On Explicit L^2-Convergence Rate Estimate for Underdamped Langevin Dynamics

Yu Cao, Jianfeng Lu & Lihan Wang

Communicated by C. Le Bris

Abstract

We provide a refined explicit estimate of the exponential decay rate of underdamped Langevin dynamics in the L^2 distance, based on a framework developed in Albritton et al. (Variational methods for the kinetic Fokker–Planck equation, arXiv arXiv:1902.04037, 2019). To achieve this, we first prove a Poincaré-type inequality with a Gibbs measure in space and a Gaussian measure in momentum. Our estimate provides a more explicit and simpler expression of the decay rate; moreover, when the potential is convex with a Poincaré constant $m \ll 1$, our estimate shows the decay rate of $O(\sqrt{m})$ after optimizing the choice of the friction coefficient, which is much faster than m for the overdamped Langevin dynamics.

1. Introduction

We consider the convergence rate for the following *underdamped Langevin dynamics* $(x_t, v_t) \in \mathbb{R}^d \times \mathbb{R}^d$:

$$
\begin{cases}
\mathrm{d}x_t = v_t \, \mathrm{d}t \\
\mathrm{d}v_t = -\nabla U(x_t) \, \mathrm{d}t - \gamma v_t \, \mathrm{d}t + \sqrt{2\gamma} \, \mathrm{d}W_t.
\end{cases}
\tag{1}
$$

Have $U(x)$ is the potential energy, $\gamma > 0$ is the friction coefficient, and W_t is a d-dimensional standard Brownian motion; the mass and temperature are set to be 1 for simplicity. The law of the process (1), $\rho(t, x, v)$, satisfies the kinetic Fokker–Planck equation

$$
\partial_t \rho = -v \cdot \nabla_x \rho + \nabla_x U \cdot \nabla_v \rho + \gamma \left(\Delta_v \rho + \nabla_v \cdot (v\rho) \right).
\tag{2}
$$

It is well-known (see for example [45, Proposition 6.1]) that under mild assumptions, (2) admits a unique stationary density function given by

$$
\mathrm{d}\rho_\infty(x, v) = \mathrm{d}\mu(x) \, \mathrm{d}\kappa(v),
\tag{3}
$$

where

$$d\mu(x) = \frac{1}{Z_U} e^{-U(x)} dx, \quad d\kappa(v) = \frac{1}{(2\pi)^{d/2}} e^{-\frac{|v|^2}{2}} dv, \quad Z_U = \int_{\mathbb{R}^d} e^{-U(x)} dx.$$

When $\gamma \to \infty$, the rescaled dynamics $x_t^{(\gamma)} := x_{\gamma t}$ converges to the *Smoluchowski SDE*, also known as the *overdamped Langevin dynamics* (see e.g., [45, Sec. 6.5]), which is given by

$$dx_t^{(\infty)} = -\nabla U(x_t^{(\infty)}) dt + \sqrt{2} dB_t.$$

An equivalent formalism of (2) is the following *backward Kolmogorov equation*:

$$\partial_t f = \mathcal{L} f, \qquad \mathcal{L} = \mathcal{L}_{\text{ham}} + \gamma \mathcal{L}_{\text{FD}}, \qquad f(0, x, v) = f_0(x, v). \tag{4}$$

Have $\mathcal{L}_{\text{ham}}$ is the Hamiltonian transport operator and $\mathcal{L}_{\text{FD}}$ is the fluctuation-dissipation term

$$\begin{cases} \mathcal{L}_{\text{ham}} = v \cdot \nabla_x - \nabla_x U \cdot \nabla_v \\ \mathcal{L}_{\text{FD}} = \Delta_v - v \cdot \nabla_v. \end{cases} \tag{5}$$

Indeed, (4) could be derived from (2) by considering $\rho(t, x, v) = f(t, x, -v)\rho_\infty$ (x, v) [45]; since by L^2-duality, $\|\rho - \rho_\infty\|_{L^2(\rho_\infty^{-1})} \equiv \|f - \int f \, d\rho_\infty\|_{L^2(\rho_\infty)}$, the exponential convergence of the solution $\rho(t, \cdot, \cdot)$ of (2) to ρ_∞ is equivalent to the exponential decay of $f(t, \cdot, \cdot)$ to zero, provided that $\int f_0 \, d\rho_\infty = 0$. Similarly, one could obtain the backward Kolmogorov equation for the overdamped Langevin dynamics, which is given by

$$\partial_t h = -\nabla_x U \cdot \nabla_x h + \Delta_x h, \qquad h(0, x) = h_0(x). \tag{6}$$

If μ satisfies a Poincaré inequality, one could show that the generator in the above equation (6) is self-adjoint and coercive with respect to $L^2(\mu)$. As a consequence, if $\int h_0 \, d\mu = 0$, then $h(t, x)$ decays to zero exponentially fast as $t \to \infty$, see for example [6, Theorem 4.2.5].

Unlike the generator of (6), the generator $\mathcal{L}$ in (4) for the underdamped Langevin is not uniformly elliptic. As a result, proving the exponential convergence of $\rho(t, \cdot, \cdot)$ to the equilibrium ρ_∞ is more challenging. With extensive works throughout the years, the exponential convergence of the underdamped Langevin dynamics is now better understood in various norms (see Section 1.2 below for a review).

Our goal in this work is to provide an explicit estimate of the decay rate in L^2 for the semigroup in (4), based on a framework proposed in [1] which implicitly uses Hörmander's bracket conditions [32]. In particular, under some mild assumptions of U, we obtain explicit estimates for some universal constant $C > 1$ independent of U, γ, d and some $\nu > 0$ such that for any possible $f = f(t, x, v)$ satisfying (4) and $\int f_0 \, d\rho_\infty = 0$, we have

$$\|f(t, \cdot, \cdot)\|_{L^2(\rho_\infty)} \leqslant C e^{-\nu t} \|f_0\|_{L^2(\rho_\infty)}. \tag{7}$$

In the rest of this section, we will first present our assumptions and main results. Next, we will briefly review existing approaches to study the exponential convergence of (4) (or equivalently (2)) in Section 1.2, and compare our estimate of the decay rate ν with some previous works aiming at explicit estimates [9,16,40,47]. We would like to comment here that convergence results are also obtained in earlier works [17,26], although their rates are only explicit in γ.

Notations

Throughout the paper we assume I to be the time interval $(0, T)$, and we use $d\lambda(t) = \frac{1}{T}\chi_{(0,T)}(t)\,dt$ to denote the rescaled Lebesgue measure on I so that $d\lambda(t)$ denotes a probability measure. For any probability measure ρ, we use $L^2(\rho)$ (and similarly $H^1(\rho)$, $H^2(\rho)$) to denote the standard Sobolev spaces, and $H^{-1}(\rho)$ to denote the dual space of $H^1(\rho)$. For the Gaussian probability measure κ in velocity space, we also use L^2_κ, H^1_κ, H^{-1}_κ to denote the corresponding spaces. Moreover, we use $H^1_0(\lambda \otimes \mu)$ to denote the $H^1(\lambda \otimes \mu)$ functions that vanish at both time boundaries $t = 0$ and $t = T$. By abuse of notation, we denote the canonical pairing $\langle \cdot, \cdot \rangle_{H^{-1}(\rho), H^1(\rho)}$ between $f \in H^1(\rho)$ and $g \in H^{-1}(\rho)$ by

$$\int fg\,d\rho := \langle g, f \rangle_{H^{-1}(\rho), H^1(\rho)}.$$

For $f \in H^{-1}(\rho)$, we use the notation $(f)_\rho := \langle f, 1 \rangle_{H^{-1}(\rho), H^1(\rho)}$. For an arbitrary Banach space V and time interval I equipped with Lebesgue measure $d\lambda(t)$, we denote by $L^p(\lambda \otimes \mu; V)$ the Banach space of functions $f(t, x, v)$ with norm

$$\|f\|_{L^p(\lambda \otimes \mu; V)} := \left(\int_{I \times \mathbb{R}^d} \|f(t, x, \cdot)\|^p_V\,d\lambda(t)\,d\mu(x) \right)^{\frac{1}{p}}.$$

Inspired by [1], we define the Banach space

$$H^1_{hyp}(\lambda \otimes \mu) := \left\{ f \in L^2(\lambda \otimes \mu; H^1_\kappa) : \partial_t f - \mathcal{L}_{ham} f \in L^2(\lambda \otimes \mu; H^{-1}_\kappa) \right\}.$$

We define a projection operator for $\phi(t, x, v) \in L^2(\lambda \otimes \rho_\infty)$ by

$$(\Pi_v \phi)(t, x) := \int_{\mathbb{R}^d} \phi(t, x, v)\,d\kappa(v). \tag{8}$$

Equivalently, Π_v is used to obtain the marginal component of ϕ in $L^2(\lambda \otimes \mu)$. By a slight abuse of notation, for $\phi(x, v) \in L^2(\rho_\infty)$, we also use the same notation Π_v to represent the similar projection, i.e., $(\Pi_v \phi)(x) := \int_{\mathbb{R}^d} \phi(x, v)\,d\kappa(v)$. The adjoints of ∇_x and ∇_v in the Hilbert space $L^2(\rho_\infty)$ are respectively given by $\nabla_x^* F = -\nabla_x \cdot F + \nabla_x U \cdot F$ and $\nabla_v^* F = -\nabla_v \cdot F + v \cdot F$ for any vector field $F(x, v) : \mathbb{R}^{2d} \to \mathbb{R}^d$. Thus we can rewrite operators $\mathcal{L}_{ham}$ and $\mathcal{L}_{FD}$ as

$$\mathcal{L}_{ham} = \nabla_v^* \nabla_x - \nabla_x^* \nabla_v, \qquad \mathcal{L}_{FD} = -\nabla_v^* \nabla_v. \tag{9}$$

For time-augmented state space $I \times \mathbb{R}^d$ equipped with measure $\lambda \otimes \mu$, we use the convention $\partial_{x_0} := \partial_t$, the short-hand notation $\overline{\nabla} := (\partial_t, \nabla_x)^\top$, and the notation $\mathscr{L} := -\partial_{tt} + \nabla_x^* \nabla_x$ to denote the "Laplace" operator on $L^2(\lambda \otimes \mu)$. We use C to denote a universal constant independent of all parameters that may change from line to line.

1.1. Assumptions and Main Results

Assumption 1. *(Poincaré inequality for μ)* Assume that the potential $U(x)$ satisfies a Poincaré inequality in space

$$\int_{\mathbb{R}^d} \left(f - \int_{\mathbb{R}^d} f \, d\mu \right)^2 d\mu \leqslant \frac{1}{m} \int_{\mathbb{R}^d} |\nabla_x f|^2 \, d\mu, \qquad \forall f \in H^1(\mu). \tag{10}$$

Assumption 2. The potential $U \in C^2(\mathbb{R}^d)$, and there exist constants $M > 0$ and $\delta \in (0, 1)$ such that

$$|\nabla_x^2 U(x)|^2 = \sum_{i,j=1}^d |\partial_{x_i x_j} U(x)|^2 \leqslant M^2(d + |\nabla_x U(x)|^2),$$

$$\text{and } \Delta_x U(x) \leqslant Md + \frac{\delta}{2} |\nabla_x U(x)|^2 \qquad \forall \, x \in \mathbb{R}^d. \tag{11}$$

for some constant $M \geqslant 1$.

Assumption 3. The embedding $H^1(\mu) \hookrightarrow L^2(\mu)$ is compact.

Remark 1.1. (i) Assumption 1 guarantees that the elliptic equation $\nabla_x^* \nabla_x u = h$ has a unique solution $u \in H^2(\mu)$ for any $h \in L^2(\mu)$ satisfying $(h)_\mu = 0$ (see for example [19, Proposition 5]). Hence, together with Assumption 3, we derive from Fredholm alternative that $L^2(\mu)$ has an orthonormal basis $\{1\} \cup \{w_\alpha\}_{\alpha > 0}$ where $w_\alpha \in H^2(\mu)$ are eigenfunctions of $\nabla_x^* \nabla_x$ with eigenvalue α^2 for a discrete set of $\alpha > 0$ (see [22, Chapter 6] for an argument with bounded domains):

$$\nabla_x^* \nabla_x w_\alpha = \alpha^2 w_\alpha.$$

Further, by Assumption 1, any eigenvalue α^2 of $\nabla_x^* \nabla_x$ satisfies $\alpha \geqslant \sqrt{m}$, in fact, the smallest α is precisely $\sqrt{m}$, the square root of the Poincaré constant; the spectrum of $\nabla_x^* \nabla_x$ is unbounded from above.

(ii) Assumption 3 is satisfied when

$$\lim_{|x| \to \infty} \frac{U(x)}{|x|^\beta} = \infty$$

for some $\beta > 1$ (see [31] for a proof). We would like to comment here that we require Assumption 3 only for technical purposes, more precisely in the proof of Lemma 2.6 where we used the spectral decomposition of the elliptic operator $\nabla_x^* \nabla_x$ to construct the test functions we desire. We believe that the assumption is not necessary for our main results to hold. We leave this for future research.

(iii) Similar versions of Assumption 2 is commonly used in the literature, see e.g., the books [45,54] and the papers [18,19], and is satisfied when U grows at most exponentially fast as $x \to \infty$. Here we adopt the more natural dimension scaling in [10, Assumption 1] (in particular, we take $c_1 = c_3 = M$ in their setting), since in the case of separable potential $U(x) = \sum_{i=1}^{d} u(x_i)$, this amounts to the more natural one-dimensional estimate $|u''|^2 \leqslant M(1 + |u'|^2)$.

Theorem 1. *Under Assumptions 1, 2, and 3, there exist a constant $v > 0$ and universal constants C, c independent of all parameters such that, for every $f(t, x, v)$ satisfying the backward Kolmogorov equation (4) with initial condition $f_0 \in L^2(\mu; H_\kappa^1)$ and*

$$(f_0)_{\rho_\infty} = 0, \tag{12}$$

we have, for every $t \in (0, \infty)$,

$$\|f(t, \cdot)\|_{L^2(\rho_\infty)} \leqslant C \exp(-vt) \|f_0\|_{L^2(\rho_\infty)}.$$

Moreover, v can be made explicit as

$$v = \frac{m\gamma}{c(\sqrt{m} + R + \gamma)^2} \tag{13}$$

with some constant $R > 0$ given by

(i) If U is convex, then

$$R = 0.$$

(ii) If the Hessian of U is bounded from below

$$\nabla_x^2 U(x) \geqslant -K \operatorname{Id}, \qquad \forall x \in \mathbb{R}^d \tag{14}$$

for some constant $K \geqslant 0$, then

$$R = \sqrt{K}.$$

Note that if $K = 0$, we recover the estimate in case (i).
(iii) In the most general case without further assumptions,

$$R = M + M^{\frac{3}{4}} d^{\frac{1}{4}}.$$

Remark 1.2. (i) If we fix $m = O(1)$, then, when $\gamma \to 0$ (resp. $\gamma \to \infty$), our estimate provides an estimate on decay rate of $O(\gamma)$ (resp. $O(\gamma^{-1})$). This is consistent with [17,26,47] and also the isotropic Gaussian case when $U(x) = \frac{m}{2}|x|^2$ (see Appendix A).

(ii) In the convex case, if we optimize with respect to γ by choosing $\gamma = \sqrt{m}$, then

$$v = \frac{\sqrt{m}}{4c}.$$

As is shown in Appendix A, the scaling on m is optimal in the regime $m \to 0$, as it is the rate even for isotropic quadratic potential. We refer the readers to Appendix B for the corresponding results from the DMS method, with a slightly more explicit estimate compared to [47].

(iii) In the case where condition (14) is satisfied, e.g. for the double well potential $U(x) = (|x|^2 - 1)^2$ with $K = 4$, our scaling on K is consistent with [36, Theorem 1] and [37, Sec. 5]. Similar assumption is also used in [44, Theorem 1] for functional inequalities.

