ON EXPLICIT L^2-CONVERGENCE RATE ESTIMATE FOR UNDERDAMPED LANGEVIN DYNAMICS

YU CAO, JIANFENG LU, AND LIHAN WANG

ABSTRACT. We provide a new explicit estimate of exponential decay rate of underdamped Langevin dynamics in L^2 distance. To achieve this, we first prove a Poincaré-type inequality with Gibbs measure in space and Gaussian measure in momentum. Our new estimate provides a more explicit and simpler expression of decay rate; moreover, when the potential is convex with Poincaré constant $m \ll 1$, our new estimate offers the decay rate of $\mathcal{O}(\sqrt{m})$ after optimizing the choice of friction coefficient, which is much faster compared to $\mathcal{O}(m)$ for the overdamped Langevin dynamics.

1. INTRODUCTION

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

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

where $U(x)$ is the potential energy, γ 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 associated Fokker-Planck (FP) equation of (1) for the probability density function $\rho(t, x, v)$ is given by

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

It is well-known that under mild assumptions the above FP equation (2) admits a unique stationary density function given by

$$\rho_\infty(x, v) = \mu_U(x)\kappa(v), \tag{3}$$

where

$$\mu_U(x) = \frac{1}{Z_U} e^{-U(x)}, \qquad \kappa(v) = \frac{1}{(2\pi)^{d/2}} e^{-\frac{|v|^2}{2}}, \qquad Z_U = \int_{\mathbb{R}^d} e^{-U(x)}\,\mathrm{d}x.$$

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., [20, Sec. 6.5]), which is given by

$$\mathrm{d}x_t^{(\infty)} = -\nabla U(x_t^{(\infty)})\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t.$$

An equivalent formalism is the following *backward Kolmogorov equation*,

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

where $\mathcal{L}_{\text{ham}}$ is the energy-conservation term and $\mathcal{L}_{\text{FD}}$ is the fluctuation-dissipation term

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

Date: February 4, 2020.
Key words and phrases. underdamped Langevin dynamics; Poincaré inequality; convergence rate; hypocoercivity.

As a remark, (4) could be derived from (2) by considering $\rho(t,x,v) = f(t,x,-v)\rho_\infty(x,v)$ [20]; since $\|\rho-\rho_\infty\|_{L^2_{\rho_\infty^{-1}}} \equiv \left\|f - \int f \,\mathrm{d}\rho_\infty\right\|_{L^2_{\rho_\infty}}$, the exponential convergence of the Fokker-Planck equation (2) is equivalent to the exponential decay of $f(t,\cdot,\cdot)$ to zero, provided that $\int f_0 \,\mathrm{d}\rho_\infty = 0$. 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_U satisfies a Poincaré inequality, one could show that the generator in the above equation is self-adjoint and coercive with respect to $L^2_{\mu_U}$. As a consequence, if $\int h_0 \,\mathrm{d}\mu_U = 0$, then $h(t,x)$ decays to zero exponentially fast as $t\to\infty$.

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. After a huge amount of works, the exponential convergence of the underdamped Langevin dynamics is now well understood in various norms (see Sec. 1.1 below for a brief review).

Our goal in this work is to provide a rather explicit estimate of the decay rate in L^2 for the semigroup in (4), based on a framework by Armstrong and Mourrat recently proposed in [2]. In particular, under some additional assumptions of U, we obtain explicit estimates for $\lambda > 0$ such that for any possible $f = f(t,x,v)$ satisfying (4) and $\int f_0 \,\mathrm{d}\rho_\infty = 0$, we have

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

In the rest of this section, we shall first briefly review existing approaches to study the exponential convergence of (4) (or equivalently (2)) in Sec. 1.1. Next, we will present in Sec. 1.2 our assumptions and main results, and compare our estimate of the decay rate λ with some previous works aiming at explicit estimates [8, 21, 6].

1.1. **A brief literature review.** 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 [26, Theorem 35]; see also [25] for a brief overview of main ideas. 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 [26, Sec. 7.2].

(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 [7, 8]. 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 [21, Sec. 2] 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 Lyaponov function E, based on [6, Sec. 2], because their setup is consistent with our $L^2_{\rho_\infty}$ estimate in Sec. 1.2 below.

As a remark, the DMS method [7, 8] has been extended or adapted to study the convergence of spherical velocity Langevin equation [12], non-equilibrium Langevin dynamics [14], Langevin dynamics with general kinetic energy [23], temperature-accelerated molecular dynamics [24], just to name a few. It might be interesting to study whether the variational framework [2] we based on can be extended to these cases.

(iii) (Convergence in Wasserstein distance). Baudoin discussed a general framework of the Bakry-Émery methodology [3] to hypoelliptic and hypocoercive operators, based on which the exponential convergence of the kinetic FP equation (quantified by a Wasserstein distance associated with a special metric) was proved under certain assumptions on the potential $U(x)$ [5, Theorem 2.6]; see also [4].

A different approach is coupling method for underdamped Langevin dynamics (1). In [6, 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 [6, 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 [9].

There are other approaches to study the long time behavior of the underdamped Langevin dynamics, e.g., Lyaponov function [18] and spectrum analysis [10, 15]. We will not go into details here.

1.2. Assumptions and main results.

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

$$\int_{\mathbb{R}^d}\left(f-\int_{\mathbb{R}^d} f\,\mathrm{d}\mu_U\right)^2 \mathrm{d}\mu_U \leqslant \frac{1}{m}\int_{\mathbb{R}^d}|\nabla f|^2\,\mathrm{d}\mu_U, \qquad \forall f\in H^1_{\mu_U}. \tag{8}$$

Assumption 2. *$U\in C^2(\mathbb{R}^d)$, and the matrix norm of $\nabla^2 U$, the Hessian of U, satisfies*

$$\|\nabla_x^2 U(x)\| \leqslant M(1+|\nabla_x U(x)|), \qquad \forall\; x\in\mathbb{R}^d. \tag{9}$$

for some constant $M\geqslant 1$.

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

The Assumption 2 is commonly used in the literature, see e.g., the books [20, 26] and the paper [21], and is satisfied when U grows at most exponentially fast as $x\to\infty$. Assumption 3 is satisfied when

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

for some $\alpha>1$ (see [13] for a proof).

The L^2_κ-dual space of H^1_κ is denoted by H^{-1}_κ. By abuse of notation, we denote the canonical pairing $\langle\cdot,\cdot\rangle_{H^1_\kappa,H^{-1}_\kappa}$ between $f\in H^1_\kappa$ and $f^*\in H^{-1}_\kappa$ by

$$\int_{\mathbb{R}^d} f f^*\,\mathrm{d}\kappa(v) := \langle f,f^*\rangle_{H^1_\kappa,H^{-1}_\kappa}.$$

For an arbitrary Banach space X and time interval I, we denote by $L^p(I\times\mu_U;X)$ the Banach space of functions $f(t,x,v)$ with norm

$$\|f\|_{L^p(I\times\mu_U;X)} := \left(\int_{I\times\mathbb{R}^d}\|f(t,x,\cdot)\|_X^p\,\mathrm{d}t\,\mathrm{d}\mu_U(x)\right)^{\frac{1}{p}}.$$

For the rest of the paper, we consider $I=(0,T)$. Finally, inspired by the work of Armstrong and Mourrat [2], we define the Banach space

$$H^1_{hyp}(I\times\mu_U) := \left\{f\in L^2(I\times\mu_U,H^1_\kappa)\ :\ \partial_t f-\mathcal{L}_{\mathrm{ham}} f\in L^2(I\times\mu_U;H^{-1}_\kappa)\right\}$$

Theorem 1. *Under Assumptions 1, 2, and 3, there exist a constant $\lambda>0$ and universal constants C_0,c_0 independent of all parameters such that, for every $T\in(0,\infty)$ and $f\in H^1_{hyp}((0,T)\times\mu_U)$ satisfying*

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} f(t=0,x,v)\,\mathrm{d}\rho_\infty(x,v)=0, \tag{10}$$

and the backward Kolmogorov equation (4), we have, for every $t\in(0,T)$,

$$\|f(t,\cdot)\|_{L^2(\mu_U;L^2_\kappa)} \leqslant C_0\exp(-\lambda t)\|f(t=0,\cdot)\|_{L^2(\mu_U;L^2_\kappa)}.$$

Moreover, λ can be made explicit as

$$\lambda = \sqrt{m}\log\Big(1 + \frac{\gamma\sqrt{m}}{c_0(\sqrt{m}+R+\gamma)^2}\Big)$$

with some constant $R > 0$ which is given by

(i) If U is convex, then

$$R = 0.$$

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

$$\nabla_x^2 U(x) \geqslant -K\,\mathrm{Id}, \qquad \forall\, x \in \mathbb{R}^d \tag{11}$$

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,

$$R = M\sqrt{d}.$$

Remark 1.1.

(i) If we fix $m = \mathcal{O}(1)$, then, when $\gamma \to 0$ (resp. $\gamma \to \infty$), our estimate provides an estimate on decay rate of $\mathcal{O}(\gamma)$ (resp. $\mathcal{O}(\gamma^{-1})$). This is consistent with [21] 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

$$\lambda = \sqrt{m}\log\big(1 + \frac{1}{4c_0}\big).$$

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. To the best of our knowledge, this optimal $\sqrt{m}$ convergence rate is new in the literature. We refer the readers to Appendix B for the corresponding results from the DMS method, with a slightly more explicit estimate compared to [21].

(iii) In case that Hessian of U is bounded from below, condition (11) 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 [16, Theorem 1] and [17, Sec. 5]. Similar assumption is also used in [19, Theorem 1] for functional inequalities.

(iv) It is well-known that for overdamped Langevin dynamics, the decay rate is simply m in $L^2_{\mu_U}$ 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 $\mathcal{O}(\sqrt{m})$, which is much faster than the overdamped Langevin dynamics.

(v) Despite there is no direct link between the decay rates in L^2 distance and Wasserstein distance, we can nonetheless compare the scaling. As we mentioned earlier, using synchronous coupling, Dalalyan and Riou-Durand [6] provided an explicit estimate of the decay rate in W_2 of $\mathcal{O}(m/\sqrt{L+m})$, under the stronger assumption that $mId \leqslant \nabla^2 U \leqslant LId$. While this agrees with our estimate when $L = \mathcal{O}(m)$, in general it is worse, especially when L is much larger.

