Notes of Geometric Ergodicity for SDEs

Xuda Ye

Peking University

March 10, 2022

1 Harris ergodic theorem

1.1 Inrtoduction

We introduce the Harris ergodic theorem for Markov chain, which is a standard probabilistic approach to study the ergodicity of Markov chains and Markov chains. The *geometric ergodicity* depicts to how fast the distribution law of a Markov chain converges to its invariant distribution. The theorem states that it only requires the minorization condition and the Lyapunov function to prove the geometric ergodicity.

Harris ergodic theorem was firstly proposed by Harris [1], then developed by Meyn and Tweedie [2] in their monograph *Markov Chains and Stochastic Stability*. Mattingly [3] employed the theorem to prove the geometric ergodicity of the second-order Langevin dynamics, which later becomes the standard approach to study the geometric ergodicity and the long time behavior of SDEs. So far, there are three major approaches to study the geometric ergodicty of SDEs:

1. **Hypocoercivity**. The convergence rate of the dynamics is reflected on the spectral gap of the generator. Readers may refer to Villani's book [4] or the review paper [5]. We also mention Lu's recent work [6], which compares the explicit convergence rates of the overdamped and underdamped Langevin dynamics.

2. **Harris ergodic theorem**. The minorization condition and the Lyapunov function imply geometric ergodicity. Except for the original work [3], Mattingly also considers the interacting particle system with singular potentials [7–9]. The Harris ergodic theorem can also be used to study the geometric ergodicity of the stochastic gradient Langevin dynamics (SGLD) [10].

3. **Reflection coupling**. A novel and purely probabilistic approach to study the geometric ergodicity. By designing a reflection coupling scheme between two duplicates, one is able to derive the exponential convergence of the distribution laws. The reflection coupling is firstly applied by Eberle [11] to study the overdamped Langevin dynamics. Eberle then applied the reflection coupling to study the interacting particle system [12], the underdamped Langevin dynamics [13], the Hamiltonian Monte Carlo [14, 15], the Andersen dynamics [16] and the mean-field dynamics [17,18]. Compared to the Harris ergodic theorem, the reflection coupling is able to obtain the explicit convergence rate independent on the number of particles, which makes it possible to study the mean-field limits. We also mention that the reflection coupling

can be used to study the ergodicity of the random batch method (RBM) [19] proposed by Jin.

In this section we will state and prove the Harris ergodic theorem, showing that the minorization condition and the Lyapunov function produce the geometric ergodicity. It is woth pointing that the minorization condition is usually easy to verify for diffusion processes, thus for a specific SDE, the only need is to find an appropriate Lyapunov function which satisfies the Lyapunov condition. In the subsequent sections, we shall see that the Harris ergdoic theorem can be applied in both the overdamped and underdamped Langevin dynamics, and can deal with singular potential functions.

1.2 Notations

Let $\{X_n\}_{n\geqslant 0}$ denote the underlying Markov chain on the state space $\mathcal{X}$, and $P(x,\cdot)$ is the transition probability. Let $\mathcal{P}(\mathcal{X})$ denote the set of all probability distributions on $\mathcal{X}$, then for each $x\in\mathcal{X}$, $P(x,\cdot)\in\mathcal{P}(\mathcal{X})$. Let the greek letters μ,ν denote the probability distributions on $\mathcal{X}$, and π denote the invariant distribution specially. When $\mathcal{X}$ is a continuous state space, for example, $\mathbb{R}^d$, the test function space is $C_b(\mathcal{X})\subset L^\infty(\mathcal{X})$, consisting of all continuous and bounded functions in $\mathcal{X}$. Since $P(x,\cdot)$ is the transition probablity in one step, define $P^k(x,\cdot)$ to be the transition probability in k steps. That is, when the initial value $X_0=x$, one has $X_k\sim P^k(x,\cdot)$. Formally, $P^0(x,\cdot)=\delta(x)$ is the Dirac distribution centered at x.