(iv) It is well-known that for overdamped Langevin dynamics, the decay rate is simply m in $L^2(\mu)$ for (6). By part (ii) of this remark, when $m \ll 1$, the underdamped Langevin dynamics (1) could converge to its equilibrium ρ_∞ at a rate $O(\sqrt{m})$ for convex potentials, which is much faster than the overdamped Langevin dynamics.

(v) Due to the relation (see e.g., [48])

$$\frac{1}{\sqrt{2}} \|\rho - \rho_\infty\|_{\mathrm{TV}} \leqslant \sqrt{\mathrm{KL}(\rho \,\|\, \rho_\infty)} \leqslant \sqrt{\chi^2(\rho, \rho_\infty)}$$

$$\equiv \|\rho - \rho_\infty\|_{L^2(\rho_\infty^{-1})} \equiv \left\| f - \int f \, d\rho_\infty \right\|_{L^2(\rho_\infty)},$$

where $f = d\rho/d\rho_\infty$, and the Talagrand inequality [44] $W_2(\rho, \rho_\infty) \leqslant \sqrt{\frac{2}{C_{LSI}} \mathrm{KL}(\rho \| \rho_\infty)}$ where C_{LSI} is the logarithmic Sobolev constant, Theorem 1 implies that $\rho(t, \cdot, \cdot)$ converges to ρ_∞ with rate 2ν in both χ^2-divergence and relative entropy, and with rate ν in total variation and (if μ satisfies log-Sobolev inequality) 2-Wasserstein distance. On the other hand, our result does not imply

$$d(\rho_t, \rho_\infty) \leqslant C \exp(-\nu t) d(\rho_0, \rho_\infty)$$

where $d(\rho, \pi) = TV(\rho, \pi)$, $W_2(\rho, \pi)$ or $\mathrm{KL}(\rho \| \pi)$. It is interesting to study if one could establish the same convergence rate with Wasserstein distance (which is the same as asking if one could establish a coupling argument for our result) or relative entropy.

Our decay estimate is based on the following Poincaré-type inequality in time-augmented space:

Theorem 2. *Under Assumptions 1, 2, and 3, there exist a universal constant C independent of all parameters, and a constant $R < \infty$ (the same constant as in Theorem 1) such that for every $f \in H^1_{hyp}(\lambda \otimes \mu)$, we have*

$$\|f - (f)_{\lambda \otimes \rho_\infty}\|_{L^2(\lambda \otimes \rho_\infty)} \leqslant C \left(\left(1 + RT + \frac{1}{(1 - e^{-\sqrt{m}T})^2} \right. \right.$$

$$\left. + \frac{R}{\sqrt{m}(1 - e^{-\sqrt{m}T})^2} \right) \|(\mathcal{I} - \Pi_v) f\|_{L^2(\lambda \otimes \rho_\infty)}$$

$$\left. + \left(\frac{1}{\sqrt{m}(1 - e^{-\sqrt{m}T})} + T \right) \|\partial_t f - \mathcal{L}_{ham} f\|_{L^2(\lambda \otimes \mu; H_\kappa^{-1})} \right). \qquad (15)$$

Let us give a brief introduction on the strategy of the proof, which is strongly motivated by the work of Armstrong and Mourrat [1]. A naive energy estimate and Gaussian Poincaré inequality yields

$$\frac{d}{dt} \|f(t, \cdot)\|^2_{L^2(\rho_\infty)} = -2\gamma \|\nabla_v f(t, \cdot)\|^2_{L^2(\rho_\infty)} \leqslant -2\gamma \|(\mathcal{I} - \Pi_v) f(t, \cdot)\|^2_{L^2(\rho_\infty)}.$$

While the above establishes the L^2 energy decay, it does not directly yield exponential decay rate. In particular, the energy dissipation is only present in velocity variable. However, instead of looking at single time slice, we should look at time intervals, since after time propagation, the dissipation in v together with the transport terms in x will lead to dissipation in x. Moreover, in the analysis, we are essentially treating the time variable t as another space variable alongside x. With the help a Poincaré-type inequality in the time-augmented state space established in Theorem 2, we can prove exponential convergence still using the standard energy estimate, in line with the moral "hypocoercivity is simply coercivity with respect to the correct norm", quoted from [1, Page 4].

To prove Theorem 2, as an educated reader might realize from [19], the elliptic regularity in x variable plays an important role in the estimates, which in Lemma 2.4 we made a mild generalization to the time-augmented space $L^2(\lambda \otimes \mu)$. However, in the proof of Theorem 2 when applying integration by parts, we need test functions that vanish at both boundary layers $t = 0$ and $t = T$, which is not necessarily satisfied by the derivatives of the solution to the elliptic equation (22). This is why we resort to Lemma 2.6 (also an extension of Bogovskii's operator [11] to $(I \times \mathbb{R}^d, \lambda \otimes \mu)$) for the solution of the divergence equation (25), which is a cornerstone of this proof. In particular, even for convex U, the constants in (15) blow up as $T \to 0$, which can be traced down to the estimate of $\psi'_{2,\alpha}$ in (35), and thus prevents us from working on single time slices.

1.2. A Literature Review and Comparison

Kinetic Fokker–Planck equation was first studied by Kolmogorov [34], and was the main motivation for Hörmander's theory on hypoelliptic equations [32], which gave an almost complete classification of second-order hypoelliptic operators. The earliest result regarding its exponential convergence were established in [52] for potentials with bounded Hessian, which was later generalized in [41,51,55]. There is a substantial amount of works in the literature for studying the exponential convergence of the underdamped Langevin dynamics. Below, we shall categorize them based on the norms and approaches to characterize the convergence.

(i) (Convergence in $H^1(\rho_\infty)$ norm). The exponential convergence of the kinetic Fokker–Planck equation in $H^1(\rho_\infty)$ was proved by Villani in [54, Theorem 35], which was inspired by early works of [27,29]. See also [53] for a brief overview of main ideas. The earlier work of [43] proved similar results on the torus without forcing term. Since $L^2(\rho_\infty)$ norm is controlled by $H^1(\rho_\infty)$ norm, this result automatically implies the convergence of (4) in $L^2(\rho_\infty)$. However, the decay rate therein is quite implicit; see [54, Sec. 7.2]. This approach is extended in [9] to possibly singular potentials with convergence rates given in certain cases.

(ii) (Convergence in a modified $L^2(\rho_\infty)$ norm). A more direct approach for convergence in $L^2(\rho_\infty)$ was developed by Dolbeault, Mouhot and Schmeiser in [18,19], see also earlier ideas in [28]. They identified a modified $L^2(\rho_\infty)$ norm, denoted by E, such that $\mathsf{E}(\rho(t, x, v)) \to 0$ exponentially fast for $\rho(t, \cdot, \cdot)$ evolving according to (2).

This hypocoercivity method was revisited and adapted in [17,26,47] to deal with the backward Kolmogorov equation (4), i.e., to show that $\mathsf{E}(f(t, \cdot, \cdot))$ decays to zero exponentially fast. In Appendix B.1, we will briefly revisit how to choose the Lyapunov function E, based on [16, Sec. 2], because their setup is consistent with our $L^2(\rho_\infty)$ estimate in Section 1.1 above. We would like to remark that while [47] gets some rate, for which the scalings in d and γ are known, it is difficult to determine the optimal γ for their convergence rate estimates.

As a remark, the DMS method [18,19] has been extended or adapted to study the convergence of spherical velocity Langevin equation [25], non-equilibrium Langevin dynamics [33], Langevin dynamics with general kinetic energy [49], temperature-accelerated molecular dynamics [50], adaptive Langevin dynamics [38], dynamics with Boltzmann-type dissipation [2], dynamics with singular potentials [12], just to name a few. It might be interesting to study whether the variational framework [1] we based on can be extended to these cases.

(iii) (Convergence in Wasserstein distance). Baudoin discussed a general framework of the Bakry–Émery methodology [5] to hypoelliptic and hypocoercive operators, based on which the exponential convergence of the kinetic Fokker–Planck equation (quantified by a Wasserstein distance associated with a special metric) was proved under certain assumptions on the potential $U(x)$ [7, Theorem 2.6]; see also [8].

A different approach is the coupling method for underdamped Langevin dynamics (1). In [16, Sec. 2], for strongly convex potential U, Dalalyan and Riou-Durand considered the mixing of the marginal distribution in the x coordinate, by a synchronous coupling argument; an estimate of the convergence rate was also explicitly provided, quantified by W_2 distance [16, Theorem 1]. For more general potentials, Eberle, Guillin and Zimmer developed a hybrid coupling method, composed of synchronous and reflection couplings, to study the exponential convergence of probability distributions for the underdamped Langevin dynamics (1), quantified by a Kantorovich semi-metric [20]. Unfortunately, their rates are dimension dependent in general.

(iv) (Convergence in relative entropy) Villani [54] obtained exponential convergence of kinetic Fokker–Planck in the case of potentials with bounded Hessian, which is extended in [8]. A more quantitative convergence rate is obtained in [40]. All of them essentially used Gamma calculus on a twisted metric so that derivatives in x direction can be introduced. In [13], exponential convergence of entropy is established for potentials that may not have bounded Hessians but satisfy a stronger weighted log-Sobolev inequality.

There are other approaches to study the long time behavior of the underdamped Langevin dynamics, e.g., Lyapunov function [4,41,51,55] and spectral analysis [21, 35]. There are also works that extend the aforementioned approaches to dynamics with singular potentials [9,12,14,15,30,39]. We will not go into details here.

While our work is not the first one that studies the exponential convergence of underdamped Langevin dynamics, our estimates are more quantitative, and in certain cases, sharper than any existing result. In particular, for a large class of convex potentials, we establish an $O(\sqrt{m})$ convergence rate after optimizing in γ, which is

Table 1. Summary of the convergence rate ν depending on d, m, L under the assumption $m\mathrm{Id} \leqslant \nabla_x^2 U \leqslant L\mathrm{Id}$ for the regime $m \ll 1 \ll L$

	Convergence rate for arbitrary γ	Convergence rate with optimal γ	Criterion
[9, Corollary 3.19]	$O(\frac{m\gamma^3}{\gamma^4+L^2})$	$O(\frac{m}{\sqrt{L}})$	twisted H^1
[16]	Only guarantees convergence for $\gamma \geqslant \sqrt{L}$	$O(\frac{m}{\sqrt{L}})$	W_2
[40, Proposition 1] (after rescaling)	Only guarantees convergence for $\gamma \geqslant \sqrt{L}$	$O(\frac{m}{\sqrt{L}})$	twisted KL
Our work	$O(\frac{m\gamma}{m+\gamma^2})$	$O(\sqrt{m})$	L^2

independent of dimension and only assumes a mild upper bound (Assumption 2) on the derivatives of the potential. To the best of our knowledge, this optimal $O(\sqrt{m})$ convergence rate is new in the literature.

Table 1 summarizes the previous results [9,16,40] under the assumption $m\mathrm{Id} \leqslant \nabla_x^2 U \leqslant L\mathrm{Id}$ (and hence guarantee Assumptions 1–3) in the most interesting regime $m \ll 1 \ll L$, with optimal choice of γ. To elaborate the comparison with result of [40], after a rescaling, they proved exponential convergence of (4) with friction parameter (using their notations) $\gamma\sqrt{\xi}$ and convergence rate $O(\frac{\lambda}{\sqrt{\xi}})$, with constraints that requires (see [40, Proof of Lemma 8])

$$\begin{cases} \frac{\xi}{2L} - \left(\frac{1}{4L} + \frac{1}{2m}\right)\lambda > 0 \\ \gamma\left(\frac{4\xi}{L} + 1\right) - \left(\frac{1}{2m} + \frac{2}{L}\right)\lambda > 0 \\ \frac{1}{2} - \frac{\xi}{2L} + \left(\frac{1}{4L} + \frac{1}{2m}\right)\lambda - \gamma\left(\frac{4\xi}{L} + 1\right) + \left(\frac{1}{2m} + \frac{2}{L}\right)\lambda \leqslant 0. \end{cases}$$

Combined, these yield $\xi \geqslant O(L)$ and $\lambda \leqslant O(m)$, which means the convergence rate cannot exceed $O(\frac{m}{\sqrt{L}})$. Moreover, they require $\gamma \geqslant O(1)$, or their friction parameter must be at least $O(\sqrt{L})$.

We also comment that in the case where $\|\nabla_x^2 U\| \leqslant L\mathrm{Id}$, but U is not necessarily convex, our convergence rate is $\nu = O(\frac{m}{\sqrt{L}})$ after optimizing in γ by choosing $\gamma \sim \sqrt{L}$, which matches the results of existing works [9,40].

2. Proofs

In this section, we present the statements and proofs of auxiliary lemmas, followed by the proofs of the two main theorems. Lemmas 2.1, 2.2 and 2.3 are the technical lemmas that prepare us for the elliptic regularity result in Lemma 2.4.

The proof of the divergence Lemma, which builds up from elliptic regularity, is presented in Lemma 2.6. The proof of Theorem 2 is then possible with the test functions obtained from Lemma 2.6. Finally we present the proof of Theorem 1 which follows from Theorem 2 and energy estimate.

We start with the Poincaré inquality on tensorized space $(I \times \mathbb{R}^d, \lambda \otimes \mu)$, which allows elliptic regularity to hold in the time-augmented state space. The proof is standard and is thus omitted.

Lemma 2.1. (Poincaré Inequality) *For* $f \in H^1(\lambda \otimes \mu)$,

$$\|f - (f)_{\lambda \otimes \mu}\|_{L^2(\lambda \otimes \mu)}^2 \leqslant \max\left\{\frac{1}{m}, \frac{T^2}{\pi^2}\right\} \left(\|\partial_t f\|_{L^2(\lambda \otimes \mu)}^2 + \|\nabla_x f\|_{L^2(\lambda \otimes \mu)}^2\right). \tag{16}$$

The next lemma is also a technical lemma, the goal of which is to show that under Assumption 2, $|\nabla^2 U|$ defines a bounded operator $H^1(\lambda \otimes \mu) \to L^2(\lambda \otimes \mu)$, which allows us to improve the regularity $u \in H^2(\lambda \otimes \mu)$ for u being the solution of (22) in the proof of Lemma 2.4.

Lemma 2.2. [54, Lemma A.24] *For any* $\phi \in H^1(\lambda \otimes \mu)$, *we have*

$$\|\phi \nabla_x U\|_{L^2(\lambda \otimes \mu)}^2 \leqslant 16\|\nabla_x \phi\|_{L^2(\lambda \otimes \mu)}^2 + 4Md\|\phi\|_{L^2(\lambda \otimes \mu)}^2, \tag{17}$$

where M is the constant in (11).

Proof.

$$\begin{aligned}
\|\phi \nabla_x U\|_{L^2(\lambda \otimes \mu)}^2 &= \int_{I \times \mathbb{R}^d} \phi^2 \nabla_x U \cdot \nabla_x U \, d\lambda(t) \, d\mu(x) \\
&= \int_{I \times \mathbb{R}^d} \nabla_x \cdot (\phi^2 \nabla_x U) \, d\lambda(t) \, d\mu(x) \\
&= 2 \int_{I \times \mathbb{R}^d} \phi \nabla_x \phi \cdot \nabla_x U \, d\lambda(t) \, d\mu(x) + \int_{I \times \mathbb{R}^d} \phi^2 \Delta_x U \, d\lambda(t) \, d\mu(x) \\
&\overset{(11)}{\leqslant} \frac{1}{4} \|\phi \nabla_x U\|_{L^2(\lambda \otimes \mu)}^2 + 4\|\nabla_x \phi\|_{L^2(\lambda \otimes \mu)}^2 \\
&\quad + Md\|\phi\|_{L^2(\lambda \otimes \mu)}^2 + \frac{\delta}{2} \int_{I \times \mathbb{R}^d} \phi^2 |\nabla_x U|^2 \, d\lambda(t) \, d\mu(x).
\end{aligned}$$

We thus finish the proof of (17) after rearranging and using $\delta < 1$. □

Next we present a technical lemma that prepares us for the (mixed space-time) H^2 estimates of u, the solution of the elliptic equation (22). This is a generalization of a similar L^2–H^2 regularity estimate in [19, Proposition 5], where only the spatial variable is considered, but our estimates are algebraically simpler thanks to Bochner's formula. Let us remark that we adopt the same scaling of parameters as [10, Lemma 3.6], especially in the most general case (iii).