Remark 1.2. *Due to the following relation (see e.g., [22])*

$$\begin{aligned}\frac{1}{\sqrt{2}}\,\|\rho-\rho_\infty\|_{TV} \leqslant & \sqrt{D\big(\rho\,||\,\rho_\infty\big)} \leqslant \sqrt{\chi^2(\rho,\rho_\infty)} \\ & \equiv \|\rho-\rho_\infty\|_{L^2_{\rho_\infty^{-1}}} \equiv \Big\| f - \int f\,\mathrm{d}\rho_\infty \Big\|_{L^2_{\rho_\infty}},\end{aligned}$$

where $f = \mathrm{d}\rho/\,\mathrm{d}\rho_\infty$. Theorem 1 implies that $\rho(t,\cdot,\cdot)$ converges to ρ_∞ with rate 2λ in both χ^2-divergence and relative entropy, and with rate λ in total variation distance.

Our decay estimate is based on the following Poincaré inequality:

Theorem 2. *Under Assumptions 1, 2, and 3, there exist a universal constant C_0 independent of all parameters, and a constant $R < \infty$ (the same constant as in Thoerem 1) such that for every $f \in H^1_{hyp}(I \times \mu_U)$, we have*

$$\|f - (f)_{I\times\mu_U}\|_{L^2(I\times\mu_U;L^2_\kappa)} \leqslant C_0\Big(\big(1 + \frac{1}{T\sqrt{m}} + R(\frac{1}{\sqrt{m}} + T)\big)\|\nabla_v f\|_{L^2(I\times\mu_U;L^2_\kappa)} + (\frac{1}{\sqrt{m}} + T)\|\partial_t f - \mathcal{L}_{ham} f\|_{L^2(I\times\mu_U;H^{-1}_\kappa)}\Big), \tag{12}$$

where

$$(f)_{I\times\mu_U} := \int_{I\times\mathbb{R}^d\times\mathbb{R}^d} f(t,x,v)\,\mathrm{d}t\,\mathrm{d}\rho_\infty(x,v).$$

Notations. We define a projection operator for $\phi(t,x,v) \in L^2(I\times\mu_U; L^2_\kappa)$ by

$$(\Pi_v\phi)(t,x) := \int_{\mathbb{R}^d} \phi(t,x,v)\,\mathrm{d}\kappa(v). \tag{13}$$

Π_v is used to obtain the marginal component of ϕ in $L^2(I\times\mu_U)$. By slight abuse of notation, for $\phi(x,v) \in L^2_{\rho_\infty}(\mathbb{R}^{2d})$, we also use the same notation Π_v to represent the similar projection, i.e., $(\Pi_v\phi)(x) := \int_{\mathbb{R}^d}\phi(x,v)\ \mathrm{d}\kappa(v)$. The adjoint of ∇_x and ∇_v in the Hilbert space $L^2(\rho_\infty)$ is 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 the vector field $F(x,v) : \mathbb{R}^{2d}\to\mathbb{R}^d$. Thus we can rewrite operators $\mathcal{L}_{\mathrm{ham}}$ and $\mathcal{L}_{\mathrm{FD}}$ by

$$\mathcal{L}_{\mathrm{ham}} = \nabla_v^*\nabla_x - \nabla_x^*\nabla_v, \qquad \mathcal{L}_{\mathrm{FD}} = -\nabla_v^*\nabla_v. \tag{14}$$

Since we are also dealing with functions that involve both time and space variables, we also introduce short-hand notations $\bar\nabla = (\partial_t, \nabla_x)^\top$ and its $L^2(I\times\mu_U)$-adjoint $\bar\nabla^* = (-\partial_t, \nabla_x^*)$. We also use the convention $\partial_{x_0} = \partial_t$.

2. Proofs

We start with a few elementary lemmas.

Lemma 2.1. *(Poincaré Inequality) For $f \in H^1(I\times\mu_U)$,*

$$\|f - (f)_{I\times\mu_U}\|^2_{L^2(I\times\mu_U)} \leqslant \max\{\frac{1}{m}, \frac{T^2}{\pi^2}\}\Big(\|\partial_t f\|^2_{L^2(I\times\mu_U)} + \|\nabla_x f\|^2_{L^2(I\times\mu_U)}\Big). \tag{15}$$

Proof. Define $(f)_x(t) = \int_{\mathbb{R}^d} f(x,t)\,\mathrm{d}\mu_U(x)$. Then,

$$\begin{aligned}
&\|f - (f)_{I\times\mu_U}\|^2_{L^2(I\times\mu_U)}\\
&= \|f - (f)_x\|^2_{L^2(I\times\mu_U)} + \|(f)_x - (f)_{I\times\mu_U}\|^2_{L^2(I\times\mu_U)}\\
&\leqslant \int_0^T \|f(t,x) - (f)_x(t)\|^2_{L^2_{\mu_U}}\ \mathrm{d}t + \int_0^T\Big(\int_{\mathbb{R}^d}\big(f(t,x) - \int_0^T f(s,x)\ \mathrm{d}s\big)\,\mathrm{d}\mu_U(x)\Big)^2\mathrm{d}t\\
&\overset{(8)}{\leqslant} \frac{1}{m}\int_0^T \|\nabla_x f\|^2_{L^2_{\mu_U}}(t)\,\mathrm{d}t + \int_{\mathbb{R}^d}\int_0^T\big(f(t,x) - \int_0^T f(s,x)\ \mathrm{d}s\big)^2\mathrm{d}t\,\mathrm{d}\mu_U(x)\\
&\leqslant \frac{1}{m}\|\nabla_x f\|^2_{L^2(I\times\mu_U)} + \frac{T^2}{\pi^2}\int_{\mathbb{R}^d}\|\partial_t f(x)\|^2_{L^2(0,T)}\,\mathrm{d}\mu_U(x)\\
&= \max\{\frac{1}{m}, \frac{T^2}{\pi^2}\}\big(\|\partial_t f\|^2_{L^2(I\times\mu_U)} + \|\nabla_x f\|^2_{L^2(I\times\mu_U)}\big),
\end{aligned}$$

where the last inequality follows the standard Poincaré inequality on $(0,T)$:

$$\|f - \int_0^T f(s)\,\mathrm{d}s\|_{L^2(0,T)} \leqslant \frac{T}{\pi}\|f'(t)\|_{L^2(0,T)}.$$

□

Lemma 2.2. *For any $\phi \in H^1(I \times \mu_U)$, we have*

$$\|\phi \nabla_x U\|^2_{L^2(I\times\mu_U)} \leqslant 8\|\nabla_x \phi\|^2_{L^2(I\times\mu_U)} + 2(M^2d^2 + Md)\|\phi\|^2_{L^2(I\times\mu_U)}, \tag{16}$$

where M is the constant in (9).

Proof.

$$\begin{aligned}
\|\phi\nabla_x U\|^2_{L^2(I\times\mu_U)} &= \int_{I\times\mathbb{R}^d} \phi^2 \nabla_x U \cdot \nabla_x U \,\mathrm{d}t\,\mathrm{d}\mu_U(x)\\
&= \int_{I\times\mathbb{R}^d} \nabla_x \cdot (\phi^2 \nabla_x U)\,\mathrm{d}t\,\mathrm{d}\mu_U(x)\\
&= 2\int_{I\times\mathbb{R}^d} \phi\nabla_x\phi\cdot\nabla_x U\,\mathrm{d}t\,\mathrm{d}\mu_U(x) + \int_{I\times\mathbb{R}^d}\phi^2\Delta_x U\,\mathrm{d}t\,\mathrm{d}\mu_U(x)\\
&\overset{(9)}{\leqslant} \frac{1}{4}\|\phi\nabla_x U\|^2_{L^2(I\times\mu_U)} + 4\|\nabla_x\phi\|^2_{L^2(I\times\mu_U)}\\
&\quad + Md\|\phi\|^2_{L^2(I\times\mu_U)} + Md\int_{I\times\mathbb{R}^d}\phi^2|\nabla_x U|\,\mathrm{d}t\,\mathrm{d}\mu_U(x)\\
&\leqslant \frac{1}{4}\|\phi\nabla_x U\|^2_{L^2(I\times\mu_U)} + 4\|\nabla_x\phi\|^2_{L^2(I\times\mu_U)} + Md\|\phi\|^2_{L^2(I\times\mu_U)}\\
&\quad + M^2d^2\|\phi\|^2_{L^2(I\times\mu_U)} + \frac{1}{4}\|\phi\nabla_x U\|^2_{L^2(I\times\mu_U)}.
\end{aligned}$$

Thus we finish the proof of (16) after rearranging. ∎

Lemma 2.3. *For any $u \in H^2(I \times \mu_U)$,*

$$\|D^2u\|^2_{L^2(I\times\mu_U)} = \sum_{i,j=0}^{d}\|\partial_{x_i}\partial_{x_j}u\|^2_{L^2(I\times\mu_U)} \leqslant C_0\Big(\|\bar\nabla^*\bar\nabla u\|^2_{L^2(I\times\mu_U)} + R^2\|\bar\nabla u\|^2_{L^2(I\times\mu_U)}\Big)$$

Here C_0 is a universal constant, and R is the same constant defined in Theorem 1.

Proof. The starting point of the proof is Bochner's formula

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

Integrate over $I \times \mu_U$ and (noticing the last term above has integral zero) we get

$$\sum_{i,j=0}^{d}\|\partial_{x_i,x_j}u\|^2_{L^2(I\times\mu_U)} = \|\bar\nabla^*\bar\nabla u\|^2_{L^2(I\times\mu_U)} - \int_{I\times\mathbb{R}^d}(\nabla_x u)^\top\nabla_x^2 U\nabla_x u\,\mathrm{d}t\,\mathrm{d}\mu_U(x). \tag{17}$$

Let us first discuss the easier case when U satisfies (11). We can estimate directly

$$\begin{aligned}
\sum_{i,j=0}^{d}\|\partial_{x_i,x_j}u\|^2_{L^2(I\times\mu_U)} &= \|\bar\nabla^*\bar\nabla u\|^2_{L^2(I\times\mu_U)} - \int_{I\times\mathbb{R}^d}(\nabla_x u)^\top\nabla_x^2 U\nabla_x u\,\mathrm{d}t\,\mathrm{d}\mu_U(x)\\
&\leqslant \|\bar\nabla^*\bar\nabla u\|^2_{L^2(I\times\mu_U)} + K\|\nabla_x u\|^2_{L^2(I\times\mu_U)}.
\end{aligned} \tag{18}$$

