delta}, u_{k\delta})$, the solution $(x_{(k+1)\delta}, u_{(k+1)\delta})$ of Eq. (7) is a Gaussian with mean,*

$$\mathbb{E}\left[u_{(k+1)\delta}\right] = u_{k\delta} e^{-2\delta} - \frac{c_\kappa}{2L}(1 - e^{-2\delta}) \nabla f(x_{k\delta})$$

$$\mathbb{E}\left[x_{(k+1)\delta}\right] = x_{k\delta} + \frac{1}{2}(1 - e^{-2\delta}) u_{k\delta} - \frac{c_\kappa}{2L}\left(\delta - \frac{1}{2}\left(1 - e^{-2\delta}\right)\right) \nabla U(x_{k\delta}),$$

and covariance,

$$\mathbb{E}\left[\left(x_{(k+1)\delta} - \mathbb{E}\left[x_{(k+1)\delta}\right]\right)\left(x_{(k+1)\delta} - \mathbb{E}\left[x_{(k+1)\delta}\right]\right)^\top\right] = \frac{c_\kappa}{L}\left[\delta - \frac{1}{4}e^{-4\delta} - \frac{3}{4} + e^{-2\delta}\right] \cdot I_{d\times d}$$

$$\mathbb{E}\left[\left(u_{(k+1)\delta} - \mathbb{E}\left[u_{(k+1)\delta}\right]\right)\left(u_{(k+1)\delta} - \mathbb{E}\left[u_{(k+1)\delta}\right]\right)^\top\right] = \frac{c_\kappa}{L}(1 - e^{-4\delta}) \cdot I_{d\times d}$$

$$\mathbb{E}\left[\left(x_{(k+1)\delta} - \mathbb{E}\left[x_{(k+1)\delta}\right]\right)\left(u_{(k+1)\delta} - \mathbb{E}\left[u_{(k+1)\delta}\right]\right)^\top\right] = \frac{c_\kappa}{2L}\left[1 + e^{-4\delta} - 2e^{-2\delta}\right] \cdot I_{d\times d}.$$

Proof Consider some $t \in [k\delta, (k+1)\delta)$.

It follows from the definition of Brownian motion that the distribution of (x_t, u_t) is a $2d$-dimensional Gaussian distribution. We will compute its moments below, using the expression in Lemma 39. Computation of the conditional means is straightforward, as we can simply ignore the zero-mean Brownian motion terms:

$$\mathbb{E}\left[u_t\right] = u_{k\delta} e^{-2(t-k\delta)} - \frac{c_\kappa}{2L}(1 - e^{-2(t-k\delta)}) \nabla U(x_{k\delta}) \tag{86}$$

$$\mathbb{E}\left[x_t\right] = x_{k\delta} + \frac{1}{2}(1 - e^{-2(t-k\delta)}) u_{k\delta} - \frac{c_\kappa}{2L}\left(t - k\delta - \frac{1}{2}\left(1 - e^{-2(t-k\delta)}\right)\right) \nabla U(x_{k\delta}). \tag{87}$$

The conditional variance for u_t only involves the Brownian motion term:

$$\begin{aligned}\mathbb{E}\left[\left(u_t - \mathbb{E}\left[u_t\right]\right)\left(u_t - \mathbb{E}\left[u_t\right]\right)^\top\right] &= \frac{4c_\kappa}{L}\mathbb{E}\left[\left(\int_{k\delta}^t e^{-2(t-s)}dB_s\right)\left(\int_{k\delta}^t e^{-2(s-t)}dB_s\right)^\top\right] \\ &= \frac{4c_\kappa}{L}\left(\int_{k\delta}^t e^{-4(t-s)}ds\right)\cdot I_{d\times d} \\ &= \frac{c_\kappa}{L}(1-e^{-4(t-k\delta)})\cdot I_{d\times d}.\end{aligned}$$

The Brownian motion term for x_t is given by

$$\begin{aligned}\sqrt{\frac{4c_\kappa}{L}}\int_{k\delta}^t\left(\int_{k\delta}^r e^{-2(r-s)}dB_s\right)dr &= \sqrt{\frac{4c_\kappa}{L}}\int_{k\delta}^t e^{2s}\left(\int_s^t e^{-2r}dr\right)dB_s \\ &= \sqrt{\frac{c_\kappa}{L}}\int_{k\delta}^t\left(1-e^{-2(t-s)}\right)dB_s.\end{aligned}$$

Here the second equality follows by Fubini's theorem. The conditional covariance for x_t now follows as

$$\begin{aligned}\mathbb{E}\left[\left(x_t - \mathbb{E}\left[x_t\right]\right)\left(x_t - \mathbb{E}\left[x_t\right]\right)^\top\right] &= \frac{c_\kappa}{L}\mathbb{E}\left[\left(\int_{k\delta}^t\left(1-e^{-2(t-s)}\right)dB_s\right)\left(\int_{k\delta}^t\left(1-e^{-2(t-s)}\right)dB_s\right)^\top\right] \\ &= \frac{c_\kappa}{L}\left[\int_{k\delta}^t\left(1-e^{-2(t-s)}\right)^2 ds\right]\cdot I_{d\times d} \\ &= \frac{c_\kappa}{L}\left[t-k\delta-\frac{1}{4}e^{-4(t-k\delta)}-\frac{3}{4}+e^{-2(t-k\delta)}\right]\cdot I_{d\times d}.\end{aligned}$$

Finally we compute the cross-covariance between x_t and u_t,

$$\begin{aligned}\mathbb{E}\left[\left(x_t - \mathbb{E}\left[x_t\right]\right)\left(u_t - \mathbb{E}\left[u_t\right]\right)^\top\right] &= \frac{2c_\kappa}{L}\mathbb{E}\left[\left(\int_{k\delta}^t\left(1-e^{-2(t-s)}\right)dB_s\right)\left(\int_{k\delta}^t e^{-2(t-s)}dB_s\right)^\top\right] \\ &= \frac{2c_\kappa}{L}\left[\int_{k\delta}^t(1-e^{-2(t-s)})(e^{-2(t-s)})ds\right]\cdot I_{d\times d} \\ &= \frac{c_\kappa}{2L}\left[1+e^{-4(t-k\delta)}-2e^{-2(t-k\delta)}\right]\cdot I_{d\times d}.\end{aligned}$$

We thus have an explicitly defined Gaussian. Notice that we can sample from this distribution in time linear in d, since all d coordinates are independent. ■