For $\mu\in\mathcal{P}(\mathcal{X})$ and $f\in L^\infty(\mathcal{X})$, define

$$\mu(f) := \int_{\mathcal{X}} f(x)\mu(\mathrm{d}x). \tag{1.1}$$

The transition probability $P(x,\cdot)$ induces a Markov operator $P: L^\infty(\mathcal{X})\to L^\infty(\mathcal{X})$:

$$(Pf)(x) := \int_{\mathcal{X}} f(y)P(x,\mathrm{d}y), \qquad \forall f\in L^\infty(\mathcal{X}). \tag{1.2}$$

If $f\geqslant 0$, there is also $Pf\geqslant 0$. According to the notation of (1.1), we can also write $P(x,f):=(Pf)(x)$. Note that $P(x,\cdot)$ also induces a Markov operator $P:\mathcal{P}(\mathcal{X})\to\mathcal{P}(\mathcal{X})$:

$$(\mu P)(A) := \int_{\mathcal{X}} P(x,A)\mu(\mathrm{d}x), \qquad \forall \mu\in\mathcal{P}(\mathcal{X}). \tag{1.3}$$

It is straightword to verify the formal associative law of multiplication

$$(\mu P)(f) = \mu(Pf) = \int_{\mathcal{X}\times\mathcal{X}} f(y)P(x,y)\mu(\mathrm{d}x). \tag{1.4}$$

The Markov operator corresponding to the transition probability $P^k(x,\cdot)$ is P^k.

1.3 Minorization condition and direct coupling

We first consider a simplifed version of the minorization condition and show that how the direct coupling approach can be used to derive the geometric ergodicity. The content here is based on [20]. Let $\{X_n\}_{n\geqslant 0}$ be the Markov chain on the state space $\mathcal{X}$ and $P(x,\cdot)$ be the transition probability.

Assumption 1.1 (minorization) *A Markov chain on $\mathcal{X}$ with transition probability $P(x,\cdot)$ is said to satisfy the minorization condition if there exists a $\rho \in \mathcal{P}(\mathcal{X})$ and $\varepsilon > 0$ such that*

$$\inf_{x\in\mathcal{X}} P(x,A) \geqslant \varepsilon\rho(A) \tag{1.5}$$

for all measurable sets $A \subset \mathcal{X}$.

The minorization condition (1.5) is *global* in $\mathcal{X}$, because the infimum is taken over $x \in \mathcal{X}$. The minorization condition (1.5) can be interpreted as, the transition probabilities have an *overlap* with positive measure. The state space $\mathcal{X}$ can either be discrete or continuous.

Example Let $\mathcal{X} = [0,2]$. Define the transition probability by

$$P(x,A) = \mathcal{N}(x,1;A) + r(x)\delta_x(A), \tag{1.6}$$

where $\mathcal{N}(x,1;A) = \mathbb{P}(z \in A)$ with $z \sim \mathcal{N}(x,1)$, and $r(x) = 1 - \mathcal{N}(x,1;\mathcal{X})$. Then $P(x,\cdot)$ satisfies the global minorization condition (1.5) by choosing ε sufficiently small and $\rho(A) = \delta_x(A)$.

The global minorization condition (1.5) allows us to directly couple two Markov chains $\{X_n\}_{n\geqslant 0}$, $\{Y_n\}_{n\geqslant 0}$ to prove the geometric ergodicity.

Theorem 1.1 (ergodicity) *If the transition probability $P(x,\cdot)$ on $\mathcal{X}$ satisfies Assumption* 1.1*, then for any $\mu,\nu \in \mathcal{P}(\mathcal{X})$, there is*

$$\|\mu P^n - \nu P^n\|_{\mathrm{TV}} \leqslant (1-\varepsilon)^n, \quad \forall n \in \mathbb{N}, \tag{1.7}$$

where $\|\cdot\|_{\mathrm{TV}}$ is the total variation of a signed measure.