Lemma 2.3. *For any* $u \in H^2(\lambda \otimes \mu)$ *such that* $\overline{\nabla} u \in H_0^1(\lambda \otimes \mu)^{d+1}$,

$$\|D^2 u\|_{L^2(\lambda \otimes \mu)}^2 = \sum_{i,j=0}^{d} \|\partial_{x_i} \partial_{x_j} u\|_{L^2(\lambda \otimes \mu)}^2 \leqslant C\left(\|\mathscr{L} u\|_{L^2(\lambda \otimes \mu)}^2 + R^2 \|\nabla_x u\|_{L^2(\lambda \otimes \mu)}^2\right), \tag{18}$$

Similarly,

$$\|\nabla_x^2 u\|_{L^2(\lambda\otimes\mu)}^2 \leqslant C\left(\|\nabla_x^* \nabla_x u\|_{L^2(\lambda\otimes\mu)}^2 + R^2\|\nabla_x u\|_{L^2(\lambda\otimes\mu)}^2\right). \qquad (19)$$

Here C is a universal constant whose the precise value can be traced in the proof under different assumptions in Theorem 1, *and R is defined in Theorem* 1.

Proof. We only prove (18) since the proof of (19) follows from a similar argument. The starting point of the proof is Bochner's formula

$$\sum_{i,j=0}^{d} |\partial_{x_i,x_j} u|^2 = \overline{\nabla} u \cdot \overline{\nabla}\mathscr{L}u - (\nabla_x u)^\top \nabla_x^2 U \nabla_x u - \mathscr{L}\frac{|\overline{\nabla} u|^2}{2}.$$

Integrate over $\lambda \otimes \mu$ and (noticing the last term above has integral zero) we get

$$\sum_{i,j=0}^{d} \|\partial_{x_i,x_j} u\|_{L^2(\lambda\otimes\mu)}^2 = \|\mathscr{L}u\|_{L^2(\lambda\otimes\mu)}^2 - \int_{I\times\mathbb{R}^d} (\nabla_x u)^\top \nabla_x^2 U \nabla_x u \, d\lambda(t) \, d\mu(x).$$

$$(20)$$

This already verifies the conclusion in cases (i) (setting $K = 0$) and (ii) with $C = 1$.

Now we deal with the more general case, without assuming (14). Using (17) with $\phi = \partial_{x_i} u$, $i = 1, \cdots, d$,

$$\int_{I\times\mathbb{R}^d} |\nabla_x u|^2 |\nabla_x U|^2 \, d\lambda(t) \, d\mu(x)$$

$$= \sum_{i=1}^{d} \int_{I\times\mathbb{R}^d} (\partial_{x_i} u)^2 |\nabla_x U|^2 \, d\lambda(t) \, d\mu(x)$$

$$\overset{(17)}{\leqslant} 16\|D_x^2 u\|_{L^2(\lambda\otimes\mu)}^2 + 4Md \int_{I\times\mathbb{R}^d} |\nabla_x u|^2 \, d\lambda(t) \, d\mu(x)$$

$$\overset{(20)}{=} 16\|\mathscr{L}u\|_{L^2(\lambda\otimes\mu)}^2 + 4Md \int_{I\times\mathbb{R}^d} |\nabla_x u|^2 \, d\lambda(t) \, d\mu(x)$$

$$- 16 \int_{I\times\mathbb{R}^d} (\nabla_x u)^\top \nabla_x^2 U \nabla_x u \, d\lambda(t) \, d\mu(x)$$

$$\overset{(11)}{\leqslant} 16\|\mathscr{L}u\|_{L^2(\lambda\otimes\mu)}^2 + 4Md \int_{I\times\mathbb{R}^d} |\nabla_x u|^2 \, d\lambda(t) \, d\mu(x)$$

$$+ 16M \int_{I\times\mathbb{R}^d} |\nabla_x u|^2 (\sqrt{d} + |\nabla_x U|) \, d\lambda(t) \, d\mu(x)$$

$$\overset{d\geqslant 1}{\leqslant} 16\|\mathscr{L}u\|_{L^2(\lambda\otimes\mu)}^2 + 20Md \int_{I\times\mathbb{R}^d} |\nabla_x u|^2 \, d\lambda(t) \, d\mu(x)$$

$$+ 128M^2 \int_{I\times\mathbb{R}^d} |\nabla_x u|^2 \, d\lambda(t) \, d\mu(x) + \frac{1}{2} \int_{I\times\mathbb{R}^d} |\nabla_x u|^2 |\nabla_x U|^2 \, d\lambda(t) \, d\mu(x).$$

Rearranging the terms, we arrive at

$$\int_{I\times\mathbb{R}^d} |\nabla_x u|^2 |\nabla_x U|^2 \, d\lambda(t) \, d\mu(x) \leqslant 32\|\mathscr{L}u\|_{L^2(\lambda\otimes\mu)}^2$$
$$+(40Md + 256M^2)\int_{I\times\mathbb{R}^d} |\nabla_x u|^2 \, d\lambda(t) \, d\mu(x). \qquad (21)$$

Therefore by (21) and triangle inequality,

$$\|D^2 u\|_{L^2(\lambda\otimes\mu)}^2 \overset{(11),(20)}{\leqslant} \|\mathscr{L}u\|_{L^2(\lambda\otimes\mu)}^2 + M\int_{I\times\mathbb{R}^d} |\nabla_x u|^2(\sqrt{d} + |\nabla_x U|) \, d\lambda(t) \, d\mu(x)$$

$$\leqslant \|\mathscr{L}u\|_{L^2(\lambda\otimes\mu)}^2$$
$$+ M\sqrt{d}\|\nabla_x u\|_{L^2(\lambda\otimes\mu)}^2 + M\|\nabla_x u\|_{L^2(\lambda\otimes\mu)}^2 \||\nabla_x u||\nabla_x U|\|_{L^2(\lambda\otimes\mu)}$$

$$\overset{(21)}{\leqslant} \|\mathscr{L}u\|_{L^2(\lambda\otimes\mu)}^2 + M\sqrt{d}\|\nabla_x u\|_{L^2(\lambda\otimes\mu)}^2$$
$$+ M\|\nabla_x u\|_{L^2(\lambda\otimes\mu)} \left(6\|\mathscr{L}u\|_{L^2(\lambda\otimes\mu)} + (16M + \sqrt{40Md})\|\nabla_x u\|_{L^2(\lambda\otimes\mu)}\right)$$

$$\leqslant 4\|\mathscr{L}u\|_{L^2(\lambda\otimes\mu)}^2 + (19M^2 + M\sqrt{40Md})\|\nabla_x u\|_{L^2(\lambda\otimes\mu)}^2.$$

One of the key lemmas of our proof is the following result on elliptic regularity on the space $(I \times \mathbb{R}^d, \lambda \otimes \mu)$ (the solution to such elliptic equation will play an important role in the proof of Lemma 2.6):

Lemma 2.4. *Consider the following elliptic equation:*

$$\begin{cases} \mathscr{L}u = h & in \ I \times \mathbb{R}^d, \\ \partial_t u(t = 0, \cdot) = \partial_t u(t = T, \cdot) = 0 & in \ \mathbb{R}^d. \end{cases} \qquad (22)$$

Assume $h \in H^{-1}(\lambda \otimes \mu)$, and $(h)_{\lambda\otimes\mu} = 0$. Define the function space

$$V = \left\{ u \in H^1(\lambda \otimes \mu) \ : \ (u)_{\lambda\otimes\mu} = 0 \right\}.$$

Then

(i) There exists a unique $u \in V$ which is a weak solution to (22). More precisely, for any $v \in H^1(\lambda \otimes \mu)$, we have

$$\int_{I\times\mathbb{R}^d} (\partial_t u \partial_t v + \nabla_x u \cdot \nabla_x v) \, d\lambda(t) \, d\mu(x) = \int_{I\times\mathbb{R}^d} h v \, d\lambda(t) \, d\mu(x).$$

Moreover, when $h \in L^2(\lambda \otimes \mu)$, we have the estimate

$$\|\partial_t u\|_{L^2(\lambda\otimes\mu)}^2 + \|\nabla_x u\|_{L^2(\lambda\otimes\mu)}^2 \leqslant \max\left\{\frac{1}{m}, \frac{T^2}{\pi^2}\right\} \|h\|_{L^2(\lambda\otimes\mu)}^2. \qquad (23)$$

(ii) If $h \in L^2(\lambda \otimes \mu)$, then the solution u to (22) satisfies $u \in H^2(\lambda \otimes \mu)$.

Remark 2.5. One could in fact estimate $\|u\|_{H^1(\lambda\otimes\mu)}$ using only $\|h\|_{H^{-1}(\lambda\otimes\mu)}$, but with a slightly worsened constant $\max\{\frac{1}{m}, \frac{T^2}{\pi^2}, 1\}$ on the rhs. Since in our applications we only use $\|h\|_{L^2(\lambda\otimes\mu)}$, we opt for the current version of (23) for simplicity.

Proof. (i) V is a linear Hilbert space and has non-zero elements (any function constant in t, and H^1 and mean zero in x is included in V). Moreover, V is a subspace of $H^1(\lambda \otimes \mu)$, and for the rest of the paper we equip it with the $H^1(\lambda \otimes \mu)$ norm. We also define the following inner-product:

$$B(u, v) := \int_{I \times \mathbb{R}^d} (\partial_t u \partial_t v + \nabla_x u \cdot \nabla_x v) \, \mathrm{d}\lambda(t) \, \mathrm{d}\mu(x).$$

One can easily verify $B(\cdot, \cdot)$ is an inner product on V. Notice that if $B(u, u) = 0$ then $\partial_t u = \nabla_x u = 0$, leaving u to be a constant, which has to be 0 since $(u)_{\lambda \otimes \mu} = 0$. If u is a weak solution of (22), then for any $v \in V$, $B(u, v) = \int_{I \times \mathbb{R}^d} h v \, \mathrm{d}\lambda(t) \, \mathrm{d}\mu(x)$, and necessarily $(h)_{\lambda \otimes \mu} = 0$ when we take $v = 1$.

Since $(u)_{\lambda \otimes \mu} = 0$, by Poincaré inequality (Lemma 2.1) we can show B is coercive under $H^1(\lambda \otimes \mu)$ norm in the sense of

$$
\begin{aligned}
B[u, u] &= \|\partial_t u\|^2_{L^2(\lambda \otimes \mu)} + \|\nabla_x u\|^2_{L^2(\lambda \otimes \mu)} \\
&\geqslant \frac{1}{C} (\|\partial_t u\|^2_{L^2(\lambda \otimes \mu)} + \|\nabla_x u\|^2_{L^2(\lambda \otimes \mu)} + \|u\|^2_{L^2(\lambda \otimes \mu)}) \\
&= \frac{1}{C} \|u\|^2_{H^1(\lambda \otimes \mu)}.
\end{aligned}
$$

We can also show B is bounded above since it is an inner-product and $B[u, u] \leqslant \|u\|^2_{H^1(\lambda \otimes \mu)}$. Define a linear functional on V: $H(v) := \int_{I \times \mathbb{R}^d} h v \, \mathrm{d}\lambda(t) \, \mathrm{d}\mu(x)$. One can verify the boundedness of H:

$$|H(v)| \leqslant \|h\|_{H^{-1}(\lambda \otimes \mu)} \|v\|_{H^1(\lambda \otimes \mu)}.$$

Thus by Lax–Milgram's Theorem, the equation (22) has a unique weak solution $u \in V$. Moreover,

$$
\begin{aligned}
(\|\partial_t u\|^2_{L^2(\lambda \otimes \mu)} + \|\nabla_x u\|^2_{L^2(\lambda \otimes \mu)})^2 &= B[u, u]^2 \\
&= \left(\int_{I \times \mathbb{R}^d} h u \, \mathrm{d}\lambda(t) \, \mathrm{d}\mu(x) \right)^2 \leqslant \|h\|^2_{L^2(\lambda \otimes \mu)} \|u\|^2_{L^2(\lambda \otimes \mu)} \\
&\overset{(16)}{\leqslant} \max\left\{ \frac{1}{m}, \frac{T^2}{\pi^2} \right\} \|h\|^2_{L^2(\lambda \otimes \mu)} \left(\|\partial_t u\|^2_{L^2(\lambda \otimes \mu)} + \|\nabla_x u\|^2_{L^2(\lambda \otimes \mu)} \right),
\end{aligned}
$$

and the desired estimate follows.

(ii) For each $i = 1, 2, \cdots, d$, consider the elliptic equation

$$
\begin{cases}
\mathscr{L} w_i = \partial_{x_i} h - \nabla_x u \cdot \nabla_x \partial_{x_i} U & \text{in } I \times \mathbb{R}^d, \\
\partial_t w_i(t = 0, \cdot) = \partial_t w_i(t = T, \cdot) = 0 & \text{in } \mathbb{R}^d.
\end{cases}
\tag{24}
$$

The motivation of considering (24) is that, if we formally differentiate (22) with respect to ∂_{x_i}, then $\partial_{x_i} u$ satisfies precisely the equation (24) for w_i. Hence, our plan is to use part (i) to establish $w_i \in H^1(\lambda \otimes \mu)$, then argue that $w_i - \partial_{x_i} u$ must be constant.

We first verify the rhs of (24) has total integral zero. Indeed

$$
\int_{I\times\mathbb{R}^d} (\partial_{x_i} h - \nabla_x u \cdot \nabla_x \partial_{x_i} U)\, d\lambda(t)\, d\mu(x)
$$
$$
= \int_{I\times\mathbb{R}^d} (h\partial_{x_i} U - \nabla_x u \cdot \nabla_x \partial_{x_i} U)\, d\lambda(t)\, d\mu(x)
$$
$$
= \int_{I\times\mathbb{R}^d} \left(\mathscr{L} u\, \partial_{x_i} U - \nabla_x u \cdot \nabla_x \partial_{x_i} U\right)\, d\lambda(t)\, d\mu(x)
$$
$$
= \int_{I\times\mathbb{R}^d} \left(\partial_t u\, \partial_{tx_i} U + \nabla_x u \cdot \nabla_x \partial_{x_i} U - \nabla_x u \cdot \nabla_x \partial_{x_i} U\right)\, d\lambda(t)\, d\mu(x) = 0.
$$

The next step is to show rhs is in $H^{-1}(\lambda\otimes\mu)$. Pick a test function $\phi\in H^1(\lambda\otimes\mu)$ with $\|\phi\|_{H^1(\lambda\otimes\mu)}=1$, and, by Lemma 2.2,

$$
\int_{I\times\mathbb{R}^d} (\partial_{x_i} h - \nabla_x u \cdot \nabla_x \partial_{x_i} U)\phi\, d\lambda(t)\, d\mu(x)
$$
$$
\leqslant \int_{I\times\mathbb{R}^d} (-h\partial_{x_i}\phi + h\phi\,\partial_{x_i} U)\, d\lambda(t)\, d\mu(x) + \int_{I\times\mathbb{R}^d} |\phi\nabla_x u||\nabla_x \partial_{x_i} U|\, d\lambda(t)\, d\mu(x)
$$
$$
\overset{(11)}{\leqslant} \|h\|_{L^2(\lambda\otimes\mu)}(1+\|\phi\partial_{x_i} U\|_{L^2(\lambda\otimes\mu)}) + M\int_{I\times\mathbb{R}^d} |\phi\nabla_x u|(\sqrt{d}+|\nabla_x U|)\, d\lambda(t)\, d\mu(x)
$$
$$
\leqslant \|h\|_{L^2(\lambda\otimes\mu)}(1+\|\phi\partial_{x_i} U\|_{L^2(\lambda\otimes\mu)}) + M\|\nabla_x u\|_{L^2(\lambda\otimes\mu)}(\sqrt{d}+\|\phi\nabla_x U\|_{L^2(\lambda\otimes\mu)})
$$
$$
\overset{(17),(23)}{\leqslant} C(M,d)\|h\|_{L^2(\lambda\otimes\mu)},
$$

where $C(M,d)>0$ is a constant depending on M,d. Therefore, by (i) we know there exists a $w_i\in V$ which is the weak solution of (24). Finally, comparing (22) and (24), we observe that $\mathscr{L}(w_i - \partial_{x_i} u)=0$ in the sense of distributions, which by (i) indicates $w_i - \partial_{x_i} u$ must be constant, which must be $-(\partial_{x_i} u)_{\lambda\otimes\mu}$, since by construction $w\in V$ and $(w)_{\lambda\otimes\mu}=0$. This also means $\partial_{x_i} u\in H^1(\lambda\otimes\mu)$ since $w_i\in H^1(\lambda\otimes\mu)$. We end the proof of $u\in H^2(\lambda\otimes\mu)$ by writing $\partial_{tt} u = \nabla_x^* \nabla_x u - h \in L^2(\lambda\otimes\mu)$. $\quad\square$

We finally need a lemma for the solution of a divergence equation with Dirichlet boundary conditions. The resolution of divergence equation is an important tool in mathematical fluid dynamics (see the book [23, Section III.3]). However, in order to obtain more natural estimate on the constants, instead of resorting to the aforementioned Bogovskii's operator, we take advantage of the structure of space $L^2(\mu)$ by eigenspace decomposition, which is made possible thanks to Assumption 3. This will provide us test functions which play a crucial role in the proof of Theorem 2.