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

Now we deal with the more general case, without assuming (11). For the rest of the proof, we let C_0 be a universal constant independent of all parameters, and may change line by line. Using (16) with $\phi = \partial_{x_i}u,\ i = 1, \cdots, d$,

$$\begin{aligned}
\int_{I\times\mathbb{R}^d}|\nabla_x u|^2|\nabla_x U|^2\,\mathrm{d}t\,\mathrm{d}\mu_U(x) &= \sum_{i=1}^{d}\int_{I\times\mathbb{R}^d}(\partial_{x_i}u)^2|\nabla_x U|^2\,\mathrm{d}t\,\mathrm{d}\mu_U(x)\\
&\overset{(16)}{\leqslant} C_0\Big(\|D^2u\|^2_{L^2(I\times\mu_U)} + M^2d^2\int_{I\times\mathbb{R}^d}|\nabla_x u|^2\,\mathrm{d}t\,\mathrm{d}\mu_U(x)\Big)
\end{aligned}$$

$$
\begin{aligned}
&\overset{(17)}{\leqslant} C_0\Big(\|\bar\nabla^*\bar\nabla u\|^2_{L^2(I\times\mu_U)} + M^2d^2\int_{I\times\mathbb{R}^d}|\nabla_x u|^2\,\mathrm{d}t\,\mathrm{d}\mu_U(x)\\
&\qquad - \int_{I\times\mathbb{R}^d}(\nabla_x u)^\top\nabla_x^2U\nabla_x u\,\mathrm{d}t\,\mathrm{d}\mu_U(x)\Big)\\
&\overset{(9)}{\leqslant} C_0\Big(\|\bar\nabla^*\bar\nabla u\|^2_{L^2(I\times\mu_U)} + M^2d^2\int_{I\times\mathbb{R}^d}|\nabla_x u|^2\,\mathrm{d}t\,\mathrm{d}\mu_U(x)\\
&\qquad + M\int_{I\times\mathbb{R}^d}|\nabla_x u|^2(1+|\nabla_x U|)\,\mathrm{d}t\,\mathrm{d}\mu_U(x)\Big)\\
&\leqslant C_0\Big(\|\bar\nabla^*\bar\nabla u\|^2_{L^2(I\times\mu_U)} + M^2d^2\int_{I\times\mathbb{R}^d}|\nabla_x u|^2\,\mathrm{d}t\,\mathrm{d}\mu_U(x)\\
&\qquad + M^2\int_{I\times\mathbb{R}^d}|\nabla_x u|^2\,\mathrm{d}t\,\mathrm{d}\mu_U(x)\Big) + \frac{1}{2}\int_{I\times\mathbb{R}^d}|\nabla_x u|^2|\nabla_x U|^2\,\mathrm{d}t\,\mathrm{d}\mu_U(x).
\end{aligned}
$$

Rearranging the terms, we arrive at

$$
\int_{I\times\mathbb{R}^d}|\nabla_x u|^2|\nabla_x U|^2\,\mathrm{d}t\,\mathrm{d}\mu_U(x) \leqslant C_0\Big(\|\bar\nabla^*\bar\nabla u\|^2_{L^2(I\times\mu_U)} + M^2d^2\int_{I\times\mathbb{R}^d}|\nabla_x u|^2\,\mathrm{d}t\,\mathrm{d}\mu_U(x)\Big). \tag{19}
$$

Therefore by (19),

$$
\begin{aligned}
\|D^2u\|^2_{L^2(I\times\mu_U)} &\overset{(9),(17)}{\leqslant} C_0\Big(\|\bar\nabla^*\bar\nabla u\|^2_{L^2(I\times\mu_U)} + M\int_{I\times\mathbb{R}^d}|\nabla_x u|^2(1+|\nabla_x U|)\,\mathrm{d}t\,\mathrm{d}\mu_U(x)\Big)\\
&\leqslant C_0\Big(\|\bar\nabla^*\bar\nabla u\|^2_{L^2(I\times\mu_U)} + M\|\nabla_x u\|^2_{L^2(I\times\mu_U)}\\
&\qquad + M\|\nabla_x u\|_{L^2(I\times\mu_U)}\||\nabla_x u||\nabla_x U|\|_{L^2(I\times\mu_U)}\Big)\\
&\overset{(19)}{\leqslant} C_0\Big(\|\bar\nabla^*\bar\nabla u\|^2_{L^2(I\times\mu_U)} + M\|\nabla_x u\|^2_{L^2(I\times\mu_U)}\\
&\qquad + M\|\nabla_x u\|_{L^2(I\times\mu_U)}(\|\bar\nabla^*\bar\nabla u\|_{L^2(I\times\mu_U)} + Md\|\nabla_x u\|_{L^2(I\times\mu_U)})\Big)\\
&\leqslant C_0\Big(\|\bar\nabla^*\bar\nabla u\|^2_{L^2(I\times\mu_U)} + M^2d\|\nabla_x u\|^2_{L^2(I\times\mu_U)}\Big).
\end{aligned}
$$

□

One of the key lemmas of our proof is the following result on elliptic regularity on $I\times\mathbb{R}^d$ with the measure $\mathrm{d}t\,\mathrm{d}\mu_U$. 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:*

$$
\begin{cases} -\partial_{tt}u + \nabla_x^*\nabla_x u = h & \text{in } I\times\mathbb{R}^d,\\ \partial_t u(t=0,\cdot) = \partial_t u(t=T,\cdot) = 0 & \text{in } \mathbb{R}^d. \end{cases} \tag{20}
$$

Assume $h\in H^{-1}(I\times\mu_U)$, and $(h)_{I\times\mu_U}=0$. Define the function space

$$
V = \big\{u\in H^1(I\times\mu_U)\ :\ \int_{I\times\mathbb{R}^d}u(t,x)\,\mathrm{d}t\,\mathrm{d}\mu_U(x) = 0\big\}.
$$

Then

(i) *There exists a unique $u\in V$ which is a weak solution to (20). More precisely, for any $v\in H^1(I\times\mu_U)$, we have*

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

When $h\in L^2(I\times\mu_U)$, we have the estimate

$$
\|\partial_t u\|^2_{L^2(I\times\mu_U)} + \|\nabla_x u\|^2_{L^2(I\times\mu_U)} \leqslant \max\{\frac{1}{m},\frac{T^2}{\pi^2}\}\|h\|^2_{L^2(I\times\mu_U)}. \tag{21}
$$

(ii) If $h \in L^2(I \times \mu_U)$, then the solution u to (20) satisfies $u \in H^2(I \times \mu_U)$, with the estimate

$$\sum_{i,j=0}^{d} \|\partial_{x_i,x_j} u\|^2_{L^2(I\times\mu_U)} \leqslant C_0\big(1 + R^2 \max\{\frac{1}{m}, T^2\}\big)\|h\|^2_{L^2(I\times\mu_U)}, \tag{22}$$

where C_0 is a universal constant and R is defined above as in Theorem 1.

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 $H^1(I \times \mu_U)$, and for the rest of the paper we equip it with the $H^1(I \times \mu_U)$ norm. We also define the following inner-product:

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

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 $\int_{I\times\mathbb{R}^d} u\, \mathrm{d}t\, \mathrm{d}\mu_U(x) = 0$. If u is a weak solution of (20), then for any $v \in V$, $B(u, v) = \int_{I\times\mathbb{R}^d} hv\, \mathrm{d}t\, \mathrm{d}\mu_U(x)$, and necessarily $\int_{I\times\mathbb{R}^d} h\, \mathrm{d}t\, \mathrm{d}\mu_U(x) = 0$ when we take $v = 1$.

Since $\int_{I\times\mathbb{R}^d} u(t, x)\, \mathrm{d}t\, \mathrm{d}\mu_U(x) = 0$, by Poincaré inequality (Lemma 2.1) we can show B is coercive under $H^1(I \times \mu_U)$ norm in the sense of

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

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

$$|H(v)| \leqslant \|h\|_{H^{-1}(I\times\mu_U)}\|v\|_{H^1(I\times\mu_U)}.$$

Thus by Lax-Milgram's Theorem, the equation (20) has a unique weak solution $u \in V$. Finally when $h \in L^2(I \times \mu_U)$

$$\begin{aligned} (\|\partial_t u\|^2_{L^2(I\times\mu_U)} + \|\nabla_x u\|^2_{L^2(I\times\mu_U)})^2 &= B[u, u]^2 \\ &= \Big(\int_{I\times\mathbb{R}^d} hu\, \mathrm{d}t\, \mathrm{d}\mu_U(x)\Big)^2 \leqslant \|h\|^2_{L^2(I\times\mu_U)}\|u\|^2_{L^2(I\times\mu_U)} \\ &\leqslant \max\{\frac{1}{m}, \frac{T^2}{\pi^2}\}\|h\|^2_{L^2(I\times\mu_U)}\big(\|\partial_t u\|^2_{L^2(I\times\mu_U)} + \|\nabla_x u\|^2_{L^2(I\times\mu_U)}\big), \end{aligned}$$

and the desired estimate follows.

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

$$\begin{cases} -\partial_{tt} w_i + \nabla_x^* \nabla_x w_i = \partial_{x_i} h - \nabla_x u \cdot \nabla_x \partial_{x_i} U & \text{in } I \times \mathbb{R}^d, \\ \partial_t w_i(t = 0, \cdot) = \partial_t w_i(t = T, \cdot) = 0 & \text{in } \mathbb{R}^d. \end{cases} \tag{23}$$

We first verify the rhs has total integral zero. Indeed

$$\begin{aligned} &\int_{I\times\mathbb{R}^d} (\partial_{x_i} h - \nabla_x u \cdot \nabla_x \partial_{x_i} U)\, \mathrm{d}t\, \mathrm{d}\mu_U(x) \\ &= \int_{I\times\mathbb{R}^d} (h\partial_{x_i} U - \nabla_x u \cdot \nabla_x \partial_{x_i} U)\, \mathrm{d}t\, \mathrm{d}\mu_U(x) \\ &= \int_{I\times\mathbb{R}^d} \big((-\partial_{tt} u + \nabla_x^* \nabla_x u)\partial_{x_i} U - \nabla_x u \cdot \nabla_x \partial_{x_i} U\big)\, \mathrm{d}t\, \mathrm{d}\mu_U(x) \end{aligned}$$

$$= \int_{I\times\mathbb{R}^d} \big(\partial_t u \partial_{tx_i} U + \nabla_x u\cdot\nabla_x\partial_{x_i}U - \nabla_x u\cdot\nabla_x\partial_{x_i}U\big)\,\mathrm{d}t\,\mathrm{d}\mu_U(x) = 0.$$

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

$$\begin{aligned}
&\int_{I\times\mathbb{R}^d}(\partial_{x_i}h - \nabla_x u\cdot\nabla_x\partial_{x_i}U)\phi\,\mathrm{d}t\,\mathrm{d}\mu_U(x)\\
&\quad\leqslant \int_{I\times\mathbb{R}^d}(-h\partial_{x_i}\phi + h\phi\partial_{x_i}U)\,\mathrm{d}t\,\mathrm{d}\mu_U(x) + \int_{I\times\mathbb{R}^d}|\phi\nabla_x u||\nabla_x\partial_{x_i}U|\,\mathrm{d}t\,\mathrm{d}\mu(x)\\
&\quad\overset{(9)}{\leqslant} \|h\|_{L^2(I\times\mu_U)}(1+\|\phi\partial_{x_i}U\|_{L^2(I\times\mu_U)}) + M\int_{I\times\mathbb{R}^d}|\phi\nabla_x u|(1+|\nabla_x U|)\,\mathrm{d}t\,\mathrm{d}\mu(x)\\
&\quad\leqslant \|h\|_{L^2(I\times\mu_U)}(1+\|\phi\partial_{x_i}U\|_{L^2(I\times\mu_U)}) + M\|\nabla_x u\|_{L^2(I\times\mu_U)}(1+\|\phi\nabla_x U\|_{L^2(I\times\mu_U)})\\
&\quad\overset{(16),(21)}{\leqslant} C(M)\|h\|_{L^2(I\times\mu_U)},
\end{aligned}$$

where $C(M)>0$ is a constant depending on M. Therefore, by (i) we know there exists a $w_i\in V$ which is the weak solution of (23).