Proof Define the modified transition probability

$$\tilde{P}(x,A) = \frac{1}{1-\varepsilon}(P(x,A) - \varepsilon\rho(A)) \tag{1.8}$$

for all measurable sets $A \subset \mathcal{X}$. The global minorization condition (1.5) ensures that $P(x,\cdot)$ is a nonnegative measure so that $\tilde{P}(x,\cdot)$ can be sampled. Define the direct coupling scheme for the Markov chains $\{X_n\}_{n\geqslant 0}, \{Y_n\}_{n\geqslant 0}$ as follows:

Algorithm 1: Direct coupling scheme

Input: $X_n = x$, $Y_n = y$, where the values $x, y \in \mathcal{X}$ are fixed.
Output: random variables X_{n+1}, Y_{n+1}.
if $x = y$ **then**
  Sample $X_{n+1} = Y_{n+1} \sim P(x,\cdot)$.
else
  Sample $\phi \in \{0,1\}$ with $\mathbb{P}(\phi = 0) = \varepsilon$.
  Sample $x' \sim \tilde{P}(x,\cdot)$, $y' \sim \tilde{P}(y,\cdot)$, $z \sim \rho(\cdot)$ independently.
  Define $X_{n+1} = \phi x' + (1-\phi)z$, $Y_{n+1} = \phi y' + (1-\phi)z$.
end

According to the direct coupling, it is easy to check the distribution law of X_{n+1} is exactly

$$(1-\varepsilon)\tilde{P}(x,A) + \varepsilon\rho(A) = P(x,A), \tag{1.9}$$

hence $\{X_n\}_{n\geqslant 0}$ is still a Markov chain with transition probability $P(x,\cdot)$. After each step, X_{n+1}, Y_{n+1} has probability ε to coincide, and then the two Markov chains become identical. Hence

$$\mathbb{P}(X_n \neq Y_n) \leqslant (1-\varepsilon)^n, \quad \forall n \in \mathbb{N}. \tag{1.10}$$

Suppose the initial values $X_0 \sim \mu$ and $Y_0 \sim \nu$, then the distribution laws of X_n and Y_n are μP^n and νP^n respectively. Hence (1.10) implies

$$\|\mu P^n - \nu P^n\|_{\mathrm{TV}} \leqslant (1-\varepsilon)^n, \tag{1.11}$$

which completes the proof. □

In the special case when ν is exactly the invariant distribution π, one obtains

Corollary 1.1 (ergodicity) *If the transition probability $P(x,\cdot)$ on $\mathcal{X}$ satisfies Assumption 1.1 and has an invariant distribution π, then for any $\mu \in \mathcal{P}(\mathcal{X})$, there is*

$$\|\mu P^n - \pi\|_{\mathrm{TV}} \leqslant (1-\varepsilon)^n, \quad \forall n \in \mathbb{N}. \tag{1.12}$$

The result also shows that the invariant distribution is unique. The global minorization condition (1.5), together with the existence of the invariant distribution, yields the geometric ergodicity.

1.4 Theorem of geometric ergodicity

The global minorization condition (1.5) is often too harsh for Markov chains, especially when the state space $\mathcal{X} = \mathbb{R}^d$. For example, consider the numerical integrator of the SDE

$$\mathrm{d}X_t = -X_t \mathrm{d}t + \sqrt{2}\mathrm{d}B_t \tag{1.13}$$

in a time step τ, the transition probability in $\mathbb{R}^d$ is given by

$$P(x,\cdot) = \mathcal{N}(e^{-\tau}x, 1-e^{-2\tau}). \tag{1.14}$$

For such a simple transition probability, it is trivial to check

$$\inf_{x\in\mathbb{R}^d} P(x,A) = 0 \tag{1.15}$$

for Borel measurable sets $A \in \mathcal{B}(\mathbb{R}^d)$. Hence it is impossible to build a global minorization condition.

In the following we always assume the state space is $\mathcal{X} = \mathbb{R}^d$. Instead of (1.1), consider

Assumption 1.2 (minorization) *A Markov chain on $\mathbb{R}^d$ with the transition probability $P(x,\cdot)$ is said to satisfy the minorization condition with respect to the compact set $C \in \mathcal{B}(\mathbb{R}^d)$ if there exists a $\rho \in \mathcal{P}(\mathbb{R}^d)$ and $\varepsilon > 0$ such that*

$$\inf_{x\in C} P(x,A) \geqslant \varepsilon\rho(A) \tag{1.16}$$

for all Borel measurable sets $A \in \mathcal{B}(\mathbb{R}^d)$.