Lemma 2.6. *For any function $f\in L^2(\lambda\otimes\mu)$ with $(f)_{\lambda\otimes\mu}=0$, there exist two functions $\phi_0\in H_0^1(\lambda\otimes\mu)$ and $\Phi\in H^2(\lambda\otimes\mu)$ such that $\nabla_x\Phi\in H_0^1(\lambda\otimes\mu)^d$ and*

$$
-\partial_t\phi_0 + \nabla_x^*\nabla_x\Phi = f \tag{25}
$$

with estimates

$$\|\phi_0\|_{L^2(\lambda\otimes\mu)} + \|\nabla_x\Phi\|_{L^2(\lambda\otimes\mu)} \leqslant C\left(\frac{1}{\sqrt{m}(1-e^{-\sqrt{m}T})} + T\right)\|f\|_{L^2(\lambda\otimes\mu)}$$

$$(26)$$

and

$$\|\nabla_x\phi_0\|_{L^2(\lambda\otimes\mu)} + \|\overline{\nabla}\nabla_x\Phi\|_{L^2(\lambda\otimes\mu)} \leqslant C\left(1 + RT + \frac{1}{(1-e^{-\sqrt{m}T})^2}\right.$$

$$\left. + \frac{R}{\sqrt{m}(1-e^{-\sqrt{m}T})^2}\right)\|f\|_{L^2(\lambda\otimes\mu)}.$$

$$(27)$$

Here C is a universal constant and R is the constant defined in Theorem 1.

Remark 2.7. We believe the correct scaling of the rhs should be $O(\frac{1}{T})$ as $T \to 0$, which we are unable to obtain, due to the pessimistic estimates in the last two lines of (31) that changed the scaling of the last two terms from $O(1)$ to $O(T^2)$, but will not pursue further since in the proof of Theorem 1 we only take $T = \frac{1}{\sqrt{m}}$. As we mentioned iearlier after Theorem 2, the scaling of $O(\frac{1}{T})$ as $T \to 0$ should come from (35).

Before we proceed to the proof, let us give a brief heuristic argument on why we need to introduce the space of harmonic functions (i.e. the space $\mathbb{H}$ that appears at the beginning of the proof) and consider orthogonal projection on it. Indeed, a direct way to look for a solution of (25) is to look for that of (22) and set $\phi_0 = \partial_t u$, $\Phi = u$. However, these test functions do not satisfy the appropriate boundary conditions. In particular, if solution of (22) satisfy $\nabla_x u(t = 0, \cdot) = \nabla_x u(t = T, \cdot) = 0$, then necessarily f has to be perpendicular to the space of harmonic functions. Meanwhile, the harmonic part of f requires special treatment from us and brings technical difficulty to the proof. However, thanks to Assumption 3, one can decompose the harmonic part of f using separation of variables, which enables us to obtain the solution of divergence equation by constructing it for each component and adding them up.

Proof. Let $\mathbb{H}$ be the subspace of $L^2(\lambda\otimes\mu)$ that consists of "harmonic functions", in other words, $f \in \mathbb{H}$ if and only if $\mathscr{L}f = 0$. We consider the decomposition $f = f^{(1)} + f^{(2)}$ where $f^{(1)} \in \mathbb{H}$ and $f^{(2)} \perp \mathbb{H}$. Since $1 \in \mathbb{H}$ we know $(f^{(2)})_{\lambda\otimes\mu} = 0$ and hence $(f^{(1)})_{\lambda\otimes\mu} = 0$. Therefore by linearity it suffices to consider $f^{(1)}$ and $f^{(2)}$ separately. For $f^{(2)}$, the equation

$$\begin{cases} \mathscr{L}u = f^{(2)} & \text{in } I \times \mathbb{R}^d, \\ \partial_t u(t = 0, \cdot) = \partial_t u(t = T, \cdot) = 0 & \text{in } \mathbb{R}^d \end{cases}$$

$$(28)$$

has a unique solution in $V \cap H^2(\lambda\otimes\mu)$ by Lemma 2.4. Moreover, for any $v \in \mathbb{H} \cap H^2(\lambda\otimes\mu)$, integration by parts yields

$$0 = \int_{I\times\mathbb{R}^d} f^{(2)}v \, d\lambda(t) \, d\mu(x) = B[u, v]$$

$$= \int_{I \times \mathbb{R}^d} u \mathscr{L} v \, d\lambda(t) \, d\mu(x) + \int_{\mathbb{R}^d} (u(T)\partial_t v(T) - u(0)\partial_t v(0)) \, d\mu(x)$$

Therefore, since v is arbitrary, we have $u(T) = u(0) = 0$, which implies $\nabla_x u \in H_0^1(\lambda \otimes \mu)^d$. Also by construction of boundary conditions $\partial_t u \in H_0^1(\lambda \otimes \mu)$. Thus for $f^{(2)}$ part, it suffices to take correspondingly $\phi_0^{(2)} = \partial_t u$, $\Phi^{(2)} = u$ with the estimates

$$\|\overline{\nabla} u\|_{L^2(\lambda \otimes \mu)}^2 \overset{(23)}{\leqslant} C \max \left\{ \frac{1}{m}, T^2 \right\} \|f^{(2)}\|_{L^2(\lambda \otimes \mu)}^2, \tag{29}$$

and

$$\|D^2 u\|_{L^2(\lambda \otimes \mu)}^2 \overset{(18),(29)}{\leqslant} C \left(1 + \frac{R^2}{m} + R^2 T^2 \right) \|f^{(2)}\|_{L^2(\lambda \otimes \mu)}^2. \tag{30}$$

We now consider the $f^{(1)}$ part. Since $\{1\} \cup \{w_\alpha\}$ forms an orthonormal basis in $L^2(\mu)$ and $(f^{(1)})_{\lambda \otimes \mu} = 0$, we have an orthogonal decomposition

$$f^{(1)}(t, x) = f_0(t) + \sum_\alpha f_\alpha(t) w_\alpha(x).$$

Since $f^{(1)}$ is harmonic,

$$0 = \mathscr{L} f^{(1)} = -f_0''(t) + \sum_\alpha (-f_\alpha''(t) + \alpha^2 f_\alpha(t)) w_\alpha(x)$$

and therefore $f_0(t)$ is an affine function $f_0(t) = c_0(t - \frac{T}{2})$ for some constant c_0, as $f_0(t)$ has integral zero. Moreover for $\alpha > 0$ there exist constants $c_\pm^\alpha$ such that

$$f_\alpha(t) = c_+^\alpha e^{-\alpha t} + c_-^\alpha e^{-\alpha(T-t)}.$$

Therefore, by orthogonality in $L^2(\lambda \otimes \mu)$, we can write for some constant $C \in (1, \infty)$,

$$\|f\|_{L^2(\lambda \otimes \mu)}^2 = \|f^{(2)}\|_{L^2(\lambda \otimes \mu)}^2 + c_0^2 \|t - \frac{T}{2}\|_{L^2(\lambda)}^2 + \sum_\alpha \|c_+^\alpha e^{-\alpha t} + c_-^\alpha e^{-\alpha(T-t)}\|_{L^2(\lambda)}^2$$

$$= \|f^{(2)}\|_{L^2(\lambda \otimes \mu)}^2 + \frac{T^2 c_0^2}{12} + \sum_\alpha \left(((c_+^\alpha)^2 + (c_-^\alpha)^2) \frac{1 - e^{-2\alpha T}}{2\alpha T} + 2c_+^\alpha c_-^\alpha e^{-\alpha T} \right)$$

$$\geqslant \|f^{(2)}\|_{L^2(\lambda \otimes \mu)}^2 + \frac{T^2 c_0^2}{12} + \sum_\alpha ((c_+^\alpha)^2 + (c_-^\alpha)^2) \left(\frac{1 - e^{-2\alpha T}}{2\alpha T} - e^{-\alpha T} \right)$$

$$\geqslant \|f^{(2)}\|_{L^2(\lambda \otimes \mu)}^2 + \frac{T^2 c_0^2}{12} + \frac{1}{C} \sum_\alpha ((c_+^\alpha)^2 + (c_-^\alpha)^2) \frac{(1 - e^{-\alpha T})^3}{\alpha T}. \tag{31}$$

The construction of test functions for $f_0(t)$ is straightforward: We simply take $\Phi^{(0)} = 0$ and $\phi_0^{(0)}(t, x) = \frac{c_0}{2}(t^2 - tT)$. We then construct $\phi_{0,\alpha}$, Φ_α for each component of the sum $e^{-\alpha t} w_\alpha(x)$, and therefore the functions $\phi_{0,\alpha}(T - t, \cdot)$, $\Phi_\alpha(T - t, \cdot)$ also apply to the component $e^{-\alpha(T-t)} w_\alpha(x)$, so that the eventual test functions

ϕ_0, Φ can be obtained after taking linear combination. The goal is to find $\phi_{0,\alpha}$, Φ_α such that

$$-\partial_t \phi_{0,\alpha} + \nabla_x^* \nabla_x \Phi_\alpha = e^{-\alpha t} w_\alpha(x).$$

Since $w_\alpha \in H^2(\lambda \otimes \mu)$, in order to eliminate the x part of the equation, we can take the natural ansatz by separation of variables $\phi_{0,\alpha} = \psi_{1,\alpha}(t) w_\alpha(x)$ and $\Phi_\alpha = \psi_{2,\alpha}(t) w_\alpha(x)$, and the two functions $\psi_{1,\alpha}(t)$, $\psi_{2,\alpha}(t)$ should satisfy $\psi_{1,\alpha}(0) = \psi_{1,\alpha}(T) = \psi_{2,\alpha}(0) = \psi_{2,\alpha}(T) = 0$ as well as the equation

$$- \psi_{1,\alpha}'(t) + \alpha^2 \psi_{2,\alpha}(t) = e^{-\alpha t}. \tag{32}$$

Integrating (32) against t, we obtain the necessary and sufficient condition

$$\int_0^T \psi_{2,\alpha}(t)\, dt = \frac{1 - e^{-\alpha T}}{\alpha^3}. \tag{33}$$

Of course there exists infinitely many possible solutions, since for any $\psi_{2,\alpha}$ that vanishes at both time boundaries and satisfies (33), the choice $\psi_{1,\alpha} = \int_0^t (\alpha^2 \psi_{2,\alpha}(\tau) - e^{-\alpha \tau})\, d\tau$ also vanishes at both time boundaries. Therefore we only need to choose a particular one to satisfy the desired estimates. Let us introduce a short-hand notation $\ell = e^{-\alpha T} \in (0, 1)$. Our idea is to find $\psi_{2,\alpha}$ of the form $\psi_{2,\alpha}(t) = \frac{1}{\alpha^2} g(e^{-\alpha t})$, which after a change of variable $s := e^{-\alpha t}$ turns the condition (33) into $\int_\ell^1 \frac{g(s)}{s}\, ds = 1 - \ell$, and the boundary conditions into $g(1) = g(\ell) = 0$. Hence, we may finish our construction by picking $g(s) = sh(s)$ with

$$h(x) = \frac{6}{(1-\ell)^2}(x - \ell)(1 - x).$$

From the expression we can directly derive (using $\alpha \geqslant \sqrt{m}$)

$$0 \leqslant g(s) \leqslant \frac{3}{2}s \quad \text{and} \quad |g'(s)| \leqslant \frac{4}{1-\ell} = \frac{4}{1 - e^{-\alpha T}}.$$

One can explicitly compute

$$\|\psi_{2,\alpha}\|_{L^2(\lambda)}^2 = \frac{1}{\alpha^4 T} \int_0^T g(e^{-\alpha t})^2\, dt = \frac{1}{\alpha^5 T} \int_\ell^1 \frac{g(s)^2}{s}\, ds = \frac{3(1 - e^{-2\alpha T})}{5\alpha^5 T}, \tag{34}$$

$$\text{and } \|\psi_{2,\alpha}'\|_{L^2(\lambda)}^2 = \frac{1}{\alpha^2 T} \int_0^T g'(e^{-\alpha t})^2 e^{-2\alpha t}\, dt = \frac{1}{\alpha^3 T} \int_\ell^1 g'(s)^2 s\, ds \leqslant \frac{8}{\alpha^3 T(1 - e^{-\alpha T})}. \tag{35}$$

Moreover since $\psi_{1,\alpha}'(t) = \alpha^2 \psi_{2,\alpha}(t) - e^{-\alpha t}$ from (32),

$$\|\psi_{1,\alpha}'\|_{L^2(\lambda)}^2 \leqslant 2\alpha^4 \|\psi_{2,\alpha}\|_{L^2(\lambda)}^2 + \frac{1 - e^{-2\alpha T}}{\alpha T} \leqslant \frac{3(1 - e^{-2\alpha T})}{\alpha T}. \tag{36}$$

Finally since

$$\psi_{1,\alpha}(t) = \int_0^t (g(e^{-\alpha s}) - e^{-\alpha s})\, ds = \frac{1}{\alpha} \int_{e^{-\alpha t}}^1 \left(\frac{g(\tau)}{\tau} - 1\right) d\tau = \frac{1}{\alpha} r(e^{-\alpha t})$$

with

$$r(s) = \int_s^1 (h(\tau) - 1)\, d\tau = \frac{(s - \ell)(1 - s)(1 + \ell - 2s)}{(1 - \ell)^2},$$

we can estimate

$$\alpha^2 \|\psi_{1,\alpha}\|_{L^2(\lambda)}^2 = \frac{1}{\alpha T} \int_\ell^1 \frac{r(t)^2}{t}\, dt = \frac{(1 - \ell)^3}{\alpha T}$$

$$\int_0^1 \frac{s^2 (1 - s)^2 (1 - 2s)^2}{(1 - \ell)s + \ell}\, ds \leqslant \frac{C(1 - e^{-\alpha T})^3}{\alpha T}. \tag{37}$$

To sum up, our construction of test functions can be write as

$$\phi_0 = \partial_t u + c_0 \frac{t^2 - tT}{2} + \sum_\alpha (c_+^\alpha \psi_{1,\alpha}(t) + c_-^\alpha \psi_{1,\alpha}(T - t)) w_\alpha(x),$$

$$\Phi = u + \sum_\alpha (c_+^\alpha \psi_{2,\alpha}(t) + c_-^\alpha \psi_{2,\alpha}(T - t)) w_\alpha(x),$$

here we recall that u is the solution of (28).