The next step is to verify $w_i = \partial_{x_i}u - \frac{1}{T}\int_{I\times\mu_U}\partial_{x_i}u\,\mathrm{d}t\,\mathrm{d}\mu_U(x)$. We know both functions are in $L^2(I\times\mu_U)$ with spatial integral zero. For derivations below, we use the short-hand notation $(f,g):=\int_{I\times\mathbb{R}^d} fg\,\mathrm{d}t\,\mathrm{d}\mu_U(x)$, and we don't specify the regularities of f and g. Take any test function $\phi\in H^2(I\times\mu_U)$. Using integration by parts it is easy to check

$$(-\partial_{tt}w_i + \nabla_x^*\nabla_x w_i,\phi) = (\partial_t w_i,\partial_t\phi) + (\nabla_x w_i,\nabla_x\phi).$$

Now we do the calculations using the equation:

$$\begin{aligned}
(-\partial_{tt}w_i + \nabla_x^*\nabla_x w_i,\phi) &= (\partial_{x_i}h - \nabla_x\partial_{x_i}U\cdot\nabla_x u,\phi)\\
&= (h,-\partial_{x_i}\phi + \phi\partial_{x_i}U) - (\nabla_x\partial_{x_i}U\cdot\nabla_x u,\phi)\\
&= (-\partial_{tt}u + \nabla_x^*\nabla_x u, -\partial_{x_i}\phi + \phi\partial_{x_i}U) - (\nabla_x\partial_{x_i}U\cdot\nabla_x u,\phi)\\
&= \big(\partial_t u,\partial_t(-\partial_{x_i}\phi + \phi\partial_{x_i}U)\big) + \big(\nabla_x u,\nabla_x(-\partial_{x_i}\phi+\phi\partial_{x_i}U)\big)\\
&\quad - (\nabla_x\partial_{x_i}U\cdot\nabla_x u,\phi)\\
&= \big(\partial_t u,\partial_t(-\partial_{x_i}\phi + \phi\partial_{x_i}U)\big) + (\nabla_x u, -\nabla_x\partial_{x_i}\phi + \partial_{x_i}U\nabla_x\phi)\\
&= (\partial_t\partial_{x_i}u,\partial_t\phi) + (\nabla_x\partial_{x_i}u,\nabla_x\phi).
\end{aligned}$$

The equality holds when we replace $\partial_{x_i}u$ by $\partial_{x_i}u - \frac{1}{T}\int_{I\times\mu_U}\partial_{x_i}u\,\mathrm{d}t\,\mathrm{d}\mu_U(x)$ as it is invariant under a constant shift. This shows that, as $L^2(I\times\mu_U)$ functions, $w_i = \partial_{x_i}u - \frac{1}{T}\int_{I\times\mu_U}\partial_{x_i}u\,\mathrm{d}t\,\mathrm{d}\mu_U(x)$. However we already have $w_i\in H^1(I\times\mu_U)$, thus we have shown that $\partial_{x_i}u\in H^1(I\times\mu_U)$ for each $i=1,\cdots,d$. Finally we use the equation $\partial_{tt}u = \nabla_x^*\nabla_x u - h$ to verify that $\partial_{tt}u\in L^2(I\times\mu_U)$. This suffices to show $u\in H^2(I\times\mu_U)$.

To estimate $\|D^2u\|^2_{L^2(I\times\mu_U)}$ we simply apply Lemma 2.3 and use $\bar\nabla^*\bar\nabla u = h$:

$$\begin{aligned}
\|D^2u\|^2_{L^2(I\times\mu_U)} &\leqslant C_0\Big(\|\bar\nabla^*\bar\nabla u\|^2_{L^2(I\times\mu_U)} + R^2\|\bar\nabla u\|^2_{L^2(I\times\mu_U)}\Big)\\
&\overset{(2.1)}{\leqslant} C_0\big(1+R^2\max\{\frac{1}{m},T^2\}\big)\|h\|^2_{L^2(I\times\mu_U)}.
\end{aligned}$$

□

Remark 2.5. *Lemma 2.4 together with Assumption 3 tells us that $L^2(\mu_U)$ has an orthonormal basis $\{1\}\cup\{w_\lambda\}$ where $w_\lambda\in H^1(\mu_U)$ are eigenfunctions of $\nabla_x^*\nabla_x$ with eigenvalue λ^2 for some $\lambda>0$:*

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

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

We finally need a lemma for the solution of a divergence equation with Dirichlet boundary conditions. This will provide us test functions which play a crucial role in the proof of Theorem 2.

Lemma 2.6. *For any function $f \in L^2(I \times \mu_U)$ with $(f)_{I\times\mu_U} = 0$, there exist $(d+1)$ functions $\bar{\phi} = (\phi_0, \phi_1, \cdots, \phi_d)^\top \in H_0^1(I \times \mu_U)^{d+1}$ such that*

$$\bar{\nabla}^* \bar{\phi} = -\partial_t \phi_0 + \sum_{i=1}^{d} \partial_{x_i}^* \phi_i = f \tag{24}$$

with estimates

$$\|\bar{\phi}\|_{L^2(I\times\mu_U)} \leqslant C_0 \max\{\frac{1}{\sqrt{m}}, T\} \|f\|_{L^2(I\times\mu_U)} \tag{25}$$

and

$$\|\bar{\nabla}\bar{\phi}\|_{L^2(I\times\mu_U)} = \Big(\sum_{i,j=0}^{d} \|\partial_i \phi_j\|_{L^2(I\times\mu_U)}^2\Big)^{1/2} \leqslant C_0\big(1 + \frac{1}{\sqrt{mT}} + \frac{R}{\sqrt{m}} + RT\big)\|f\|_{L^2(I\times\mu_U)}. \tag{26}$$

Here C_0 is a universal constant and R is exactly the constant in Theorem 1.

Proof. Let $\mathbb{H}$ be the subspace of $L^2(I \times \mu_U)$ that consists of "harmonic functions", in other words, $f \in \mathbb{H}$ if and only if $\bar{\nabla}^*\bar{\nabla} f = 0$. 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 $(f^{(2)})_{I\times\mu_U} = 0$ and hence $(f^{(1)})_{I\times\mu_U} = 0$. Therefore by linearity it suffices to consider $f^{(1)}$ and $f^{(2)}$ separately. For $f^{(2)}$, the equation

$$\begin{cases} -\partial_{tt} u + \nabla_x^* \nabla_x u = f^{(2)} & \text{in } I \times \mathbb{R}^d, \\ \partial_t u(t=0,\cdot) = \partial_t u(t=T,\cdot) = 0 & \text{in } \mathbb{R}^d \end{cases} \tag{27}$$

has a unique solution in V by Lemma 2.4. Moreover, for any $v \in \mathbb{H} \cap H^1(I \times \mu_U)$, integration by parts yields

$$\begin{aligned} 0 = \int_{I\times\mathbb{R}^d} f^{(2)} v \,\mathrm{d}t\,\mathrm{d}\mu(x) = B[u,v] \\ = \int_{I\times\mathbb{R}^d} u \bar{\nabla}^*\bar{\nabla} v \,\mathrm{d}t\,\mathrm{d}\mu(x) + \int_{\mathbb{R}^d} \big(u(T)\partial_t v(T) - u(0)\partial_t v(0)\big)\,\mathrm{d}\mu(x) \end{aligned}$$

Therefore, since v is arbitrary, we have $u(T) = u(0) = 0$, meaning $\nabla_x u \in H_0^1(I \times \mu_U)^d$. Also $\partial_t u \in H_0^1(I \times \mu_U)$. Thus for $f^{(2)}$ part, it suffices to take correspondingly $\phi_0^{(2)} = \partial_t u$, $\phi_i^{(2)} = \partial_{x_i} u$ with the estimates

$$\|\bar{\phi}^{(2)}\|_{L^2(I\times\mu_U)}^2 = \|\bar{\nabla} u\|_{L^2(I\times\mu_U)}^2 \overset{(21)}{\leqslant} \max\{\frac{1}{m}, \frac{T^2}{\pi^2}\} \|f^{(2)}\|_{L^2(I\times\mu_U)}^2, \tag{28}$$

and

$$\|\bar{\nabla}\bar{\phi}^{(2)}\|_{L^2(I\times\mu_U)}^2 = \|D^2 u\|_{L^2(I\times\mu_U)}^2 \overset{(22)}{\leqslant} C_0(1 + \frac{R^2}{m} + R^2 T^2) \|f^{(2)}\|_{L^2(I\times\mu_U)}^2. \tag{29}$$

Now we consider the $f^{(1)}$ part. Since $\{1\} \cup \{w_\lambda\}$ forms an orthonormal basis in $L^2(I \times \mu_U)$ and $(f^{(1)})_{I\times\mu_U} = 0$, we have an orthogonal decomposition

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

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

$$0 = \bar{\nabla}^*\bar{\nabla} f^{(1)} = -f_0''(t) + \sum_\lambda (-f_\lambda''(t) + \lambda^2 f_\lambda(t)) w_\lambda(x)$$

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

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

The construction of $\bar\phi^{(0)}$ for $f_0(t)$ is easy: We can simply take $\phi_i^{(0)} = 0$ for $1 \leqslant i \leqslant d$ and $\phi_0^{(0)}(t,x) = \int_0^t f_0(\tau)\,\mathrm{d}\tau$ (so that $\phi_0^{(0)}(T,x) = 0$ since f_0 integrates to 0), and thus

$$\|\bar\phi^{(0)}\|_{L^2(I\times\mu_U)} \leqslant \frac{T}{\pi}\|f_0\|_{L^2(I\times\mu_U)}, \tag{30}$$

and

$$\|\bar\nabla\bar\phi^{(0)}\|_{L^2(I\times\mu_U)} \leqslant \|f_0\|_{L^2(I\times\mu_U)}. \tag{31}$$