The minorization condition (1.16) is called *local* because it only requires the infimum over the compact set C. For most diffusion processes including the overdamped and underdamped Langevin dynamics, the local minorization condition (1.16) is easy to satisfy. Roughly speaking, $P(x,\cdot)$ can be viewed as approximately a Gaussian distribution with parameters continuously depending on x, thus it is natural for the distribution family $\{P(x,\cdot)\}_{x\in C}$ to have a common lower bound, even when the potential function is singular.

Remark *In the Mattingly's original paper* [3], *the distribution ρ is required to satisfy $\rho(C) = 1$ and $\rho(C^c) = 0$. Since this condition is not necessary in the proof, we remove it here. Also in* [8,9], *the minorization condition does not require $\rho(C) = 1$.*

Other than the local minorization condition (1.16), we further require the Lyapunov condition to prove the ergodicity.

Assumption 1.3 (Lyapunov) *For the transition probability $P(x, \cdot)$, there is a function $W : \mathbb{R}^d \to [1, +\infty)$ with $\lim_{x\to\infty} W(x) = +\infty$, and real numbers $\alpha \in (0, 1), \beta \geqslant 0$ such that*

$$\mathbb{E}_{z\sim P(x,\cdot)}[W(z)] \leqslant \alpha W(x) + \beta, \qquad \forall x \in \mathbb{R}^d. \tag{1.17}$$

The Lyapunov condition (1.17) implies that $\mathbb{E}[W(X_{n+1})] \leqslant \alpha\mathbb{E}[W(X_n)] + \beta$. When the compact set C is sufficiently large, the Lyapunov condition ensures that C is approximately *recurrent.*

Now we state the Harris ergodic theorem, which exploits the local minorization condition (1.16) and the Lyapunov condition (1.17) to prove the geometric ergodicity.

Theorem 1.2 (ergodicity) *Let $\{X_n\}_{n\geqslant 0}$ be the Markov chain with transition probability $P(x, \cdot)$. Given $\gamma \in (\alpha^{\frac{1}{2}}, 1)$, suppose $P(x, \cdot)$ satisfies Assumption* 1.2 *and Assumption* 1.3 *with C given by*

$$C := \left\{ x \in \mathbb{R}^d : W(x) \leqslant \frac{2\beta}{\gamma - \alpha} \right\}, \tag{1.18}$$

then $P(x, \cdot)$ possesses a unique invariant distribution π. Furthermore, there is $r(\gamma) \in (0, 1)$ and $\kappa(\gamma) \in (0, +\infty)$ such that for all measurable functions $f : |f| \leqslant W$

$$|\mathbb{E}^x[f(X_n)] - \pi(f)| \leqslant \kappa r^n W(x), \qquad \forall x \in \mathbb{R}^d, \tag{1.19}$$

where $\mathbb{E}^x$ denotes the expectation under the condition $X_0 = x$.

Compared to Corollary 1.1, Theorem 1.2 measures the difference in distributions by the weak error, rather than the total distance.

Before going into the details, we first briefly describe the underlying idea of the proof. Similar to Theorem 1.1, we aim to define a coupling scheme between two duplicate Markov chains $\{X_n\}_{n\geqslant 0}$ and $\{Y_n\}_{n\geqslant 0}$ with transition probability $P(x, \cdot)$, then estimate the coinciding time. In general, there are two mechanisms promoting the Markov chains $\{X_n\}_{n\geqslant 0}$ and $\{Y_n\}_{n\geqslant 0}$ to coincide:

1. **Recurrency of the compact set C.** The Lyapunov condition (1.17) forces X_n, Y_n to have low probability far from the orgin. By choosing the compact C sufficiently large, we deduce that C is *recurrent* and X_n, Y_n have high probability stay inside C, creating conditions for the local minorization condition (1.16) to play its role.