We now establish the estimates by direct calculations, which is possible since the variables are separated. Notice that for α, β,

$$\langle \nabla_x w_\alpha, \nabla_x w_\beta \rangle_{L^2(\mu)} = \langle w_\alpha, \nabla_x^* \nabla_x w_\beta \rangle_{L^2(\mu)} = \beta^2 \langle w_\alpha, w_\beta \rangle_{L^2(\mu)} = \alpha^2 \delta_{\alpha,\beta},$$

hence cross terms in the expansion of $\| \sum_\alpha (c_+^\alpha \psi_{2,\alpha}(t) + c_-^\alpha \psi_{2,\alpha}(T - t)) \nabla_x w_\alpha(x) \|_{L^2(\lambda \otimes \mu)}^2$ vanish. Therefore, we can estimate

$$\|\phi_0\|_{L^2(\lambda \otimes \mu)}^2 + \|\nabla_x \Phi\|_{L^2(\lambda \otimes \mu)}^2$$

$$\leqslant 3 \left(\|\partial_t u\|_{L^2(\lambda \otimes \mu)}^2 + \frac{c_0^2}{4} \|t^2 - tT\|_{L^2(\lambda)}^2 \right.$$

$$+ \sum_\alpha \|c_+^\alpha \psi_{1,\alpha}(t) + c_-^\alpha \psi_{1,\alpha}(T - t)\|_{L^2(\lambda)}^2 \|w_\alpha\|_{L^2(\mu)}^2$$

$$\left. + \|\nabla_x u\|_{L^2(\lambda \otimes \mu)}^2 + \left\| \sum_\alpha (c_+^\alpha \psi_{2,\alpha}(t) + c_-^\alpha \psi_{2,\alpha}(T - t)) \nabla_x w_\alpha \right\|_{L^2(\lambda \otimes \mu)}^2 \right)$$

$$\overset{(23)}{\leqslant} 6 \left(\max\{\frac{1}{m}, T^2\} \|f^{(2)}\|_{L^2(\lambda \otimes \mu)}^2 + \frac{c_0^2 T^4}{120} + \sum_\alpha ((c_+^\alpha)^2 + (c_-^\alpha)^2) \|\psi_{1,\alpha}\|_{L^2(\lambda)}^2 \right.$$

$$\left. + \sum_\alpha \|c_+^\alpha \psi_{2,\alpha}(t) + c_-^\alpha \psi_{2,\alpha}(T - t)\|_{L^2(\lambda)}^2 \|\nabla_x w_\alpha\|_{L^2(\mu)}^2 \right)$$

$$\leqslant C \left(\max\{\frac{1}{m}, T^2\} \|f^{(2)}\|_{L^2(\lambda \otimes \mu)}^2 + c_0^2 T^4 \right.$$

$$\left. + \sum_\alpha ((c_+^\alpha)^2 + (c_-^\alpha)^2)(\|\psi_{1,\alpha}\|_{L^2(\lambda)}^2 + \alpha^2 \|\psi_{2,\alpha}\|_{L^2(\lambda)}^2) \right)$$

$$\overset{(37),(34)}{\leqslant} C\left(\max\{\frac{1}{m}, T^2\}\left\|f^{(2)}\right\|^2_{L^2(\lambda\otimes\mu)} + c_0^2 T^4 \right.$$
$$\left. + \sum_\alpha \frac{1}{\alpha^2}((c_+^\alpha)^2 + (c_-^\alpha)^2)\frac{(1-e^{-\alpha T})^3 + 1 - e^{-2\alpha T}}{\alpha T}\right)$$
$$\overset{(31)}{\leqslant} C \max\left\{\frac{1}{m(1-e^{-\sqrt{m}T})^2}, T^2\right\}\|f\|^2_{L^2(\lambda\otimes\mu)}. \tag{38}$$

Here in the last line when we used (31), the worse factor $(1-e^{-\sqrt{m}T})^{-2}$ comes only from the last term on the line above. This establishes (26). Using similar arguments, we can estimate

$$\|\nabla_x \phi_0\|^2_{L^2(\lambda\otimes\mu)} = \left\|\nabla_x \partial_t u + \sum_\alpha (c_+^\alpha \psi_{1,\alpha}(t) - c_-^\alpha \psi_{1,\alpha}(T-t))\nabla_x w_\alpha(x)\right\|^2_{L^2(\lambda\otimes\mu)}$$
$$\leqslant 2\left(\|\nabla_x \partial_t u\|^2_{L^2(\lambda\otimes\mu)} + \sum_\alpha \alpha^2\|c_+^\alpha \psi_{1,\alpha}(t) + c_-^\alpha \psi_{1,\alpha}(T-t)\|^2_{L^2(\lambda)}\right)$$
$$\leqslant C\left(\|\nabla_x \partial_t u\|^2_{L^2(\lambda\otimes\mu)} + \sum_\alpha ((c_+^\alpha)^2 + (c_-^\alpha)^2)\alpha^2\|\psi_{1,\alpha}\|^2_{L^2(\lambda)}\right)$$
$$\overset{(37)}{\leqslant} C\left(\|\nabla_x \partial_t u\|^2_{L^2(\lambda\otimes\mu)} + \sum_\alpha ((c_+^\alpha)^2 + (c_-^\alpha)^2)\frac{(1-e^{-\alpha T})^3}{\alpha T}\right), \tag{39}$$

as well as

$$\|\partial_t \nabla_x \Phi\|^2_{L^2(\lambda\otimes\mu)} = \left\|\nabla_x \partial_t u + \sum_\alpha (c_+^\alpha \psi'_{2,\alpha}(t) - c_-^\alpha \psi'_{2,\alpha}(T-t))\nabla_x w_\alpha(x)\right\|^2_{L^2(\lambda\otimes\mu)}$$
$$\leqslant 2\left(\|\nabla_x \partial_t u\|^2_{L^2(\lambda\otimes\mu)} + \sum_\alpha \|c_+^\alpha \psi'_{2,\alpha}(t)\right.$$
$$\left. - c_-^\alpha \psi'_{2,\alpha}(T-t)\|^2_{L^2(\lambda)}\|\nabla_x w_\alpha\|^2_{L^2(\mu)}\right)$$
$$\leqslant C\left(\|\nabla_x \partial_t u\|^2_{L^2(\lambda\otimes\mu)} + \sum_\alpha \alpha^2((c_+^\alpha)^2 + (c_-^\alpha)^2)\|\psi'_{2,\alpha}\|^2_{L^2(\lambda)}\right)$$
$$\overset{(35)}{\leqslant} C\left(\|\nabla_x \partial_t u\|^2_{L^2(\lambda\otimes\mu)} + \sum_\alpha ((c_+^\alpha)^2 + (c_-^\alpha)^2)\frac{1}{\alpha T(1-e^{-\alpha T})}\right). \tag{40}$$

We finally treat the terms from $\nabla_x^2 \Phi$:

$$\|\nabla_x^2 \Phi\|^2_{L^2(\lambda\otimes\mu)} \overset{(19)}{\leqslant} C\left(\|\nabla_x^* \nabla_x \Phi\|^2_{L^2(\lambda\otimes\mu)} + R^2\|\nabla_x \Phi\|^2_{L^2(\lambda\otimes\mu)}\right)$$
$$\overset{(25),(38)}{\leqslant} C\left(\left\|f + \partial_{tt}u + c_0(t - \frac{T}{2}) + \sum_\alpha (c_+^\alpha \psi'_{1,\alpha}(t)\right.\right.$$
$$\left.\left. - c_-^\alpha \psi'_{1,\alpha}(T-t))w_\alpha(x)\right\|^2_{L^2(\lambda\otimes\mu)}\right.$$

$$+ R^2 \left(T^2 + \frac{1}{m(1 - e^{-\sqrt{m}T})^2} \right) \|f\|^2_{L^2(\lambda \otimes \mu)} \right)$$

$$\leqslant C \left(\|\partial_{tt} u\|^2_{L^2(\lambda \otimes \mu)} + c_0^2 T^2 + \sum_\alpha ((c_+^\alpha)^2 + (c_-^\alpha)^2) \|\psi_{1,\alpha}'\|^2_{L^2(\lambda)} \right.$$

$$+ \left(1 + R^2 T^2 + \frac{R^2}{m(1 - e^{-\sqrt{m}T})^2} \right) \|f\|^2_{L^2(\lambda \otimes \mu)} \right)$$

$$\overset{(36)}{\leqslant} C \left(\|\partial_{tt} u\|^2_{L^2(\lambda \otimes \mu)} + c_0^2 T^2 + \sum_\alpha ((c_+^\alpha)^2 + (c_-^\alpha)^2) \frac{1 - e^{-2\alpha T}}{\alpha T} \right.$$

$$+ \left(1 + R^2 T^2 + \frac{R^2}{m(1 - e^{-\sqrt{m}T})^2} \right) \|f\|^2_{L^2(\lambda \otimes \mu)} \right). \qquad (41)$$

Adding together (39), (40) and (41), we arrive at

$$\|\nabla_x \phi_0\|^2_{L^2(\lambda \otimes \mu)} + \|\overline{\nabla} \nabla_x \Phi\|^2_{L^2(\lambda \otimes \mu)}$$

$$\leqslant C \left(\|D^2 u\|^2_{L^2(\lambda \otimes \mu)} + c_0^2 T^2 + \sum_\alpha ((c_+^\alpha)^2 + (c_-^\alpha)^2) \frac{1}{\alpha T(1 - e^{-\alpha T})} \right.$$

$$+ \left(1 + R^2 T^2 + \frac{R^2}{m(1 - e^{-\sqrt{m}T})^2} \right) \|f\|^2_{L^2(\lambda \otimes \mu)} \right)$$

$$\overset{(30),(31)}{\leqslant} C \left(1 + R^2 T^2 + \frac{1}{(1 - e^{-\sqrt{m}T})^4} + \frac{R^2}{m(1 - e^{-\sqrt{m}T})^4} \right) \|f\|^2_{L^2(\lambda \otimes \mu)}.$$

We are now ready to prove the main results of the paper. The proof is essentially inspired from that of [1, Proof of Theorem 3]. In particular, to retrieve the $L^2(\lambda \otimes \mu; H_\kappa^{-1})$ norm, we need to construct a test function that is in $L^2(\lambda \otimes \mu; H_\kappa^1)$, which is highly related to the test functions constructed in Lemma 2.6. The differences of these two proofs are: (1) we choose the test functions explicitly $\xi_0 = 1$ and $\xi_i = v_i$, which are orthogonal to each other and have explicit expressions for up to fourth moments (in particular any first and third moments vanish); (2) Instead of using $\|\overline{\nabla} \Pi_v f\|_{H^{-1}(\lambda \otimes \mu)}$ as an intermediate step, we proceed as (42) and control the $L^2(\lambda \otimes \mu; H_\kappa^1)$ norm of another explicitly constructed function, in order to minimize the usage of Cauchy–Schwarz inequalities and track the dimension dependence of constants carefully.

Proof of Theorem 2. Without loss of generality, assume $(f)_{\lambda \otimes \rho_\infty} = 0$. which indicates $(\Pi_v f)_{\lambda \otimes \mu} = 0$. Therefore, we can take ϕ_0, Φ as in Lemma 2.6 with $\Pi_v f$ in place of f, so that $-\partial_t \phi_0 + \nabla_x^* \nabla_x \Phi = \Pi_v f$. The trick in our following step is to introduce v variable in the calculation. Notice, by Gaussianity, that

$$\int_{\mathbb{R}^d} v_i \, d\kappa(v) = 0, \qquad \int_{\mathbb{R}^d} v_i v_j \, d\kappa(v) = \delta_{i,j},$$

where $\delta_{i,j}$ is the Kronecker symbol which equals to 1 if $i = j$ and 0 otherwise. Thus,

$$
\begin{aligned}
\|\Pi_v f\|^2_{L^2(\lambda\otimes\mu)} &= \int_{I\times\mathbb{R}^d} \Pi_v f(-\partial_t\phi_0 + \nabla^*_x\nabla_x\Phi)\,\mathrm{d}\lambda(t)\,\mathrm{d}\mu(x) \\
&= \int_{I\times\mathbb{R}^{2d}} \Pi_v f(-\partial_t\phi_0 + v\cdot\nabla_x\phi_0 + v\cdot\partial_t\nabla_x\Phi \\
&\quad - v\cdot\nabla^2_x\Phi\cdot v + \nabla_x\Phi\cdot\nabla_x U)\,\mathrm{d}\lambda(t)\,\mathrm{d}\rho_\infty(x,v) \\
&= \int_{I\times\mathbb{R}^{2d}} f(-\partial_t\phi_0 + v\cdot\nabla_x\phi_0 + v\cdot\partial_t\nabla_x\Phi \\
&\quad - v\cdot\nabla^2_x\Phi\cdot v + \nabla_x\Phi\cdot\nabla_x U)\,\mathrm{d}\lambda(t)\,\mathrm{d}\rho_\infty(x,v) \\
&\quad + \int_{I\times\mathbb{R}^{2d}} (\partial_t\phi_0 - v\cdot\nabla_x\phi_0 - v\cdot\partial_t\nabla_x\Phi + v\cdot\nabla^2_x\Phi\cdot v \\
&\quad - \nabla_x\Phi\cdot\nabla_x U)(f - \Pi_v f)\,\mathrm{d}\lambda(t)\,\mathrm{d}\rho_\infty(x,v).
\end{aligned}
\tag{42}
$$

For the first integral on the right hand side, we use integration by parts, where it is important that the test functions $(\phi_0, \nabla_x\Phi)$ have Dirichlet boundary conditions in time:

$$
\begin{aligned}
&\int_{I\times\mathbb{R}^{2d}} f(-\partial_t\phi_0 + v\cdot\nabla_x\phi_0 + v\cdot\partial_t\nabla_x\Phi - v\cdot\nabla^2_x\Phi\cdot v + \nabla_x\Phi\cdot\nabla_x U)\,\mathrm{d}\lambda(t)\,\mathrm{d}\rho_\infty(x,v) \\
&= \int_{I\times\mathbb{R}^{2d}} (\partial_t f\,\phi_0 - \partial_t f(v\cdot\nabla_x\Phi) - \phi_0(v\cdot\nabla_x f) + f\phi_0(v\cdot\nabla_x U) \\
&\quad + (v\cdot\nabla_x f)(v\cdot\nabla_x\Phi) - f(v\cdot\nabla_x\Phi)(v\cdot\nabla_x U) + f\nabla_x\Phi\cdot\nabla_x U)\,\mathrm{d}\lambda(t)\,\mathrm{d}\rho_\infty(x,v) \\
&= \int_{I\times\mathbb{R}^{2d}} (\partial_t f\,\phi_0 - \partial_t f(v\cdot\nabla_x\Phi) - \phi_0(v\cdot\nabla_x f) + \phi_0(\nabla_v f\cdot\nabla_x U) \\
&\quad + (v\cdot\nabla_x f)(v\cdot\nabla_x\Phi) - \nabla_v\cdot((v\cdot\nabla_x\Phi)f\nabla_x U) + f\nabla_x\Phi\cdot\nabla_x U)\,\mathrm{d}\lambda(t)\,\mathrm{d}\rho_\infty(x,v) \\
&= \int_{I\times\mathbb{R}^{2d}} ((\partial_t f - v\cdot\nabla_x f + \nabla_x U\cdot\nabla_v f)(\phi_0 - v\cdot\nabla_x\Phi))\,\mathrm{d}\lambda(t)\,\mathrm{d}\rho_\infty(x,v) \\
&\leqslant \|\partial_t f - \mathcal{L}_{\mathrm{ham}} f\|_{L^2(\lambda\otimes\mu;H^{-1}_\kappa)}\|\phi_0 - v\cdot\nabla_x\Phi\|_{L^2(\lambda\otimes\mu;H^1_\kappa)}.
\end{aligned}
$$