Now without loss of generality, we assume $f^{(1)} = e^{\lambda t}w_\lambda(x)$ for a particular λ (the case $f^{(1)} = e^{\lambda(T-t)}w_\lambda(x)$ can be treated similarly by flipping time and the general case follows from a linear combination). Thus $\|f^{(1)}\|^2_{L^2(I\times\mu_U)} = \frac{e^{2\lambda T}-1}{2\lambda}$. We need to find $\bar\phi^{(1)}$ such that

$$-\partial_t\phi_0^{(1)} + \sum_{i=1}^d \partial^*_{x_i}\phi_i^{(1)} = e^{\lambda t}w_\lambda(x).$$

By Lemma 2.4 $w_\lambda \in H^2(I\times\mu_U)$, thus we can take the ansatz $\phi_0^{(1)} = \psi_1(t)w_\lambda(x)$ and for $i = 1,\cdots,d$, $\phi_i^{(1)} = \psi_2(t)\partial_{x_i}w_\lambda(x)$, and the two functions $\psi_1(t),\psi_2(t)$ should satisfy $\psi_1(0) = \psi_1(T) = \psi_2(0) = \psi_2(T) = 0$ as well as the equation

$$-\psi_1'(t) + \lambda^2\psi_2(t) = e^{\lambda t}. \tag{32}$$

Of course there exists infinitely many possible solutions, we will choose a particular one so that $\bar\phi^{(1)}$ satisfies the desired estimates. Let us introduce a short-hand notation $L = e^{\lambda T}$, so $\|f^{(1)}\|^2_{L^2(I\times\mu_U)} = \frac{L^2-1}{2\lambda}$. Let

$$h(x) = \begin{cases} \dfrac{4}{L-1}(x-1), & x\in[1,\dfrac{L+1}{2}];\\ \dfrac{4}{L-1}(L-x), & x\in[\dfrac{L+1}{2},L]. \end{cases}$$

Pick $g(s) = sh(s)$, then $g(1) = g(L) = 0$, and $\int_1^L \frac{g(s)}{s}\,\mathrm{d}s = L-1$. From the expression we can directly derive (using $\lambda \geqslant \sqrt{m}$)

$$g(s) \leqslant 2s \ \text{ and } \ g'(s) \leqslant \frac{L}{L-1} = 1 + \frac{1}{e^{\lambda T}-1} \leqslant 1 + \frac{1}{\sqrt{m}T}.$$

Then pick $\psi_2(t) = \dfrac{1}{\lambda^2}g(e^{\lambda t})$. It is easy to compute

$$\begin{aligned} \|\psi_2\|^2_{L^2(I)} &= \frac{1}{\lambda^4}\int_0^T g(e^{\lambda t})^2\,\mathrm{d}t = \frac{1}{\lambda^5}\int_1^L \frac{g(s)^2}{s}\,\mathrm{d}s \leqslant \frac{2(L^2-1)}{\lambda^5},\\ \text{and } \|\psi_2'\|^2_{L^2(I)} &= \frac{1}{\lambda^2}\int_0^T g'(e^{\lambda t})^2 e^{2\lambda t}\,\mathrm{d}t = \frac{1}{\lambda^3}\int_1^L g'(s)^2 s\,\mathrm{d}s \leqslant \frac{L^2-1}{2\lambda^3}(1+\frac{1}{mT^2}). \end{aligned}$$

Moreover since $\psi_1'(t) = \lambda^2\psi_2(t) - e^{\lambda t}$ from (32),

$$\|\psi_1'\|^2_{L^2(I)} \leqslant 2\lambda^4\|\psi_2\|^2_{L^2(I)} + \frac{e^{2\lambda T}-1}{\lambda} \leqslant \frac{5(L^2-1)}{\lambda}.$$

Finally since

$$|\psi_1(t)| = |\int_0^t (g(e^{\lambda s}) - e^{\lambda s})\,\mathrm{d}s| = \frac{1}{\lambda}|\int_1^{e^{\lambda t}}(\frac{g(\tau)}{\tau}-1)\,\mathrm{d}\tau| \leqslant \frac{3}{\lambda}(e^{\lambda t}-1)$$

we can estimate

$$\lambda^2\|\psi_1\|^2_{L^2(I)} \leqslant 9\int_0^T (e^{\lambda t}-1)^2\,\mathrm{d}t \leqslant 9\frac{L^2-1+2\ln L}{\lambda} \leqslant 18\frac{L^2-1}{\lambda}.$$

Now using $\|w_\lambda\|_{L^2(I\times\mu_U)} = 1$ and $\|\nabla_x w_\lambda\|_{L^2(I\times\mu_U)} = \lambda$, we estimate $\|\bar{\phi}^{(1)}\|_{L^2(I\times\mu_U)}$:

$$\begin{aligned}\|\bar{\phi}^{(1)}\|^2_{L^2(I\times\mu_U)} &= \|\phi_0^{(1)}\|^2_{L^2(I\times\mu_U)} + \sum_{i=1}^d \|\phi_i^{(1)}\|^2_{L^2(I\times\mu_U)} \\ &= \|\psi_1 w_\lambda\|^2_{L^2(I\times\mu_U)} + \|\psi_2 \nabla_x w_\lambda\|^2_{L^2(I\times\mu_U)} \\ &= \|\psi_1\|^2_{L^2(I)} + \lambda^2\|\psi_2\|^2_{L^2(I)} \leqslant \frac{C}{\lambda^2}\|f^{(1)}\|^2_{L^2(I\times\mu_U)} \leqslant \frac{C}{m}\|f^{(1)}\|^2_{L^2(I\times\mu_U)}.\end{aligned} \tag{33}$$

Combining (28), (30) and (33), we can derive (25).

Now we shift our focus to $\|\bar{\nabla}\bar{\phi}^{(1)}\|_{L^2(I\times\mu_U)}$. By Lemma 2.3 and $\bar{\nabla}^*\bar{\nabla} w_\lambda = \lambda^2 w_\lambda$,

$$\begin{aligned}\|\bar{\nabla}\bar{\phi}^{(1)}\|^2_{L^2(I\times\mu_U)} &= \|\bar{\nabla}\phi_0^{(1)}\|^2_{L^2(I\times\mu_U)} + \sum_{i=1}^d \|\bar{\nabla}\phi_i^{(1)}\|^2_{L^2(I\times\mu_U)} \\ &= \|\psi_1' w_\lambda\|^2_{L^2(I\times\mu_U)} + \|\psi_1\nabla_x w_\lambda\|^2_{L^2(I\times\mu_U)} + \|\psi_2'\nabla_x w_\lambda\|^2_{L^2(I\times\mu_U)} + \|\psi_2 D_x^2 w_\lambda\|^2_{L^2(I\times\mu_U)} \\ &= \|\psi_1'\|^2_{L^2(I)} + \lambda^2\|\psi_1\|^2_{L^2(I)} + \lambda^2\|\psi_2'\|^2_{L^2(I)} + \|\psi_2\|^2_{L^2(I)}(\lambda^4 + R^2\lambda^2) \\ &\leqslant C_0(1 + \frac{1}{mT^2} + \frac{R^2}{\lambda^2})\|f^{(1)}\|^2_{L^2(I\times\mu_U)}\end{aligned}$$

Combining above with (29) and (31), and noticing $\frac{1}{\lambda^2} \leqslant \frac{1}{m}$, we finish the proof for (26). □

Proof of Theorem 2. Without loss of generality, assume $(f)_{I\times\mu_U} = 0$. We start with

$$\|f - \Pi_v f\|_{L^2(I\times\mu_U;L^2_\kappa)} \leqslant \|\nabla_v f\|_{L^2(I\times\mu_U;L^2_\kappa)}. \tag{34}$$

The proof is simple: for every x we have Gaussian Poincaré inequality, and then integrate over $(t,x) \in I\times\mathbb{R}^d$ against $\mathrm{d}t\,\mathrm{d}\mu_U(x)$.

Obviously $\int_{I\times\mathbb{R}^d} \Pi_v f\,\mathrm{d}t\,\mathrm{d}\mu_U(x) = (f)_{I\times\mu_U} = 0$. Therefore, we can take ϕ_i as in Lemma 2.6 with $\Pi_v f$ in place of f, so that $\bar{\nabla}^*\bar{\phi} = \Pi_v f$. The trick in our following step is to introduce v variable in the calculation. Notice by Gaussianity

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

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

$$\begin{aligned}&\|\Pi_v f\|^2_{L^2(I\times\mu_U)} = \int_{I\times\mathbb{R}^d} \Pi_v f\bar{\nabla}^*\bar{\phi}\,\mathrm{d}t\,\mathrm{d}\mu_U(x) \\ &= \int_{I\times\mathbb{R}^d\times\mathbb{R}^d} \Pi_v f(-\partial_t\phi_0 + v\cdot\nabla_x\phi_0 + v\cdot\partial_t\phi - \sum_i v_i v\cdot\partial_{x_i}\phi + \phi\cdot\nabla_x U)\,\mathrm{d}t\,\mathrm{d}\rho_\infty(x,v) \\ &= \int_{I\times\mathbb{R}^d\times\mathbb{R}^d} f(-\partial_t\phi_0 + v\cdot\nabla_x\phi_0 + v\cdot\partial_t\phi - \sum_i v_i v\cdot\partial_{x_i}\phi + \phi\cdot\nabla_x U)\,\mathrm{d}t\,\mathrm{d}\rho_\infty(x,v) \\ &\quad - \int_{I\times\mathbb{R}^d\times\mathbb{R}^d} (-\partial_t\phi_0 + v\cdot\nabla_x\phi_0 + v\cdot\partial_t\phi - \sum_i v_i v\cdot\partial_{x_i}\phi + \phi\cdot\nabla_x U)(f - \Pi_v f)\,\mathrm{d}t\,\mathrm{d}\rho_\infty(x,v).\end{aligned} \tag{35}$$