2. **Direct coupling inside C.** Similar to the direct coupling in the proof of Theorem 1.1, we may define the direct coupling inside C. At least under the condition both $X_n, Y_n \in C$, the local minorization condition (1.16) allows X_n, Y_n to have positive probability to coincide in the next step. Since C is recurrent, we can always wait until X_n, Y_n enter C.

The two mechanisms together make it possible to estimate the coinciding time ζ of the coupled Markov chains $\{X_n\}_{n\geqslant 0}$ and $\{Y_n\}_{n\geqslant 0}$.

Remark *Mattingly in* [21] *presents an alternative proof of the Harris ergodic theorem, which is shorter than the proof in* [3]*. Still, we apply the original proof in* [3] *because the direct coupling is easier for me to understand. Also, it is possible to give explicit estimate of the convergence rate r once the minorization constant ε in* (1.16) *is known. Unfortunately, it is very difficult to determine the constant ε for SDEs, hence the Harris ergodic theorem can hardly produce quantitative estimate of the convergence rate for SDEs.*

1.5 Proof of the theorem

Now we prove the Harris ergodic theorem (Theorem 1.2). The proof in this note is mainly based on Mattingly [3], and more the techinal details can be found in Sections 12 & 15 in Meyn and Tweedie's book [2]. The proof below contains two key conclusions: existence of the invariant distribution and estimation of the coinciding time.

1.5.1 Preliminary results

First rewrite the Lyapunov condition (1.17) in terms of the compact set C.

Lemma 1.1 (Lyapunov) *Under Assumption* 1.3*, given constants $\gamma \in (\alpha, 1), s \in [1, +\infty)$, define*

$$c(s) := \frac{s\beta}{\gamma - \alpha}, \qquad C(s) := \{x \in \mathbb{R}^d : W(x) \leqslant c(s)\}, \tag{1.20}$$

then

$$\mathbb{E}_{z \sim P(x,\cdot)}[W(z)] \leqslant \gamma W(x) + s\beta 1_{C(s)}(x), \qquad \forall x \in \mathbb{R}^d. \tag{1.21}$$

In the following, under the Lyapunov condition 1.17, we shall choose the compact set C to be $C(2)$. Induction of the inequality (1.21) on n directly yields the following result.

Lemma 1.2 (Lyapunov) *Let N be any stopping time and fix an $n \geqslant 0$. Under Assumption* 1.3*, there exists a constant κ such that*

$$\mathbb{E}\big[W(X_n)1_{N \geqslant n}\big] \leqslant \kappa\gamma^n \left[W(x_0) + \mathbb{E}\left\{\sum_{j=1}^{N \wedge n} \gamma^{-j} 1_C(X_{j-1})\right\}\right] \leqslant \frac{\kappa[\gamma^n W(X_0) + 1]}{1-\gamma}. \tag{1.22}$$

Utilizing the Lyapunov function, Lemma 1.2 presents a useful tool to characterize the stopping times relating to the compact set C, because the exponential tails involve the indicative functions of C. In particular, the first visit time τ_C can be estimated as follows.

Corollary 1.2 (visit time) *Under Assumption* 1.3*, let $\tau_C := \inf\{n > 0 : x_n \in C\}$ be the next visit time. There exists a constant κ such that for any $n > 0$ one has*

$$\mathbb{P}(\tau_C > n) \leqslant \kappa\gamma^n[W(X_0) + 1] \tag{1.23}$$

and

$$\mathbb{E}\left(\frac{1}{\gamma}\right)^{\tau_C} \leqslant \kappa[W(X_0) + 1]. \tag{1.24}$$

Proof Since τ_C is the visit time, one has $X_j \notin C$ for $j = 1, \cdots, \tau_C - 1$. Hence the exponential tail in (1.22) becomes

$$\sum_{j=1}^{n\wedge\tau_C} \gamma^{n-j} 1_C(X_{j-1}) = \gamma^{n-1} 1_C(X_0) \leqslant \gamma^{n-1}. \tag{1.25}$$