We further estimate the term $\|\phi_0 - v\cdot\nabla_x\Phi\|_{L^2(\lambda\otimes\mu;H^1_\kappa)}$ by explicit integration, noticing (ϕ_0, Φ) do not depend on v so that explicit moments of v can be directly calculated:

$$
\begin{aligned}
\|\phi_0 - v\cdot\nabla_x\Phi\|^2_{L^2(\lambda\mu;H^1_\kappa)} &= \int_{I\times\mathbb{R}^d} \|\phi_0 - v\cdot\nabla_x\Phi\|^2_{H^1_\kappa}\,\mathrm{d}\lambda(t)\,\mathrm{d}\mu(x) \\
&= \int_{I\times\mathbb{R}^d} \Big(\|\phi_0 - v\cdot\nabla_x\Phi\|^2_{L^2_\kappa} \\
&\quad + \|\nabla_v(\phi_0 - v\cdot\nabla_x\Phi)\|^2_{L^2_\kappa}\Big)\,\mathrm{d}\lambda(t)\,\mathrm{d}\mu(x) \\
&= \int_{I\times\mathbb{R}^d} \Big(\int_{\mathbb{R}^d}(\phi_0 - v\cdot\nabla_x\Phi)^2\,\mathrm{d}\kappa(v) \\
&\quad + \int_{\mathbb{R}^d}|\nabla_x\Phi|^2\,\mathrm{d}\kappa(v)\Big)\,\mathrm{d}\lambda(t)\,\mathrm{d}\mu(x)
\end{aligned}
$$

$$= \int_{I \times \mathbb{R}^d} \left(\phi_0^2 + 2|\nabla_x \Phi|^2 \right) d\lambda(t) d\mu(x)$$

$$\overset{(26)}{\leqslant} C \left(\frac{1}{m(1 - e^{-\sqrt{m}T})^2} + T^2 \right) \|\Pi_v f\|_{L^2(\lambda \otimes \mu)}^2.$$

For the second integral in (42), we estimate again by explicit expansion in v, which is possible since we have explicit up to fourth moments of v:

$$\|\partial_t \phi_0 - v \cdot \nabla_x \phi_0 - v \cdot \partial_t \nabla_x \Phi + v \cdot \nabla_x^2 \Phi \cdot v - \nabla_x \Phi \cdot \nabla_x U\|_{L^2(\lambda \otimes \rho_\infty)}^2$$

$$= \int_{I \times \mathbb{R}^{2d}} (\partial_t \phi_0 - v \cdot \nabla_x \phi_0 - v \cdot \partial_t \nabla_x \Phi$$

$$+ v \cdot \nabla_x^2 \Phi \cdot v - \nabla_x \Phi \cdot \nabla_x U)^2 \, d\lambda(t) \, d\rho_\infty(x, v)$$

$$= \int_{I \times \mathbb{R}^{2d}} \Big((\partial_t \phi_0 - \nabla_x \Phi \cdot \nabla_x U)^2 - 2(\partial_t \phi_0 - \nabla_x \Phi \cdot \nabla_x U)(v \cdot \nabla_x \phi_0)$$

$$- 2(\partial_t \phi_0 - \nabla_x \Phi \cdot \nabla_x U)(v \cdot \partial_t \nabla_x \Phi) + (v \cdot \nabla_x \phi_0)^2 + (v \cdot \partial_t \nabla_x \Phi)^2$$

$$+ 2(\partial_t \phi_0 - \nabla_x \Phi \cdot \nabla_x U)v \cdot \nabla_x^2 \Phi \cdot v + 2(v \cdot \partial_t \nabla_x \Phi)(v \cdot \nabla_x \phi_0)$$

$$+ (v \cdot \nabla_x^2 \Phi \cdot v)^2 - 2(v \cdot \partial_t \nabla_x \Phi)(v \cdot \nabla_x^2 \Phi \cdot v)$$

$$-2(v \cdot \partial_{x_k} \phi_0)(v \cdot \nabla_x^2 \Phi \cdot v) \Big) \, d\lambda(t) \, d\rho_\infty(x, v)$$

$$= \int_{I \times \mathbb{R}^{2d}} \Big((\partial_t \phi_0 - \nabla_x \Phi \cdot \nabla_x U)^2$$

$$+ \sum_i v_i^2 \left((\partial_{x_i} \phi_0)^2 + (\partial_t \partial_{x_i} \Phi)^2 + 2\partial_{x_i} \phi_0 \partial_t \partial_{x_i} \Phi \right)$$

$$+ 2(\partial_t \phi_0 - \nabla_x \Phi \cdot \nabla_x U) \sum_i v_i^2 \partial_{x_i x_i} \Phi$$

$$+ \sum_i v_i^4 (\partial_{x_i x_i} \Phi_i)^2 + 2 \sum_{i \neq j} v_i^2 v_j^2 (\partial_{x_i x_j} \Phi)^2$$

$$+ \sum_{i \neq j} v_i^2 v_j^2 \partial_{x_i x_i} \Phi \partial_{x_j x_j} \Phi \Bigg) \, d\lambda(t) \, d\rho_\infty(x, v)$$

$$= \int_{I \times \mathbb{R}^d} \Big((\partial_t \phi_0 - \nabla_x \Phi \cdot \nabla_x U)^2 + |\nabla_x \phi_0 + \partial_t \nabla_x \Phi|^2 + 3 \sum_i (\partial_{x_i x_i} \Phi)^2$$

$$+ 2 \sum_{i \neq j} (\partial_{x_i x_j} \Phi)^2 + 2(\partial_t \phi_0 - \nabla_x \Phi \cdot \nabla_x U)\Delta_x \Phi + \sum_{i \neq j} \partial_{x_i x_i} \Phi \partial_{x_j x_j} \Phi \Bigg) \, d\lambda(t) \, d\mu(x)$$

$$\leqslant \int_{I \times \mathbb{R}^d} \Big((\partial_t \phi_0 - \nabla_x \Phi \cdot \nabla_x U + \Delta_x \Phi)^2 + 2|\nabla_x \phi_0|^2 + 2|\overline{\nabla} \nabla_x \Phi|^2 \Big) \, d\lambda(t) \, d\mu(x)$$

$$\overset{(28)}{=} \|\Pi_v f\|_{L^2(\lambda \otimes \mu)}^2 + 2\|\nabla_x \phi_0\|_{L^2(\lambda \otimes \mu)}^2 + 2\|\overline{\nabla} \nabla_x \Phi\|_{L^2(\lambda \otimes \mu)}^2$$

$$\overset{(27)}{\leqslant} C \left(1 + R^2 T^2 + \frac{1}{(1 - e^{-\sqrt{m}T})^4} + \frac{R^2}{m(1 - e^{-\sqrt{m}T})^4} \right) \|\Pi_v f\|_{L^2(\lambda \otimes \mu)}^2.$$

Combining the above estimates, we arrive at

$$\|\Pi_v f\|_{L^2(\lambda\otimes\mu)}^2 \leqslant \|\partial_t f - \mathcal{L}_{\text{ham}} f\|_{L^2(\lambda\otimes\mu; H_\kappa^{-1})} \|\phi_0 - v\cdot\nabla_x\Phi\|_{L^2(\lambda\otimes\mu; H_\kappa^1)}$$

$$+ \|\partial_t\phi_0 - v\cdot\nabla_x\phi_0 - v\cdot\partial_t\nabla_x\Phi + v\cdot\nabla_x^2\Phi\cdot v$$

$$- \nabla_x\Phi\cdot\nabla_x U\|_{L^2(\lambda\otimes\rho_\infty)} \|f - \Pi_v f\|_{L^2(\lambda\otimes\rho_\infty)}$$

$$\leqslant C\left(\left(\frac{1}{\sqrt{m}(1-e^{-\sqrt{m}T})} + T\right)\|\partial_t f - \mathcal{L}_{\text{ham}} f\|_{L^2(\lambda\otimes\mu; H_\kappa^{-1})}\|\Pi_v f\|_{L^2(\lambda\otimes\mu)}\right.$$

$$+ \left(1 + RT + \frac{1}{(1-e^{-\sqrt{m}T})^2} + \frac{R}{\sqrt{m}(1-e^{-\sqrt{m}T})^2}\right)$$

$$\left. \|(Id - \Pi_v)f\|_{L^2(\lambda\otimes\rho_\infty)}\|\Pi_v f\|_{L^2(\lambda\otimes\mu)}\right).$$

Finally

$$\|f\|_{L^2(\lambda\otimes\rho_\infty)} \leqslant \|(Id - \Pi_v)f\|_{L^2(\lambda\otimes\rho_\infty)} + \|\Pi_v f\|_{L^2(\lambda\otimes\mu)}$$

$$\leqslant C\left(\left(\frac{1}{\sqrt{m}(1-e^{-\sqrt{m}T})} + T\right)\|\partial_t f - \mathcal{L}_{\text{ham}} f\|_{L^2(\lambda\otimes\mu; H_\kappa^{-1})}\right.$$

$$+ \left.\left(1 + RT + \frac{1}{(1-e^{-\sqrt{m}T})^2} + \frac{R}{\sqrt{m}(1-e^{-\sqrt{m}T})^2}\right)\|(Id - \Pi_v)f\|_{L^2(\lambda\otimes\rho_\infty)}\right),$$

as claimed.   □

With Theorem 2, we are now able to prove exponential relaxation to equilibrium claimed in Theorem 1, which essentially follows from a standard energy estimate.

Proof of Theorem 1. We first notice that the solution $f \in H_{hyp}^1((0, T)\otimes\mu)$ for all $T > 0$. Indeed, as long as $f_0 \in L^2(\mu; H_\kappa^1)$, we have $f(t, \cdot, \cdot) \in L^2(\mu; H_\kappa^1)$ for any $t > 0$ (see for example [54, Theorem 35]), and hence $\partial_t f - \mathcal{L}_{\text{ham}} f = -\gamma\nabla_v^*\nabla_v f \in L^2(\lambda\otimes\mu; H_\kappa^{-1})$. We also have that (12) implies

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} f(t, x, v)\,d\rho_\infty(x, v) = 0$$

for all $t \in (0, T)$. This follows from

$$\frac{d}{dt}\int_{\mathbb{R}^d\times\mathbb{R}^d} f(t, x, v)\,d\rho_\infty(x, v) = 0,$$

using the equation (4) and integration by parts.

For every $0 < s < t$, we have the typical energy estimate (hereafter we use $L^2((s, t)\otimes\rho_\infty)$ to denote $L^2(\lambda_{(s,t)}\otimes\rho_\infty)$):

$$\|f(t, \cdot)\|_{L^2(\rho_\infty)}^2 - \|f(s, \cdot)\|_{L^2(\rho_\infty)}^2 = -2\gamma\|\nabla_v f\|_{L^2((s,t)\otimes\rho_\infty)}^2. \tag{43}$$

In particular,

$$\text{the mapping } t \mapsto \|f(t, \cdot)\|_{L^2(\rho_\infty)}^2 \text{ is nonincreasing.} \tag{44}$$

Since by equation (4),

$$-\gamma\nabla_v^*\nabla_v f = \partial_t f - \mathcal{L}_{\text{ham}} f,$$

we have

$$\|\partial_t f - \mathcal{L}_{\text{ham}} f\|_{L^2((s,t)\otimes\mu, H_\kappa^{-1})} = \gamma \|\nabla_v^* \nabla_v f\|_{L^2((s,t)\otimes\mu, H_\kappa^{-1})} \leqslant \gamma \|\nabla_v f\|_{L^2((s,t)\otimes\rho_\infty)}.$$

Now fix T to be the length of the time interval. Denote $b_1 = C(\frac{1}{\sqrt{m}(1-e^{-\sqrt{m}T})} + T)$ and $b_2 = C(1 + RT + \frac{1}{(1-e^{-\sqrt{m}T})^2} + \frac{R}{\sqrt{m}(1-e^{-\sqrt{m}T})^2})$, and thus by Theorem 2, (43) and (44), and Gaussian Poincaré inequality

$$\|(Id - \Pi_v) f\|_{L^2(\lambda\otimes\rho_\infty)} \leqslant \|\nabla_v f\|_{L^2(\lambda\otimes\rho_\infty)},$$

we have for time stamps $t_k = kT$

$$
\begin{aligned}
&\|f(t_k, \cdot)\|_{L^2(\rho_\infty)}^2 - \|f(t_{k-1}, \cdot)\|_{L^2(\rho_\infty)}^2 \\
&\leqslant -\frac{2\gamma}{(b_1\gamma + b_2)^2} \Big(b_2 \|\nabla_v f\|_{L^2((t_{k-1}, t_k)\otimes\rho_\infty)} + b_1 \|\partial_t f \\
&\qquad - \mathcal{L}_{\text{ham}} f\|_{L^2((t_{k-1}, t_k)\otimes\mu, H_\kappa^{-1})} \Big)^2 \\
&\leqslant -\frac{2\gamma}{(b_1\gamma + b_2)^2} \Big(b_2 \|(Id - \Pi_v) f\|_{L^2((t_{k-1}, t_k)\otimes\rho_\infty)} + b_1 \|\partial_t f \\
&\qquad - \mathcal{L}_{\text{ham}} f\|_{L^2((t_{k-1}, t_k)\otimes\mu, H_\kappa^{-1})} \Big)^2 \\
&\leqslant -\frac{2\gamma}{(b_1\gamma + b_2)^2} \|f\|_{L^2((t_{k-1}, t_k)\otimes\rho_\infty)}^2 \\
&\leqslant -\frac{2\gamma T}{(b_1\gamma + b_2)^2} \|f(t_k, \cdot)\|_{L^2(\rho_\infty)}^2.
\end{aligned}
$$

Now for any $t > 0$, we pick the integer k satisfying $t_k \leqslant t < t_{k+1}$, so that $\|f(t, \cdot)\|_{L^2(\rho_\infty)} \leqslant \|f(t_k, \cdot)\|_{L^2(\rho_\infty)}$. Applying above inequality iteratively and using the monoticity (44), we obtain

$$
\begin{aligned}
\|f(t, \cdot)\|_{L^2(\rho_\infty)}^2 &\leqslant \left(1 + \frac{2\gamma T}{(b_1\gamma + b_2)^2}\right)^{-k} \|f_0\|_{L^2(\rho_\infty)}^2 \\
&\leqslant \left(1 + \frac{2\gamma T}{(b_1\gamma + b_2)^2}\right)^{-\frac{t}{T}+1} \|f_0\|_{L^2(\rho_\infty)}^2 \\
&= \left(1 + \frac{2\gamma T}{(b_1\gamma + b_2)^2}\right) \exp\left(-\frac{t}{T} \log\left(1 + \frac{2\gamma T}{(b_1\gamma + b_2)^2}\right)\right) \|f_0\|_{L^2(\rho_\infty)}^2.
\end{aligned}
$$

The prefactor

$$1 + \frac{2\gamma T}{(b_1\gamma + b_2)^2} \leqslant C\left(1 + \frac{\gamma T}{\left(\frac{\gamma}{\sqrt{m}} + \gamma T + 1\right)^2}\right)$$

is bounded above by a constant. Using $\log(1 + x) \geqslant \frac{1}{C}x$ for $x \in [0, \frac{1}{C}]$ for some universal constant C, and then pick $T = \frac{1}{\sqrt{m}}$, this yields exponential decay with rate

$$\nu \geqslant C\frac{\gamma}{(b_1\gamma + b_2)^2} \geqslant C\frac{\gamma m}{(\gamma + R + \sqrt{m})^2},$$

which is precisely (13).   □

Data Availability Statement This manuscript has no associated data.

Declarations
Conflict of interest Statement The authors have no conflict of interest.

Appendix A. The Decay Rate for Isotropic Quadratic Potential

For isotropic quadratic potential, an explicit expression for the spectral gap of $\mathcal{L}$ is available (thus also the decay rate in (7)). Note that while the result is stated for $d = 1$, it trivially extends to arbitrary dimension for isotropic quadratic potential as different coordinates are independent. The spectrum is also explicitly known for $V = 0$ and $x \in \mathbb{T}^d$ on a torus, see [35].