For the first integral on the right hand side, we use integration by parts:

$$\begin{aligned}&\int_{I\times\mathbb{R}^d\times\mathbb{R}^d} f(-\partial_t\phi_0 + v\cdot\nabla_x\phi_0 + v\cdot\partial_t\phi - \sum_i v_i v\cdot\partial_{x_i}\phi + \phi\cdot\nabla_x U)\,\mathrm{d}t\,\mathrm{d}\rho_\infty(x,v) \\ &\quad = \int_{I\times\mathbb{R}^d\times\mathbb{R}^d} \Big(\partial_t f\phi_0 - \partial_t f(v\cdot\phi) - \phi_0(v\cdot\nabla_x f) + f\phi_0(v\cdot\nabla_x U) \\ &\qquad + (v\cdot\nabla_x f)(v\cdot\phi) - f(v\cdot\phi)(v\cdot\nabla_x U) + f\phi\cdot\nabla_x U\Big)\,\mathrm{d}t\,\mathrm{d}\rho_\infty(x,v)\end{aligned}$$

$$
\begin{aligned}
&= \int_{I\times\mathbb{R}^d\times\mathbb{R}^d} \Big(\partial_t f\phi_0 - \partial_t f(v\cdot\phi) - \phi_0(v\cdot\nabla_x f) + \phi_0(\nabla_v f\cdot\nabla_x U)\\
&\qquad + (v\cdot\nabla_x f)(v\cdot\phi) - \nabla_v\cdot((v\cdot\phi)f\nabla_x U) + f\phi\cdot\nabla_x U\Big)\,\mathrm{d}t\,\mathrm{d}\rho_\infty(x,v)\\
&= \int_{I\times\mathbb{R}^d\times\mathbb{R}^d} \Big((\partial_t f - v\cdot\nabla_x f + \nabla_x U\cdot\nabla_v f)(\phi_0 - v\cdot\phi)\Big)\,\mathrm{d}t\,\mathrm{d}\rho_\infty(x,v)\\
&\leqslant \|\partial_t f - \mathcal{L}_{\mathrm{ham}} f\|_{L^2(I\times\mu_U;H_\kappa^{-1})}\|\phi_0 - v\cdot\phi\|_{L^2(I\times\mu_U;H_\kappa^1)}.
\end{aligned}
$$

We further estimate the term $\|\phi_0 - v\cdot\phi\|_{L^2(I\times\mu_U;H^1_\kappa)}$:

$$
\begin{aligned}
\|\phi_0 - v\cdot\phi\|^2_{L^2(I\times\mu_U;H^1_\kappa)} &= \int_{I\times\mathbb{R}^d} \|\phi_0 - v\cdot\phi\|^2_{H^1_\kappa}\,\mathrm{d}t\,\mathrm{d}\mu_U(x)\\
&= \int_{I\times\mathbb{R}^d} \Big(\|\phi_0 - v\cdot\phi\|^2_{L^2_\kappa} + \|\nabla_v(\phi_0 - v\cdot\phi)\|^2_{L^2_\kappa}\Big)\,\mathrm{d}t\,\mathrm{d}\mu_U(x)\\
&= \int_{I\times\mathbb{R}^d} \Big(\int_{\mathbb{R}^d}(\phi_0 - v\cdot\phi)^2\,\mathrm{d}\kappa(v) + \int_{\mathbb{R}^d}|\phi|^2\,\mathrm{d}\kappa(v)\Big)\,\mathrm{d}t\,\mathrm{d}\mu_U(x)\\
&= \int_{I\times\mathbb{R}^d} (\phi_0^2 + 2|\phi|^2)\,\mathrm{d}t\,\mathrm{d}\mu_U(x) \overset{(25)}{\leqslant} C_0(\frac{1}{m} + T^2)\}\|\Pi_v f\|^2_{L^2(I\times\mu_U)}.
\end{aligned}
$$

For the second integral in (35), we estimate (again using the convention $\partial_{x_0} = \partial_t$)

$$
\begin{aligned}
&\| -\partial_t\phi_0 + v\cdot\nabla_x\phi_0 + v\cdot\partial_t\phi - \sum_i v_i v\cdot\partial_{x_i}\phi + \phi\cdot\nabla_x U\|^2_{L^2(I\times\mu_U;L^2_\kappa)}\\
&= \int_{I\times\mathbb{R}^d\times\mathbb{R}^d} (\partial_t\phi_0 - v\cdot\nabla_x\phi_0 - v\cdot\partial_t\phi + \sum_i v_i v\cdot\partial_{x_i}\phi - \phi\cdot\nabla_x U)^2\,\mathrm{d}t\,\mathrm{d}\rho_\infty(x,v)\\
&= \int_{I\times\mathbb{R}^d\times\mathbb{R}^d} \Big((\partial_t\phi_0 - \phi\cdot\nabla_x U)^2 - 2(\partial_t\phi_0 - \phi\cdot\nabla_x U)(v\cdot\nabla_x\phi_0) - 2(\partial_t\phi_0 - \phi\cdot\nabla_x U)(v\cdot\partial_t\phi)\\
&\qquad + (v\cdot\nabla_x\phi_0)^2 + (v\cdot\partial_t\phi)^2 + 2(\partial_t\phi_0 - \phi\cdot\nabla_x U)\sum_{i,j} v_i v_j\partial_{x_i}\phi_j + 2(v\cdot\partial_t\phi)(v\cdot\nabla_x\phi_0)\\
&\qquad + \sum_{i,j,k,l} v_i v_j v_k v_l\partial_{x_i}\phi_j\partial_{x_k}\phi_l - 2\sum_{i,j,k} v_i v_j v_k\partial_t\phi_k\partial_{x_i}\phi_j - 2\sum_{i,j,k} v_i v_j v_k\partial_{x_k}\phi_0\partial_{x_i}\phi_j\Big)\,\mathrm{d}t\,\mathrm{d}\rho_\infty(x,v)\\
&= \int_{I\times\mathbb{R}^d\times\mathbb{R}^d} \Big((\partial_t\phi_0 - \phi\cdot\nabla_x U)^2 + \sum_i v_i^2\big((\partial_{x_i}\phi_0)^2 + (\partial_t\phi_i)^2 + 2\partial_{x_i}\phi_0\partial_t\phi_i\big)\\
&\qquad + 2(\partial_t\phi_0 - \phi\cdot\nabla_x U)\sum_i v_i^2\partial_{x_i}\phi_i + \sum_i v_i^4(\partial_{x_i}\phi_i)^2 + \sum_{i\neq j} v_i^2 v_j^2(\partial_{x_i}\phi_j)^2\\
&\qquad + \sum_{i\neq j} v_i^2 v_j^2\partial_{x_i}\phi_i\partial_{x_j}\phi_j + \sum_{i\neq j} v_i^2 v_j^2\partial_{x_i}\phi_j\partial_{x_j}\phi_i\Big)\,\mathrm{d}t\,\mathrm{d}\rho_\infty(x,v)\\
&= \int_{I\times\mathbb{R}^d} \Big((\partial_t\phi_0 - \phi\cdot\nabla_x U)^2 + |\nabla_x\phi_0|^2 + |\partial_t\phi|^2 + 2\nabla_x\phi_0\cdot\partial_t\phi + 2(\partial_t\phi_0 - \phi\cdot\nabla_x U)\sum_i\partial_{x_i}\phi_i\\
&\qquad + 3\sum_i(\partial_{x_i}\phi_i)^2 + \sum_{i\neq j}(\partial_{x_i}\phi_j)^2 + \sum_{i\neq j}\partial_{x_i}\phi_i\partial_{x_j}\phi_j + \sum_{i\neq j}\partial_{x_i}\phi_j\partial_{x_j}\phi_i\Big)\,\mathrm{d}t\,\mathrm{d}\mu_U(x)\\
&\leqslant \int_{I\times\mathbb{R}^d} \Big((\partial_t\phi_0 - \phi\cdot\nabla_x U + \sum_i\partial_{x_i}\phi_i)^2 + 2\sum_{(i,j)\neq(0,0)}^{d}|\partial_{x_i}\phi_j|^2\Big)\,\mathrm{d}t\,\mathrm{d}\mu_U(x)\\
&\overset{(27),(17)}{=} \|\Pi_v f\|^2_{L^2(I\times\mu_U)} + 2\|\bar\nabla\bar\phi\|^2_{L^2(I\times\mu_U)} \leqslant C_0(1 + \frac{1}{\sqrt{m}T} + \frac{R}{\sqrt{m}} + RT)\|\Pi_v f\|^2_{L^2(I\times\mu_U)}.
\end{aligned}
$$

Combining the above estimates, we arrive at

$$
\|\Pi_v f\|^2_{L^2(I\times\mu_U)} \leqslant \|\partial_t f - \mathcal{L}_{\mathrm{ham}} f\|_{L^2(I\times\mu_U;H^{-1}_\kappa)}\|\phi_0 - v\cdot\phi\|_{L^2(I\times\mu_U;H^1_\kappa)}
$$

$$+ \| -\partial_t\phi_0 + v\cdot\nabla_x\phi_0 + v\cdot\partial_t\phi - \sum_i v_i v\cdot\partial_{x_i}\phi + \phi\cdot\nabla_x U\|_{L^2(I\times\mu_U;L^2_\kappa)}\|f-\Pi_v f\|_{L^2(I\times\mu_U;L^2_\kappa)}$$

$$\leqslant C_0\Big((\frac{1}{\sqrt{m}} + T\|\partial_t f - \mathcal{L}_{\mathrm{ham}} f\|_{L^2(I\times\mu_U;H^{-1}_\kappa)}\|\Pi_v f\|_{L^2(I\times\mu_U)}$$

$$+ (1+\frac{1}{\sqrt{m}T} + \frac{R}{\sqrt{m}} + RT)\|\nabla_v f\|_{L^2(I\times\mu_U)}\|\Pi_v f\|_{L^2(I\times\mu_U)}\Big).$$

Finally

$$\|f\|_{L^2(I\times\mu_U;L^2_\kappa)} \leqslant \|f-\Pi_v f\|_{L^2(I\times\mu_U;L^2_\kappa)} + \|\Pi_v f\|_{L^2(I\times\mu_U)}$$

$$\leqslant C_0\Big((\frac{1}{\sqrt{m}} + T)\|\partial_t f - \mathcal{L}_{\mathrm{ham}} f\|_{L^2(I\times\mu_U;H^{-1}_\kappa)} + (1+\frac{1}{\sqrt{m}T} + \frac{R}{\sqrt{m}} + RT)\|\nabla_v f\|_{L^2(I\times\mu_U)}\Big),$$

as claimed. ☐

With Theorem 2, it is easy to prove exponential decay to equilibrium.

Proof of Theorem 1. We first notice that (10) implies

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

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

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

using the equation (4) and integration by parts.