Also

$$\mathbb{E}\big[W(X_n)1_{\tau_C>n}\big] \geqslant \frac{s\beta}{\gamma-\alpha}\mathbb{E}[1_{\tau_C>n}] = \frac{s\beta}{\gamma-\alpha}\mathbb{P}(\tau_C > n), \tag{1.26}$$

hence the first conclusion (1.23) holds. For the second conclusion (1.25), pick $s \in [1, 2)$ and define

$$\gamma' = \alpha + \frac{s}{2}(\gamma - \alpha) \in (\alpha, \gamma). \tag{1.27}$$

It is easy to check the compact set induced by s and γ' is exactly the original C since

$$\frac{s\beta}{\gamma'-\alpha} = \frac{2\beta}{\gamma-\alpha}. \tag{1.28}$$

Therefore, the first conclusion (1.23) holds for any $\gamma' \in (\alpha, \gamma)$. Hence

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{\gamma}\right)^{\tau_C} &= \sum_{n=1}^{\infty} \frac{1}{\gamma^n}\mathbb{P}(\tau_C = n) \\
&\leqslant \sum_{n=1}^{\infty} \frac{1}{\gamma^n}\mathbb{P}(\tau_C > n-1) \\
&\leqslant \kappa \sum_{n=1}^{\infty} \frac{1}{\gamma^n}(\gamma')^n[W(X_0)+1] \\
&\leqslant \kappa[W(X_0)+1].
\end{aligned}$$

□

The estimate of τ_C shows that the visit time τ_C approximately obeys the exponential distribution, implying that the compact set C is recurrent in the Markov chain $\{X_n\}_{n\geqslant 0}$.

1.5.2 Existence of the invariant distribution

The existence of the invariant distribution can be proved only using the Lyapunov condition. Fix the initial value $x_0 \in \mathbb{R}^d$ and define the distribution $\mu_n \in \mathcal{P}(\mathbb{R}^d)$ by

$$\mu_n = \frac{1}{n}\sum_{k=1}^{n} P^k(x_0, \cdot). \tag{1.29}$$

Recall that $P^k(x, \cdot)$ is the transition probability in k steps. Equivalently, μ_n can be interpreted as

$$\mu_n(A) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{P}(X_k \in A) \tag{1.30}$$

for measurable sets $A \in \mathcal{B}(\mathbb{R}^d)$. Roughly speaking, $\mu_n(A)$ characterizes the proportion of time that the Markov chain $\{X_n\}_{n\geqslant 0}$ stays inside A, hence we expect the limit of μ_n to be the invariant distribution w.r.t. the transition probability $P(x, \cdot)$.

We show that μ_n is a tight sequence of measures. The definition of the tightness for a family of sequences can be found at [22]. In fact, we shall show that for any $\varepsilon > 0$, there exists a compact set $C_\varepsilon \in \mathcal{B}(\mathbb{R}^d)$ such that

$$\mu_n(C_\varepsilon^c) \leqslant \varepsilon, \qquad \forall n \in \mathbb{N}. \tag{1.31}$$

To prove (1.31), choose C_ε to be the level set of the Lyapunov function $W(x)$:

$$C_\varepsilon := \big\{x \in \mathbb{R}^d : W(x) \leqslant L_\varepsilon\big\}, \tag{1.32}$$

where $L_\varepsilon > 0$ is a constant which will determined later. Then using Chebyshev's inequality,

$$\begin{aligned} \mu_n(C_\varepsilon^c) &= \frac{1}{n}\sum_{k=1}^n \mathbb{P}(X_k \notin C_\varepsilon) \\ &= \frac{1}{n}\sum_{k=1}^n \mathbb{P}(W(X_k) > L_\varepsilon) \\ &\leqslant \frac{1}{nL_\varepsilon}\sum_{k=1}^n \mathbb{E}[W(X_k)] \leqslant \frac{\kappa(W(x_0)+1)}{L_\varepsilon}. \end{aligned}$$

Therefore, we may just choose

$$L_\varepsilon = \frac{\kappa(W(x_0)+1)}{\varepsilon} \tag{1.33}$$

to obtain the tightness (1.31).