Theorem 3. ([46, (10.83)], [42, Theorem 3.1]) *When* $U(x) = \frac{m}{2}|x|^2$, $d = 1$, *the spectrum of the operator* $-\mathcal{L}$ *is given by*

$$\left\{\lambda_{i,j} := \frac{\gamma}{2}(i + j) + \frac{\sqrt{\gamma^2 - 4m}}{2}(i - j), \quad i, j = 0, 1, 2, \cdots \right\}.$$

Let λ_{exact} be the spectral gap for the real component of $\{\lambda_{i,j}\}_{i,j \geqslant 0}$. Notice that the spectral gap is always achieved when $i = 0$ and $j = 1$, thus

$$\lambda_{\text{exact}} = \mathfrak{Re}\left(\frac{\gamma}{2} - \frac{\sqrt{\gamma^2 - 4m}}{2}\right). \tag{45}$$

Corollary A.1. *For any dimension d, for isotropic potential* $U(x) = \frac{m}{2}|x|^2$, *(7) holds with the decay rate* λ_{exact}.

Appendix B. The DMS Hypocoercive Estimation

In this section, we will revisit the decay rate by DMS estimation [18,19], adapted and summarized for underdamped Langevin equation in [47, Sec. 2]. In the first part of this section, we will review the main result based on [47]; in addition, we will provide a new estimate of the operator norm of $\|\mathcal{A}\mathcal{L}_{\mathrm{ham}}(1 - \Pi_v)\|_{L^2(\rho_\infty) \to L^2(\rho_\infty)}$, which leads into a more explicit expression of the decay rate. In the second part, we will present the asymptotic analysis of the decay rate with respect to m and γ, under the assumption that $\nabla_x^2 U \geqslant -2\,\mathrm{Id}$.

B.1. Revisiting the DMS Hypocoercive Estimation in $L^2(\rho_\infty)$

Let us first define an operator

$$\mathcal{A} = \left(1 + (\mathcal{L}_{\mathrm{ham}}\Pi_v)^*(\mathcal{L}_{\mathrm{ham}}\Pi_v)\right)^{-1} (\mathcal{L}_{\mathrm{ham}}\Pi_v)^* \tag{46}$$

and a Lyapunov function E for $\phi(x, v)$ by

$$\mathsf{E}(\phi) = \frac{1}{2} \|\phi\|_{L^2(\rho_\infty)}^2 - \epsilon \, (\mathcal{A}\phi, \phi)_{L^2(\rho_\infty)}, \tag{47}$$

where $\epsilon \in (-1, 1)$ is some quantity depending on $\mathcal{L}$, to be specified below. The functional E is equivalent to $L^2(\rho_\infty)$ norm in the following sense (see e.g., [47, Eq. (17)]),

$$\frac{1 - |\epsilon|}{2} \|\phi\|_{L^2(\rho_\infty)}^2 \leqslant \mathsf{E}(\phi) \leqslant \frac{1 + |\epsilon|}{2} \|\phi\|_{L^2(\rho_\infty)}^2. \tag{48}$$

Theorem 4. (See [47, Theorem 1]) *Assume that the Poincaré inequality* (10) *holds and there exists* $\mathsf{R}_{ham} < \infty$ *such that*

$$\|\mathcal{A}\mathcal{L}_{ham}(1 - \Pi_v)\|_{L^2(\rho_\infty) \to L^2(\rho_\infty)} \leqslant \mathsf{R}_{ham}. \tag{49}$$

Suppose $\epsilon \in (-1, 1)$ *is chosen such that* $\lambda_{DMS} = \lambda_{DMS}(\gamma, m, \mathsf{R}_{ham}, \epsilon) > 0$, *where*

$$\lambda_{DMS} := \frac{\gamma - \frac{\epsilon}{1+m} - \sqrt{\epsilon^2(\mathsf{R}_{ham} + \frac{\gamma}{2})^2 + \left(\gamma - \frac{2m+1}{m+1}\epsilon\right)^2}}{2(1 + |\epsilon|)}. \tag{50}$$

Then for any solution $f(t, x, v)$ *of* (4) *with* $\int f_0 \, \mathrm{d}\rho_\infty = 0$, *we have*

$$\|f(t, \cdot, \cdot)\|_{L^2(\rho_\infty)} \leqslant \sqrt{\frac{1 + |\epsilon|}{1 - |\epsilon|}} \, \|f_0\|_{L^2(\rho_\infty)} \, e^{-\lambda_{DMS} \, t}.$$

Notice that when $\epsilon = 0$, the rate $\lambda_{\text{DMS}} = 0$, which reduces to the conclusion that $\|f(t, \cdot, \cdot)\|_{L^2(\rho_\infty)}$ is non-increasing in time t. The existence of R_{ham} has been studied under fairly general assumptions on the potential $U(x)$ in [19, Sec. 2]. In the Proposition B.1 below, we provide a simpler estimation of R_{ham} only under the assumption of lower bound on Hessian; see the Appendix B.3 for its proof. The first part of the proof is the same as [19, Lemma 4]; the simplicity in our approach comes from the application of *Bochner's formula*. It is interesting to observe that R_{ham} does not depend on m when U is an isotropic quadratic potential.

Proposition B.1. *Assume there exists* $K \in \mathbb{R}$ *such that* $\nabla_x^2 U \geqslant -K \,\text{Id}$ *for all* $x \in \mathbb{R}^d$, *then we can choose*

$$\mathsf{R}_{ham} = \sqrt{\max\{K, 2\}}. \tag{51}$$

such that (49) is satisfied.

For the isotropic case $U(x) = \frac{m}{2}|x|^2$, *we have*

$$\|\mathcal{AL}_{ham}(1 - \Pi_v)\|_{L^2(\rho_\infty) \to L^2(\rho_\infty)} = \sqrt{2}.$$

Thus the optimal choice of R_{ham} *is* $\sqrt{2}$ *and (51) is tight in this case.*

As an immediate consequence, if it holds that $\nabla_x^2 U \geqslant -2 \,\text{Id}$, we can take $\mathsf{R}_{\text{ham}} = \sqrt{2}$, which is tight for the isotropic case.

B.2. Asymptotic Analysis of the Decay Rate

In this subsection, we shall assume that $\nabla_x^2 U \geqslant -2 \,\text{Id}$, thus we can choose $\mathsf{R}_{\text{ham}} = \sqrt{2}$, according to the Proposition B.1. To remove the dependence on the parameter ϵ and to find the optimal decay rate, let us introduce

$$\Lambda_{\text{DMS}}(\gamma, m) := \sup_{\epsilon \in (-1,1)} \lambda_{\text{DMS}}(\gamma, m, \sqrt{2}, \epsilon)$$

$$= \sup_{\epsilon \in (-1,1)} \frac{\gamma - \frac{\epsilon}{1+m} - \sqrt{\epsilon^2(\sqrt{2} + \frac{\gamma}{2})^2 + \left(\gamma - \frac{2m+1}{m+1}\epsilon\right)^2}}{2(1 + |\epsilon|)}, \tag{52}$$

provided that the supremum is not achieved at the boundary i.e., $\epsilon = 1^-$ or $\epsilon = (-1)^+$. Observe that

- When $\epsilon = 0$, $\lambda_{\text{DMS}}(\gamma, m, \sqrt{2}, 0) = 0$;
- When $\epsilon = (-1)^+$, $\lambda_{\text{DMS}}(\gamma, m, \sqrt{2}, (-1)^+) < 0$.

Therefore, the supremum can only be achieved at $\epsilon = 1^-$, or the critical points of the expression on the right hand side of (52). In general, it is hard to obtain a simple explicit expression of $\Lambda_{\text{DMS}}(\gamma, m)$. Therefore, we shall consider the following asymptotic regions:

Proposition B.2. *(i) For fixed $m = O(1)$, we have*

$$\Lambda_{DMS}(\gamma, m) = \begin{cases} \left(\dfrac{-(1+m)\sqrt{3m^2 + 4m + 1} + 3m^2 + 3m + 1}{6m^2 + 8m + 3} \right) \gamma + O(\gamma^2), & \text{when } \gamma \to 0; \\ \dfrac{4m^2}{(1+m)^2} \gamma^{-1} + O(\gamma^{-2}), & \text{when } \gamma \to \infty. \end{cases} \tag{53}$$

(ii) *Consider coupled asymptotic regime $\gamma = b\sqrt{m}$ (or equivalently $m = (\gamma/b)^2$) for some $b = O(1)$, we have*

$$\Lambda_{DMS}(\gamma, m) = \begin{cases} \dfrac{\gamma^5}{2b^4} + O(\gamma^6), & \text{when } \gamma \to 0; \\ \dfrac{4}{\gamma} + O(\gamma^{-2}), & \text{when } \gamma \to \infty. \end{cases} \tag{54}$$

The proof can be found in Appendix B.3. The scaling in the first case is already known in e.g., [17,26,47]; in the above proposition, we simply explicitly calculate the leading order term. The second case is relevant when we choose γ to optimize the convergence rate according to m and for the regime $m \to 0$.

B.3. Proofs of the Propositions in Appendix

Proof of Proposition B.1. We first consider the case that Hessian is bounded from below. It is equivalent to consider the operator norm of

$$-(1 - \Pi_v)\mathcal{L}_{\text{ham}}\mathcal{A}^* = -(1 - \Pi_v)\mathcal{L}_{\text{ham}}^2 \Pi_v \left(1 + (\mathcal{L}_{\text{ham}}\Pi_v)^*(\mathcal{L}_{\text{ham}}\Pi_v)\right)^{-1}.$$

Notice that this operator is supported on $\text{Ran}(\Pi_v)$ from the observation that $\mathcal{A} = \Pi_v \mathcal{A}$, it is then equivalent to find the smallest R_{ham} such that for any $\phi(x, v)$ with $\Pi_v \phi = \phi$ (i.e., $\phi(x, v) \equiv \phi(x)$ is a function of x only), we have

$$\left\| -(1 - \Pi_v)\mathcal{L}_{\text{ham}}\mathcal{A}^* \phi \right\|_{L^2(\rho_\infty)} \leqslant \mathsf{R}_{\text{ham}} \|\phi\|_{L^2(\rho_\infty)} = \mathsf{R}_{\text{ham}} \|\phi\|_{L^2(\mu)}. \tag{55}$$

Given such a function ϕ with $\Pi_v \phi = \phi$, define

$$\varphi := \left(1 + (\mathcal{L}_{\text{ham}}\Pi_v)^*(\mathcal{L}_{\text{ham}}\Pi_v)\right)^{-1} \phi.$$

It is easy to check that $\Pi_v \varphi = \varphi$. By simplifying the above equation with (5) and (9),

$$\phi(x) = \varphi(x) - \Delta_x \varphi(x) + \nabla_x U(x) \cdot \nabla_x \varphi = \varphi(x) + \nabla_x^* \nabla_x \varphi(x). \tag{56}$$

Furthermore, by some straightforward calculation, we have

$$-(1 - \Pi_v)\mathcal{L}_{\text{ham}}\mathcal{A}^* \phi = -(1 - \Pi_v)\mathcal{L}_{\text{ham}}^2 \Pi_v \varphi = -\sum_{i,j}(v_i v_j - \delta_{i,j})\partial_{x_i, x_j}\varphi.$$

Thus

$$\left\| -(1 - \Pi_v)\mathcal{L}_{\text{ham}}\mathcal{A}^* \phi \right\|_{L^2(\rho_\infty)}^2 = \int \left(\sum_{i,j}(v_i v_j - \delta_{i,j})\partial_{x_i, x_j}\varphi \right)^2 \, d\rho_\infty$$

$$= 2 \sum_{i,j} \int \left(\partial_{x_i, x_j} \varphi \right)^2 \, d\mu.$$

Then, by Bochner's formula,

$$\left\| -(1 - \Pi_v) \mathcal{L}_{\text{ham}} \mathcal{A}^*(\phi) \right\|_{L^2(\rho_\infty)}^2$$

$$= 2 \int \nabla_x \varphi \cdot \nabla_x \nabla_x^* \nabla_x \varphi - \nabla_x \varphi \cdot \nabla_x^2 U \nabla_x \varphi - \nabla_x^* \nabla_x \left(\frac{|\nabla_x \varphi|^2}{2} \right) \, d\mu$$

$$= 2 \int |\nabla_x^* \nabla_x \varphi|^2 - \nabla_x \varphi \cdot \nabla_x^2 U \nabla_x \varphi \, d\mu$$

$$\leqslant 2 \left(\int |\nabla_x^* \nabla_x \varphi|^2 \, d\mu + K \int |\nabla_x \varphi|^2 \, d\mu \right)$$

$$\leqslant \max \{K, 2\} \left(\int |\nabla_x^* \nabla_x \varphi|^2 \, d\mu + 2 \int |\nabla_x \varphi|^2 \, d\mu \right).$$

From (56), we have

$$\|\phi\|_{L^2(\mu)}^2 = \int \varphi^2 + 2\varphi \, \nabla_x^* \nabla_x \varphi + \left| \nabla_x^* \nabla_x \varphi \right|^2 \, d\mu$$

$$\geqslant 2 \int |\nabla_x \varphi|^2 \, d\mu + \int |\nabla_x^* \nabla_x \varphi|^2 \, d\mu.$$

By combining the last two equations,

$$\left\| -(1 - \Pi_v) \mathcal{L}_{\text{ham}} \mathcal{A}^*(\phi) \right\|_{L^2(\rho_\infty)}^2 \leqslant \max\{K, 2\} \|\phi\|_{L^2(\mu)}^2,$$

which yields (51).

We now consider the isotropic case. Recall that the operator norm of $\mathcal{A} \mathcal{L}_{\text{ham}} (1 - \Pi_v)$ is the smallest R_{ham} such that (55) holds. Let us consider the elliptic PDE (56). By the choice $U(x) = \frac{m}{2} |x|^2$,

$$\phi(x) = \left(1 + m(x - \frac{1}{m} \nabla_x) \cdot \nabla_x \right) \varphi(x).$$

Then by rescaling the variable $x = \frac{y}{\sqrt{m}}$ and rescaling the functions $\overline{\phi}(y) := \phi(x) = \phi\left(\frac{y}{\sqrt{m}} \right)$, $\overline{\varphi}(y) := \varphi(x) = \varphi\left(\frac{y}{\sqrt{m}} \right)$, we have

$$\overline{\phi}(y) = \left(1 + m(y - \nabla_y) \cdot \nabla_y \right) \overline{\varphi}(y). \tag{57}$$

In addition, by rewriting (55), we need to find the smallest R_{ham} such that

$$2m^2 \sum_{i,j} \int \left| \partial_{y_i, y_j} \overline{\varphi}(y) \right|^2 \, e^{-\frac{|y|^2}{2}} \, dy \leqslant \mathsf{R}_{\text{ham}}^2 \int \overline{\phi}(y)^2 e^{-\frac{|y|^2}{2}} \, dy. \tag{58}$$

Next, let us expand the last equation by probabilists' Hermite polynomials $H_k(z) := (z - \frac{d}{dz})^k \cdot 1$ for integers $k \geqslant 0$. Recall two important properties

$$H_k'(z) = k H_{k-1}(z), \qquad \frac{1}{\sqrt{2\pi}} \int H_j(z) H_k(z) e^{-\frac{z^2}{2}} \, dz = k! \, \delta_{j,k}.$$

Given $\boldsymbol{n} = (n_1, n_2, \cdots, n_d)$, define

$$H_{\boldsymbol{n}}(y) := H_{n_1}(y_1) H_{n_2}(y_2) \cdots H_{n_d}(y_d).$$

By the above properties, it is easy to show that if $\overline{\varphi} = H_{\boldsymbol{n}}$, then $\overline{\phi} = N_{\boldsymbol{n}} H_{\boldsymbol{n}}$, where $N_{\boldsymbol{n}} := 1 + m \sum_i n_i$. Thus if $\overline{\varphi}(y) = \sum_{\boldsymbol{n}} a_{\boldsymbol{n}} H_{\boldsymbol{n}}$, then we have $\overline{\phi} = \sum_{\boldsymbol{n}} a_{\boldsymbol{n}} N_{\boldsymbol{n}} H_{\boldsymbol{n}}$. By such an expansion, (58) can be rewritten as

$$2m^2 \sum_{i,j} \sum_{\boldsymbol{n}} a_{\boldsymbol{n}}^2 (n_i n_j - \delta_{i,j} n_i) \prod_{k=1}^{d} n_k! \leqslant \mathsf{R}_{\mathrm{ham}}^2 \sum_{\boldsymbol{n}} a_{\boldsymbol{n}}^2 N_{\boldsymbol{n}}^2 \prod_{k=1}^{d} n_k!$$

Then finding the operator norm of $\mathcal{A}\mathcal{L}_{\mathrm{ham}}(1 - \Pi_v)$ is equivalent to finding the smallest $\mathsf{R}_{\mathrm{ham}}$ such that for any $\boldsymbol{n}$, one has

$$\sum_{i,j} (n_i n_j - \delta_{i,j} n_i) \leqslant \frac{\mathsf{R}_{\mathrm{ham}}^2}{2m^2} N_{\boldsymbol{n}}^2 \equiv \frac{\mathsf{R}_{\mathrm{ham}}^2}{2m^2} \left(1 + m \sum_i n_i\right)^2.$$

When $n_1 \to \infty$ and $n_2, n_3, \cdots, n_d = 0$, we know that $\frac{\mathsf{R}_{\mathrm{ham}}^2}{2} \geqslant 1$. Also observe that

$$\sum_{i,j} (n_i n_j - \delta_{i,j} n_i) \leqslant \left(\sum_i n_i\right)^2 = \frac{1}{m^2} \left(m \sum_i n_i\right)^2 \leqslant \frac{1}{m^2} \left(1 + m \sum_i n_i\right)^2.$$

Therefore, $\frac{\mathsf{R}_{\mathrm{ham}}^2}{2} = 1$ is sufficient.