For every $0<s<t<T$, we have the typical energy estimate:

$$\|f(t,\cdot)\|^2_{L^2(\mu_U;L^2_\kappa)} - \|f(s,\cdot)\|^2_{L^2(\mu_U;L^2_\kappa)} = -2\gamma\|\nabla_v f\|^2_{L^2((s,t)\times\mu_U;L^2_\kappa)}. \tag{36}$$

In particular,

$$\text{the mapping } t\mapsto \|f(t,\cdot)\|^2_{L^2(\mu_U;L^2_\kappa)} \text{ is nonincreasing.} \tag{37}$$

Since by equation (4),

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

we have

$$\|\partial_t f - \mathcal{L}_{\mathrm{ham}} f\|_{L^2((s,t)\times\mu_U,H^{-1}_\kappa)} = \gamma\|\nabla_v^*\nabla_v f\|_{L^2((s,t)\times\mu_U,H^{-1}_\kappa)} \leqslant \gamma\|\nabla_v f\|_{L^2((s,t)\times\mu_U,L^2_\kappa)}.$$

Fix some time interval length $t_0>0$ that will be specified later. Denote $b_1 = C_0(\frac{1}{\sqrt{m}} + t_0)$ and $b_2 = C_0(1+\dfrac{1}{\sqrt{m}t_0} + \dfrac{R}{\sqrt{m}} + Rt_0)$, and thus by Theorem 2, (36) and (37),

$$\|f(t,\cdot)\|^2_{L^2(\mu_U;L^2_\kappa)} - \|f(t-t_0,\cdot)\|^2_{L^2(\mu_U;L^2_\kappa)}$$

$$\leqslant -\frac{2\gamma}{(b_1\gamma+b_2)^2}\Big(b_2\|\nabla_v f\|_{L^2((t-t_0,t)\times\mu_U,L^2_\kappa)} + b_1\|\partial_t f - \mathcal{L}_{\mathrm{ham}} f\|_{L^2((t-t_0,t)\times\mu_U,H^{-1}_\kappa)}\Big)^2$$

$$\leqslant -\frac{2\gamma}{(b_1\gamma+b_2)^2}\|f\|^2_{L^2((t-t_0,t)\times\mu_U,L^2_\kappa)} \leqslant -\frac{2\gamma t_0}{(b_1\gamma+b_2)^2}\|f(t,\cdot)\|^2_{L^2(\mu_U,L^2_\kappa)}.$$

Now for any $t>0$, pick the integer k satisfying $kt_0\leqslant t<(k+1)t_0$. Applying the above, we get

$$\|f(t,\cdot)\|^2_{L^2(\mu_U;L^2_\kappa)} \leqslant \Big(1+\frac{2\gamma t_0}{(b_1\gamma+b_2)^2}\Big)^{-k}\|f(0,\cdot)\|^2_{L^2(\mu_U;L^2_\kappa)}$$

$$\leqslant \Big(1+\frac{2\gamma t_0}{(b_1\gamma+b_2)^2}\Big)^{-\frac{t}{t_0}+1}\|f(0,\cdot)\|^2_{L^2(\mu_U;L^2_\kappa)}$$

$$= \Big(1+\frac{2\gamma t_0}{(b_1\gamma+b_2)^2}\Big)\exp\Big(-\frac{t}{t_0}\log\big(1+\frac{2\gamma t_0}{(b_1\gamma+b_2)^2}\big)\Big)\|f(0,\cdot)\|^2_{L^2(\mu_U;L^2_\kappa)}.$$

Picking t_0 to optimize the convergence rate $\frac{1}{2t_0}\log(1+\frac{2\gamma t_0}{(b_1\gamma+b_2)^2})$ might be difficult, we will simply choose $t_0=\frac{1}{2\sqrt{m}}$, so that there exists some universal constant c_0 such that

$$\|f(t,\cdot)\|_{L^2(\mu_U;L^2_\kappa)} \leqslant \sqrt{1+\frac{\gamma\sqrt{m}}{c_0(\sqrt{m}+R+\gamma)^2}}\exp\Big(-t\sqrt{m}\log\big(1+\frac{\gamma\sqrt{m}}{c_0(\sqrt{m}+R+\gamma)^2}\big)\Big)\|f(0,\cdot)\|_{L^2(\mu_U;L^2_\kappa)}$$

and we finish the proof since the expression under the square root is bounded above by a constant. □

Acknowledgment

This research is supported in part by National Science Foundation via grants DMS-1454939. We would like to thank Rong Ge, Yulong Lu, Jonathan Mattingly, Jean-Christophe Mourrat, and Gabriel Stoltz for helpful discussions, and thank Felix Otto for discussions and providing an idea leading to the proof of Lemma 2.6. LW would also like to thank Scott Armstrong [1] and Nicola Gigli [11] for answering our question on MathOverflow, which lead to our proof of Lemma 2.4 (ii).

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.

Theorem 3 ([20, Theorem 6.4])**.** *When $U(x)=\frac{m}{2}|x|^2$, $d=1$, the spectrum of the operator $-\mathcal{L}$ is given by*

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

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

$$\lambda_{\text{exact}} = \mathfrak{Re}\Big(\frac{\gamma}{2}-\frac{\sqrt{\gamma^2-4m}}{2}\Big). \tag{38}$$

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 [7, 8], adapted and summarized for underdamped Langevin equation in [21, Sec. 2]. In the first part of this section, we will review the main result based on [21]; in addition, we will provide a new estimate of the operator norm of $\|\mathcal{A}\mathcal{L}_{\text{ham}}(1-\Pi_v)\|_{L^2_{\rho_\infty}\to L^2_{\rho_\infty}}$, 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 $\nabla_x^2 U \geqslant -2\,\text{Id}$.

B.1. **Revisiting the DMS hypocoercive estimation in $L^2_{\rho_\infty}$.** Let us first define an operator

$$\mathcal{A} = \left(1+(\mathcal{L}_{\text{ham}}\Pi_v)^*(\mathcal{L}_{\text{ham}}\Pi_v)\right)^{-1}(\mathcal{L}_{\text{ham}}\Pi_v)^* \tag{39}$$

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

$$\mathsf{E}(\phi) = \frac{1}{2}\|\phi\|^2_{L^2_{\rho_\infty}} - \epsilon\,(\mathcal{A}\phi,\phi)_{L^2_{\rho_\infty}}, \tag{40}$$

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., [21, Eq. (17)]),

$$\frac{1-|\epsilon|}{2}\,\|\phi\|^2_{L^2_{\rho_\infty}} \leqslant \mathsf{E}(\phi) \leqslant \frac{1+|\epsilon|}{2}\,\|\phi\|^2_{L^2_{\rho_\infty}}. \tag{41}$$

Theorem 4 (See [21, Theorem 1])**.** *Assume that the Poincaré inequality (8) holds and there exists* $\mathsf{R}_{ham} < \infty$ *such that*

$$\|\mathcal{A}\mathcal{L}_{ham}(1-\Pi_v)\|_{L^2_{\rho_\infty}\to L^2_{\rho_\infty}} \leqslant \mathsf{R}_{ham}. \tag{42}$$

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

$$\lambda_{DMS} := \frac{\gamma - \frac{\epsilon}{1+m} - \sqrt{\epsilon^2(\mathsf{R}_{ham} + \frac{\gamma}{2})^2 + \left(\gamma - \frac{2m+1}{m+1}\epsilon\right)^2}}{2(1+|\epsilon|)}. \tag{43}$$

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 $\lambda_{\mathrm{DMS}} = 0$, which reduces to the conclusion that $\|f(t,\cdot,\cdot)\|_{L^2_{\rho_\infty}}$ is non-increasing in time t. The existence of $\mathsf{R}_{\mathrm{ham}}$ has been studied under fairly general assumptions on the potential $U(x)$ in [8, Sec. 2]. In the Proposition B.1 below, we provide a simpler estimation of $\mathsf{R}_{\mathrm{ham}}$ 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 [8, Lemma 4]; the simplicity in our approach comes from the application of *Bochner's formula*. It is interesting to observe that $\mathsf{R}_{\mathrm{ham}}$ does not depend on m when U is an isotropic quadratic potential.

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

$$\mathsf{R}_{ham} = \sqrt{\max\{K, 2\}}. \tag{44}$$

such that (42) is satisfied.

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

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

Thus the optimal choice of R_{ham} *is* $\sqrt{2}$ *and (44) is tight in this case.*

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

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

$$\begin{aligned}\Lambda_{\mathrm{DMS}}(\gamma, m) &:= \sup_{\epsilon\in(-1,1)} \lambda_{\mathrm{DMS}}(\gamma, m, \sqrt{2}, \epsilon)\\ &= \sup_{\epsilon\in(-1,1)} \frac{\gamma - \frac{\epsilon}{1+m} - \sqrt{\epsilon^2(\sqrt{2} + \frac{\gamma}{2})^2 + \left(\gamma - \frac{2m+1}{m+1}\epsilon\right)^2}}{2(1+|\epsilon|)},\end{aligned} \tag{45}$$

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

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

Therefore, the supremum can only be achieved at $\epsilon = 1^-$, or the critical points of the expression on the right hand side of (45). In general, it is hard to obtain a simple explicit expression of $\Lambda_{\mathrm{DMS}}(\gamma, m)$. Therefore, we shall consider the following asymptotic regions.

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

$$\Lambda_{DMS}(\gamma, m) = \begin{cases} \Big(\dfrac{-(1+m)\sqrt{3m^2+4m+1}+3m^2+3m+1}{6m^2+8m+3}\Big)\gamma + \mathcal{O}(\gamma^2), & \text{when } \gamma \to 0; \\ \dfrac{4m^2}{(1+m)^2}\gamma^{-1} + \mathcal{O}(\gamma^{-2}), & \text{when } \gamma \to \infty. \end{cases} \tag{46}$$

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

$$\Lambda_{DMS}(\gamma, m) = \begin{cases} \dfrac{\gamma^5}{2b^4} + \mathcal{O}(\gamma^6), & \text{when } \gamma \to 0; \\ \dfrac{4}{\gamma} + \mathcal{O}(\gamma^{-2}), & \text{when } \gamma \to \infty. \end{cases} \tag{47}$$

The proof can be found in Appendix B.3. The scaling in the first case is already known in e.g., [21]; 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

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

Notice that this operator is supported on $\mathrm{Ran}(\Pi_v)$ from the observation that $\mathcal{A} = \Pi_v\mathcal{A}$, it is then equivalent to find the smallest $\mathsf{R}_{\mathrm{ham}}$ 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

$$\left\|-(1-\Pi_v)\mathcal{L}_{\mathrm{ham}}\mathcal{A}^*\phi\right\|_{L^2_{\rho_\infty}} \leqslant \mathsf{R}_{\mathrm{ham}}\left\|\phi\right\|_{L^2_{\rho_\infty}} = \mathsf{R}_{\mathrm{ham}}\left\|\phi\right\|_{L^2_{\mu_U}}. \tag{48}$$