By Prokhorov's theorem [22], there exists a subsequene of μ_n which converges weakly to $\pi \in \mathcal{P}(\mathbb{R}^d)$. Let us assume

$$\mu_{n_k} \xrightarrow{w} \pi, \qquad \text{as } k \to \infty. \tag{1.34}$$

We show that π is the invariant distribution. For any test function $f \in C_b(\mathbb{R}^d)$, notice that

$$\mu_n(Pf) = (\mu_n P)(f) = \frac{1}{n}\sum_{k=1}^n P^{k+1}(x_0, f) = \frac{1}{n}\sum_{k=2}^{n+1} P^k(x_0, f),$$

hence we have the identity

$$\mu_n(Pf) - \mu_n(f) = \frac{1}{n}\Big(P^{n+1}(x_0, f) - P(x_0, f)\Big) \tag{1.35}$$

and thus

$$|\mu_n(Pf) - \mu_n(f)| \leqslant \frac{2}{n}\|f\|_{L^\infty(\mathbb{R}^d)}. \tag{1.36}$$

Let $n = n_k$ and $k \to \infty$, we obtain

$$\lim_{k\to\infty}\Big(\mu_{n_k}(Pf) - \mu_{n_k}(f)\Big) = 0. \tag{1.37}$$

Since the weak limit of μ_{n_k} is π, we obtain for any $f \in C_b(\mathbb{R}^d)$,

$$\pi(Pf) = \pi(f), \tag{1.38}$$

which implies $\pi P = \pi$, and π is the invariant distribution.

In this part we only proved the existence of the invariant distribution. The Lyapunov condition is not sufficient to guarantee the uniqueness.

1.5.3 Direct coupling scheme inside C

We aim to define the coupling scheme between the Markov chains $\{X_n\}_{n\geqslant 0}, \{Y_n\}_{n\geqslant 0}$, and estimate the their coinciding time ζ. Similar to the proof of Theorem 1.1, define the modified transition probability $\tilde{P}(x,\cdot)$ by

$$\tilde{P}(x,A) = \begin{cases} P(x,A), & \text{if } x \in C^c \\ \dfrac{1}{1-\varepsilon}(P(x,A) - \varepsilon\rho(A)), & \text{if } x \in C \end{cases} \tag{1.39}$$

for all measurable sets $A \in \mathcal{B}(\mathbb{R}^d)$. The local minorization condition (1.16) ensures that $\tilde{P}(x,\cdot)$ is a nonnegative measure for any $x \in \mathbb{R}^d$.

Define the direct coupling scheme for the Markov chains $\{X_n\}_{n\geqslant 0}, \{Y_n\}_{n\geqslant 0}$ as follows:

Algorithm 2: Direct coupling scheme

Input: $X_n = x$, $Y_n = y$, where the values $x, y \in \mathcal{X}$ are fixed.
Output: random variables X_{n+1}, Y_{n+1}.

if $x = y$ **then**
  Sample $X_{n+1} = Y_{n+1} \sim P(x,\cdot)$.
else
  Sample $\phi \in \{0,1\}$ with $\mathbb{P}(\phi = 0) = \varepsilon$.
  Sample $x' \sim \tilde{P}(x,\cdot)$, $y' \sim \tilde{P}(y,\cdot)$, $z \sim \rho(\cdot)$ independently.
  Define $\begin{cases} X_{n+1} = 1_C(x)\big(\phi x' + (1-\phi)z\big) + (1 - 1_C(x))x' \\ Y_{n+1} = 1_C(y)\big(\phi y' + (1-\phi)z\big) + (1 - 1_C(y))y' \end{cases}$
end

It is clear that the transition probability of $\{X_n\}_{n\geqslant 0}$ is exactly

$$\begin{aligned} & 1_C(x)((1-\varepsilon)\tilde{P}(x,A) + \varepsilon\rho(A)) + (1 - 1_C(x))\tilde{P}(x,A) \\ = {} & 1_C(x)P(x,A) + (1 - 1_C(x))P(x,A) = P(x,A). \end{aligned}$$

We present some observations of the direct coupling scheme above:

1. If $(x,y) \notin C \times C$, then $X_{n+1}|X_n = x$ and $Y_{n+1}|Y_n = y$ are mutually independent.

2. If $(x,y) \in C \times C$, then $X_{n+1}|X_n = x$ and $Y_{n+1}|Y_n = y$ has ε probability to coincide. Once $X_{n+1} = Y_{n+1}$ happens, $\{X_n\}_{n\geqslant 0}$ and $\{Y_n\}_{n\geqslant 0}$ will be identical in the remaining time.