In summary, $\|\mathcal{A}\mathcal{L}_{\mathrm{ham}}(1 - \Pi_v)\|_{L^2(\rho_\infty) \to L^2(\rho_\infty)} = \sqrt{2}$ and the optimal choice of $\mathsf{R}_{\mathrm{ham}}$ is $\sqrt{2}$. $\quad\square$

Proof of Proposition B.2. We used Maple software to help verify the asymptotic expansion.
Part (i): $m = O(1)$.

- **(when $\gamma \to 0$).** Via asymptotic expansion, we have

$$\lambda_{\mathrm{DMS}}(\gamma, m, \sqrt{2}, 1^-) = -\frac{1 + \sqrt{6m^2 + 8m + 3}}{4(1 + m)} + O(\gamma) < 0.$$

Thus the supremum is not obtained at $\epsilon = 1^-$. Then let us consider critical points within the domain $(-1, 1)$, whose asymptotic expansions are

$$\epsilon_\pm = \frac{(6m^2 + 5m + 1 \pm \sqrt{3m^2 + 4m + 1})(1 + m)}{18m^3 + 30m^2 + 17m + 3} \gamma + O(\gamma^2) > 0.$$

After comparison, the larger decay rate is obtained at ϵ_- with the value in (53).

- **(when $\gamma \to \infty$)**. Similarly, via asymptotic expansion, we have

$$\lambda_{\text{DMS}}(\gamma, m, \sqrt{2}, 1^-) = -\frac{\frac{\sqrt{5}}{2} - 1}{4}\gamma + O(1) < 0.$$

Thus we need to consider the critical points. It turns out, there is only one critical point within the domain $(-1, 1)$, which is $\varepsilon = \frac{8m}{1+m}\gamma^{-1} + O(\gamma^{-2})$ with the decay rate in (53).

Part (ii): $\gamma = b\sqrt{m}$ with $b = O(1)$.

- **(when $\gamma \to 0$)**. Via asymptotic expansion, one could check that

$$\lambda_{\text{DMS}}(\gamma, m = (\gamma/b)^2, \sqrt{2}, 1^-) = -\frac{1 + \sqrt{3}}{4} + O(\gamma) < 0.$$

Thus, we only need to consider the decay rate at critical points, which are given by

$$\epsilon_1 = \frac{\gamma^3}{b^2} + O(\gamma^4), \qquad \epsilon_2 = \frac{2}{3}\gamma + O(\gamma^2).$$

and the associated decay rates are

$$\lambda_{\text{DMS}}(\gamma, m = (\gamma/b)^2, \sqrt{2}, \epsilon_1) = \frac{\gamma^5}{2b^4} + O(\gamma^6) > 0;$$

$$\lambda_{\text{DMS}}(\gamma, m = (\gamma/b)^2, \sqrt{2}, \epsilon_2) = -\frac{1}{3}\gamma + O(\gamma^2) < 0.$$

Therefore, the optimal decay rate is obtained at ϵ_1, which gives (54).

- **(when $\gamma \to \infty$)**. Via asymptotic expansion, one could obtain

$$\lambda_{\text{DMS}}(\gamma, m = (\gamma/b)^2, \sqrt{2}, 1^-) = -\frac{\sqrt{5} - 2}{8}\gamma + O(1) < 0.$$

Thus the supremum in (52) cannot be obtained at $\epsilon = 1^-$. Then, let us look at the critical points. It turns out there is only one within the interval $(-1, 1)$, which is $\epsilon_1 = \frac{8}{\gamma} + O(\gamma^{-2})$. The optimal decay rate must be achieved at ϵ_1, with the expression given in (54). □

References

1. ALBRITTON, D., ARMSTRONG, S., MOURRAT, J.-C., NOVACK, M.: *Variational methods for the kinetic Fokker–Planck equation*, arXiv preprint arXiv:1902.04037, 2019
2. ANDRIEU, C., DURMUS, A., NÜSKEN, N., ROUSSEL, J.: Hypocoercivity of piecewise deterministic Markov process-Monte Carlo. *Ann. Appl. Probab.* **31**(5), 2478–2517, 2021

3. ARMSTRONG, S.: *Answer to "Elliptic regularity with Gibbs measure satisfying Bakry–Emery condition"*, MathOverflow. https://mathoverflow.net/q/335599 (version: 2019-07-06)
4. BAKRY, D., CATTIAUX, P., GUILLIN, A.: Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. *J. Funct. Anal.* **254**(3), 727–759, 2008
5. BAKRY, D., ÉMERY, M.: Diffusions hypercontractives, Séminaire de Probabilités XIX 1983/84, Springer, 177–206, 1985
6. BAKRY, D., GENTIL, I., LEDOUX, M.: *Analysis and Geometry of Markov Diffusion Operators*, Springer, Cham, 2014
7. BAUDOIN, F.: Wasserstein contraction properties for hypoelliptic diffusions, arXiv:1602.04177 [math], 2016
8. BAUDOIN, F.: Bakry–émery meet Villani. *J. Funct. Anal.* **273**(7), 2275–2291, 2017
9. BAUDOIN, F., GORDINA, M., HERZOG, D.P.: Gamma calculus beyond Villani and explicit convergence estimates for Langevin dynamics with singular potentials. *Arch. Rational Mech. Anal.* **241**(2), 765–804, 2021
10. BERNARD, É., FATHI, M., LEVITT, A., STOLTZ, G.: Hypocoercivity with Schur complements. *Annales Henri Lebesgue* **5**, 523–557, 2022
11. BOGOVSKII, M.E.: Solution of the First Boundary Value Problem for the Equation of Continuity of an Incompressible Medium, Doklady Akademii Nauk, vol. 248, Russian Academy of Sciences, 1037–1040, 1979
12. CAMRUD, E., HERZOG, D.P., STOLTZ, G., GORDINA, M.: Weighted L^2-contractivity of Langevin dynamics with singular potentials. *Nonlinearity* **35**(2), 998, 2021
13. CATTIAUX, P., GUILLIN, A., MONMARCHÉ, P., ZHANG, C.: Entropic multipliers method for Langevin diffusion and weighted log Sobolev inequalities. *J. Funct. Anal.* **277**(11), 108288, 2019
14. CONRAD, F., GROTHAUS, M.: Construction, ergodicity and rate of convergence of N-particle Langevin dynamics with singular potentials. *J. Evol. Equ.* **10**(3), 623–662, 2010
15. COOKE, B., HERZOG, D.P., MATTINGLY, J.C., MCKINLEY, S.A., SCHMIDLER, S.C.: Geometric ergodicity of two-dimensional Hamiltonian systems with a Lennard–Jones-like repulsive potential. *Commun. Math. Sci.* **15**(7), 1987–2025, 2017
16. DALALYAN, A.S., RIOU-DURAND, L.: On sampling from a log-concave density using kinetic Langevin diffusions. *Bernoulli* **26**(3), 1956–1988, 2020
17. DOLBEAULT, J., KLAR, A., MOUHOT, C., SCHMEISER, C.: Exponential rate of convergence to equilibrium for a model describing fiber lay-down processes. *Appl. Math. Res. eXpress* **2013**(2), 165–175, 2013
18. DOLBEAULT, J., MOUHOT, C., SCHMEISER, C.: Hypocoercivity for kinetic equations with linear relaxation terms. *Comptes Rendus Mathematique* **347**(9), 511–516, 2009
19. DOLBEAULT, J., MOUHOT, C., SCHMEISER, C.: Hypocoercivity for linear kinetic equations conserving mass. *Trans. Am. Math. Soc.* **367**(6), 3807–3828, 2015
20. EBERLE, A., GUILLIN, A., ZIMMER, R.: Couplings and quantitative contraction rates for Langevin dynamics. *Ann. Probab.* **47**(4), 1982–2010, 2019
21. ECKMANN, J.-P., HAIRER, M.: Spectral properties of hypoelliptic operators. *Commun. Math. Phys.* **235**(2), 233–253, 2003
22. EVANS, L.C.: *Partial Differential Equations*, vol. 19, American Mathematical Society, Philadelphia, 2010
23. GALDI, G.: *An Introduction to the Mathematical Theory of the Navier–Stokes Equations: Steady-State Problems*, Springer, Berlin, 2011
24. GIGLI, N.: *Answer to "Elliptic regularity with Gibbs measure satisfying Bakry–Emery condition"*, MathOverflow. https://mathoverflow.net/q/335608 (version: 2019-07-06)
25. GROTHAUS, M., STILGENBAUER, P.: Hypocoercivity for Kolmogorov backward evolution equations and applications. *J. Funct. Anal.* **267**(10), 3515–3556, 2014
26. GROTHAUS, M., STILGENBAUER, P.: Hilbert space hypocoercivity for the Langevin dynamics revisited. *Methods Funct. Anal. Topol.* **22**(02), 152–168, 2016
27. HELFFER, B., NIER, F.: *Hypoelliptic Estimates and Spectral Theory for Fokker–Planck Operators and Witten Laplacians*, vol. 1862, Springer, 2005

28. HÉRAU, F.: Hypocoercivity and exponential time decay for the linear inhomogeneous relaxation Boltzmann equation. *Asympt. Anal.* **46**(3–4), 349–359, 2006
29. HÉRAU, F., NIER, F.: Isotropic hypoellipticity and trend to equilibrium for the Fokker–Planck equation with a high-degree potential. *Arch. Ration. Mech. Anal.* **171**(2), 151–218, 2004
30. HERZOG, D.P., MATTINGLY, J.C.: Ergodicity and Lyapunov functions for Langevin dynamics with singular potentials. *Commun. Pure Appl. Math.* **72**(10), 2231–2255, 2019
31. HOOTON, J.G.: Compact Sobolev imbeddings on finite measure spaces. *J. Math. Anal. Appl.* **83**, 570–581, 1981
32. HÖRMANDER, L.: Hypoelliptic second order differential equations. *Acta Math.* **119**, 147–171, 1967
33. IACOBUCCI, A., OLLA, S., STOLTZ, G.: Convergence rates for nonequilibrium Langevin dynamics. *Annales mathématiques du Québec* **43**(1), 73–98, 2019
34. KOLMOGOROV, A.: Zufallige bewegungen (zur theorie der Brownschen bewegung). *Ann. Math.* **66**, 116–117, 1934
35. KOZLOV, S.M.: Effective diffusion in the Fokker–Planck equation. *Math. Notes Acad. Sci. USSR* **45**(5), 360–368, 1989
36. LEDOUX, M.: A simple analytic proof of an inequality by P. Buser. *Proc. Am. Math. Soc.* **121**(3), 951–959, 1994
37. LEDOUX, M.: Spectral gap, logarithmic Sobolev constant, and geometric bounds. *Surv. Differ. Geom.* **9**(1), 219–240, 2004
38. LEIMKUHLER, B., SACHS, M., STOLTZ, G.: Hypocoercivity properties of adaptive Langevin dynamics. *SIAM J. Appl. Math.* **80**(3), 1197–1222, 2020
39. LU, Y., MATTINGLY, J.C.: Geometric ergodicity of Langevin dynamics with Coulomb interactions. *Nonlinearity* **33**(2), 675, 2019
40. MA, Y.-A., CHATTERJI, N.S., CHENG, X., FLAMMARION, N., BARTLETT, P.L., JORDAN, M.I.: Is there an analog of Nesterov acceleration for gradient-based MCMC? *Bernoulli* **27**(3), 1942–1992, 2021
41. MATTINGLY, J.C., STUART, A.M., HIGHAM, D.J.: Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stoch. Process. Appl.* **101**(2), 185–232, 2002
42. METAFUNE, G., PALLARA, D., PRIOLA, E.: Spectrum of Ornstein–Uhlenbeck operators in lp spaces with respect to invariant measures. *J. Funct. Anal.* **196**(1), 40–60, 2002
43. MOUHOT, C., NEUMANN, L.: Quantitative perturbative study of convergence to equilibrium for collisional kinetic models in the torus. *Nonlinearity* **19**(4), 969, 2006
44. OTTO, F., VILLANI, C.: Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.* **173**(2), 361–400, 2000
45. PAVLIOTIS, G.A.: Stochastic Processes and Applications: Diffusion Processes, the Fokker–Planck and Langevin Equations, vol. 60, Springer, 2014
46. RISKEN, H.: *Fokker–Planck Equation: Methods of Solution and Applications, Springer Series in Synergetics*, 1989
47. ROUSSEL, J., STOLTZ, G.: Spectral methods for Langevin dynamics and associated error estimates. *ESAIM Math. Model. Numer. Anal.* **52**(3), 1051–1083, 2018
48. SASON, I., VERDÚ, S.: f-divergence inequalities. *IEEE Trans. Inf. Theory* **62**(11), 5973–6006, 2016
49. STOLTZ, G., TRSTANOVA, Z.: Langevin dynamics with general kinetic energies. *Multiscale Model. Simul.* **16**(2), 777–806, 2018
50. STOLTZ, G., VANDEN-EIJNDEN, E.: Longtime convergence of the temperature-accelerated molecular dynamics method. *Nonlinearity* **31**(8), 3748–3769, 2018
51. TALAY, D.: Stochastic Hamiltonian systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. *Markov Process. Relat. Fields* **8**(2), 163–198, 2002
52. TROPPER, M.M.: Ergodic and quasideterministic properties of finite-dimensional stochastic systems. *J. Stat. Phys.* **17**(6), 491–509, 1977

53. VILLANI, C.: Hypocoercive diffusion operators. *Bollettino dell'Unione Matematica Italiana* **10-B**(2), 257–275, 2007
54. VILLANI, C.: Hypocoercivity. *Mem. Am. Math. Soc.* **202**(950), 66, 2009
55. WU, L.: Large and moderate deviations and exponential convergence for stochastic damping Hamiltonian systems. *Stoch. Process. Appl.* **91**(2), 205–238, 2001

Yu Cao
Institute of Natural Sciences, School of Mathematical Sciences,
Shanghai Jiao Tong University,
Shanghai
200240 China. e-mail: yucao@sjtu.edu.cn

and

Jianfeng Lu
Department of Mathematics,
Duke University,
Durham
NC
27708 USA. e-mail: jianfeng@math.duke.edu

and

Department of Physics,
Duke University,
Durham
NC
27708 USA.

and

Department of Chemistry,
Duke University,
Durham
NC
27708 USA.

and

Lihan Wang
Department of Mathematical Sciences,
Carnegie Mellon University,
Pittsburgh
PA
15213 USA. e-mail: lihanw@andrew.cmu.edu