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

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

It is easy to check that $\Pi_v\varphi = \varphi$. By simplifying the above equation with (5) and (14),

$$\phi(x) = \varphi(x) - \Delta_x\varphi(x) + \nabla_x U(x)\cdot\nabla_x\varphi = \varphi(x) + \nabla_x^*\nabla_x\varphi(x). \tag{49}$$

Furthermore, by some straightforward calculation, we have

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

Thus

$$\begin{aligned}\left\|-(1-\Pi_v)\mathcal{L}_{\mathrm{ham}}\mathcal{A}^*\phi\right\|^2_{L^2_{\rho_\infty}} &= \int\Big(\sum_{i,j}(v_iv_j-\delta_{i,j})\partial_{x_i,x_j}\varphi\Big)^2\,\mathrm{d}\rho_\infty \\ &= 2\sum_{i,j}\int\left(\partial_{x_i,x_j}\varphi\right)^2\,\mathrm{d}\mu_U.\end{aligned}$$

Then by Bochner's formula,

$$\begin{aligned}&\left\|-(1-\Pi_v)\mathcal{L}_{\mathrm{ham}}\mathcal{A}^*(\phi)\right\|^2_{L^2_{\rho_\infty}} \\ &\quad= 2\int\nabla_x\varphi\cdot\nabla_x\nabla_x^*\nabla_x\varphi - \nabla_x\varphi\cdot\nabla_x^2U\nabla_x\varphi - \nabla_x^*\nabla_x\left(\frac{|\nabla_x\varphi|^2}{2}\right)\mathrm{d}\mu_U\end{aligned}$$

$$\begin{aligned}
&= 2\int |\nabla_x^*\nabla_x\varphi|^2 - \nabla_x\varphi\cdot\nabla_x^2 U\nabla_x\varphi \,\mathrm{d}\mu_U \\
&\leqslant 2\left(\int |\nabla_x^*\nabla_x\varphi|^2 \,\mathrm{d}\mu_U + K\int |\nabla_x\varphi|^2\,\mathrm{d}\mu_U\right) \\
&\leqslant \max\{K,2\}\left(\int |\nabla_x^*\nabla_x\varphi|^2\,\mathrm{d}\mu_U + 2\int |\nabla_x\varphi|^2\,\mathrm{d}\mu_U\right).
\end{aligned}$$

From (49), we have

$$\begin{aligned}
\|\phi\|_{L^2_{\mu_U}}^2 &= \int \varphi^2 + 2\varphi\ \nabla_x^*\nabla_x\varphi + \left|\nabla_x^*\nabla_x\varphi\right|^2\ \mathrm{d}\mu_U \\
&\geqslant 2\int |\nabla_x\varphi|^2\,\mathrm{d}\mu_U + \int |\nabla_x^*\nabla_x\varphi|^2\,\mathrm{d}\mu_U.
\end{aligned}$$

By combining the last two equations,

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

which yields (44).

We now consider the isotropic case. Recall that the operator norm of $\mathcal{A}\mathcal{L}_{\mathrm{ham}}(1-\Pi_v)$ is the smallest $\mathsf{R}_{\mathrm{ham}}$ such that (48) holds. Let us consider the elliptic PDE (49). By the choice $U(x) = \frac{m}{2}|x|^2$,

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

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

$$\bar\phi(y) = \Big(1 + m(y-\nabla_y)\cdot\nabla_y\Big)\bar\varphi(y). \tag{50}$$

In addition, by rewriting (48), we need to find the smallest $\mathsf{R}_{\mathrm{ham}}$ such that

$$2m^2\sum_{i,j}\int \left|\partial_{y_i,y_j}\bar\varphi(y)\right|^2\ e^{-\frac{|y|^2}{2}}\,\mathrm{d}y \leqslant \mathsf{R}_{\mathrm{ham}}^2 \int \bar\phi(y)^2 e^{-\frac{|y|^2}{2}}\,\mathrm{d}y. \tag{51}$$

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

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

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 $\bar\varphi = H_{\boldsymbol{n}}$, then $\bar\phi = N_{\boldsymbol{n}}H_{\boldsymbol{n}}$, where $N_{\boldsymbol{n}} := 1 + m\sum_i n_i$. Thus if $\bar\varphi(y) = \sum_{\boldsymbol{n}} a_{\boldsymbol{n}}H_{\boldsymbol{n}}$, then we have $\bar\phi = \sum_n a_{\boldsymbol{n}}N_{\boldsymbol{n}}H_{\boldsymbol{n}}$. By such an expansion, (51) can be rewritten as

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

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

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

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

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

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

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

Proof of Proposition B.2. We used Maple software to help verify the asymptotic expansion.

Part (i): $m = \mathcal{O}(1)$**.**

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

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

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

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

After comparison, the larger decay rate is obtained at ϵ_- with the value in (46).

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

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

Thus we need to consider the critical points. It turns out, there is only one critical point within the domain $(-1,1)$, which is $\epsilon = \frac{8m}{1+m}\gamma^{-1} + \mathcal{O}(\gamma^{-2})$ with the decay rate in (46).

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

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

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

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

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

and the associated decay rates are

$$\lambda_{\text{DMS}}(\gamma, m = (\gamma/b)^2, \sqrt{2}, \epsilon_1) = \frac{\gamma^5}{2b^4} + \mathcal{O}(\gamma^6) > 0;$$
$$\lambda_{\text{DMS}}(\gamma, m = (\gamma/b)^2, \sqrt{2}, \epsilon_2) = -\frac{1}{3}\gamma + \mathcal{O}(\gamma^2) < 0.$$

Therefore, the optimal decay rate is obtained at ϵ_1, which gives (47).

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

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

Thus the supremum in (45) 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 $\epsilon_1 = \frac{8}{\gamma} + \mathcal{O}(\gamma^{-2})$. The optimal decay rate must be achieved at ϵ_1, with the expression given in (47). □

References

1. Scott Armstrong, *Answer to "Elliptic regularity with Gibbs measure satisfying Bakry-Emery condition"*, MathOverflow, https://mathoverflow.net/q/335599 (version: 2019-07-06).
2. Scott Armstrong and Jean-Christophe Mourrat, *Variational methods for the kinetic Fokker-Planck equation*, arXiv preprint arXiv:1902.04037 (2019).
3. Dominique Bakry and Michel Émery, *Diffusions hypercontractives*, Séminaire de Probabilités XIX 1983/84, Springer, 1985, pp. 177–206.
4. Fabrice Baudoin, *Bakry-Emery meet Villani*, arXiv:1308.4938 [math-ph] (2013), arXiv: 1308.4938.
5. ______, *Wasserstein contraction properties for hypoelliptic diffusions*, arXiv:1602.04177 [math] (2016).
6. Arnak S Dalalyan and Lionel Riou-Durand, *On sampling from a log-concave density using kinetic Langevin diffusions*, arXiv preprint arXiv:1807.09382 (2018).
7. Jean Dolbeault, Clément Mouhot, and Christian Schmeiser, *Hypocoercivity for kinetic equations with linear relaxation terms*, Comptes Rendus Mathematique **347** (2009), no. 9, 511–516.
8. ______, *Hypocoercivity for linear kinetic equations conserving mass*, Transactions of the American Mathematical Society **367** (2015), no. 6, 3807–3828.
9. Andreas Eberle, Arnaud Guillin, and Raphael Zimmer, *Couplings and quantitative contraction rates for Langevin dynamics*, The Annals of Probability **47** (2019), no. 4, 1982–2010.
10. J.-P. Eckmann and M. Hairer, *Spectral properties of hypoelliptic operators*, Communications in Mathematical Physics **235** (2003), no. 2, 233–253.
11. Nicola Gigli, *Answer to "Elliptic regularity with Gibbs measure satisfying Bakry-Emery condition"*, MathOverflow, https://mathoverflow.net/q/335608 (version: 2019-07-06).
12. Martin Grothaus and Patrik Stilgenbauer, *Hypocoercivity for Kolmogorov backward evolution equations and applications*, Journal of Functional Analysis **267** (2014), no. 10, 3515–3556.
13. James G Hooton, *Compact sobolev imbeddings on finite measure spaces*, Journal of Mathematical Analysis and Applications **83** (1981), no. 2, 570–581.
14. A. Iacobucci, S. Olla, and G. Stoltz, *Convergence rates for nonequilibrium Langevin dynamics*, Annales mathématiques du Québec **43** (2019), no. 1, 73–98.
15. S. M. Kozlov, *Effective diffusion in the Fokker-Planck equation*, Mathematical notes of the Academy of Sciences of the USSR **45** (1989), no. 5, 360–368.
16. Michel Ledoux, *A simple analytic proof of an inequality by P. Buser*, Proceedings of the American Mathematical Society **121** (1994), no. 3, 951–959.
17. ______, *Spectral gap, logarithmic Sobolev constant, and geometric bounds*, Surveys in Differential Geometry **9** (2004), no. 1, 219–240.
18. J. C. Mattingly, A. M. Stuart, and D. J. Higham, *Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise*, Stochastic Processes and their Applications **101** (2002), no. 2, 185–232.
19. Felix Otto and Cédric Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality*, Journal of Functional Analysis **173** (2000), no. 2, 361–400.
20. Grigorios A Pavliotis, *Stochastic processes and applications: diffusion processes, the Fokker-Planck and Langevin equations*, vol. 60, Springer, 2014.
21. Julien Roussel and Gabriel Stoltz, *Spectral methods for Langevin dynamics and associated error estimates*, ESAIM: Mathematical Modelling and Numerical Analysis **52** (2018), no. 3, 1051–1083.
22. I. Sason and S. Verdú, *f-divergence inequalities*, IEEE Transactions on Information Theory **62** (2016), no. 11, 5973–6006.
23. Gabriel Stoltz and Zofia Trstanova, *Langevin dynamics with general kinetic energies*, Multiscale Modeling & Simulation **16** (2018), no. 2, 777–806.
24. Gabriel Stoltz and Eric Vanden-Eijnden, *Longtime convergence of the temperature-accelerated molecular dynamics method*, Nonlinearity **31** (2018), no. 8, 3748–3769.
25. Cédric Villani, *Hypocoercive diffusion operators*, Bollettino dell'Unione Matematica Italiana **10-B** (2007), no. 2, 257–275.
26. ______, *Hypocoercivity*, Memoirs of the American Mathematical Society **202** (2009), no. 950.

Department of Mathematics, Duke University, Durham NC 27708
E-mail address: yucao@math.duke.edu

Department of Mathematics, Department of Physics, and Department of Chemistry, Duke University, Durham NC 27708
E-mail address: jianfeng@math.duke.edu

Department of Mathematics, Duke University, Durham NC 27708
E-mail address: lihan@math.duke.edu