In this way, the direct coupling scheme manages to bring X_n and Y_n together once they both enter the compact set C. Recalling the Lyapunov condition (1.17), there are two different mechanisms helping the coupled Markov chains $\{X_n\}_{n\geqslant 0}$ and $\{Y_n\}_{n\geqslant 0}$ coincide:

1. **Recurrency of the compact set C.** The first visit time τ_C obeys the exponential distribution approximately, hence $\{X_n\}_{n\geqslant 0}, \{Y_n\}_{n\geqslant 0}$ shall enter C frequently.

2. **Direct coupling inside C.** Once both X_n, Y_n stay inside C, there is ε probability for them to coincide. If they do not coincide, wait for the next time they stay inside C.

Now we are ready to prove the theorem. Suppose $\{X_n\}_{n\geqslant 0}$ and $\{Y_n\}_{n\geqslant 0}$ are two Markov chains evolved by the direct coupling scheme, and the initial values $X_0 = x_0$ and $Y_0 = y_0$ in $\mathbb{R}^d$ are given. Define the coinciding time of the Markov chains by

$$\zeta = \inf_{n\geqslant 0}\{(X_n, Y_n) \in C \times C, \phi_n = 0\}. \tag{1.40}$$

Here, the subscript n in the random variable ϕ represents the direct coupling scheme at the n-th step. For any test function $f \in C_b(\mathbb{R}^d)$ with $|f| \leqslant W$, write $f = f^+ - f^-$, where $f^+, f^- \geqslant 0$. Then

$$\mathbb{E}f^+(X_n) = \mathbb{E}f^+(X_n)1_{n\geqslant\zeta} + \mathbb{E}f^+(X_n)1_{n<\zeta}. \tag{1.41}$$

After the coinciding time ζ, the Markov chains X_n, Y_n become idential, hence

$$\mathbb{E}f^+(X_n)1_{n\geqslant\zeta} = \mathbb{E}f^+(Y_n)1_{n\geqslant\zeta} \leqslant \mathbb{E}f^+(Y_n). \tag{1.42}$$

Combining (1.41)(1.42) gives

$$\mathbb{E}f^+(X_n) - \mathbb{E}f^+(Y_n) \leqslant \mathbb{E}f^+(X_n)1_{n<\zeta} \leqslant \mathbb{E}W(X_n)1_{n<\zeta}, \tag{1.43}$$

thus

$$|\mathbb{E}f^+(X_n) - \mathbb{E}f^+(Y_n)| \leqslant \max\{\mathbb{E}W(X_n)1_{n<\zeta}, \mathbb{E}W(Y_n)1_{n<\zeta}\}. \tag{1.44}$$

Similarly

$$|\mathbb{E}f^-(X_n) - \mathbb{E}f^-(Y_n)| \leqslant \max\{\mathbb{E}W(X_n)1_{n<\zeta}, \mathbb{E}W(Y_n)1_{n<\zeta}\}. \tag{1.45}$$

Hence

$$|\mathbb{E}f(X_n) - \mathbb{E}f(Y_n)| \leqslant 2\max\{\mathbb{E}W(X_n)1_{n<\zeta}, \mathbb{E}W(Y_n)1_{n<\zeta}\}. \tag{1.46}$$

Our final task is to estimate the coinciding time ζ in the RHS of (1.46).

1.5.4 Estimation of the coinciding time ζ

Since $W(x) \geqslant 1$, it is necessary to prove

$$\mathbb{E}1_{n<\zeta} = \mathbb{P}(\zeta > n) \tag{1.47}$$

has exponential decay as n increases. Roughly speaking, we need to prove ζ obeys the exponential distribution approximately. However, unlike the visit time τ_C, there are two different mechanisms impelling X_n and Y_n to coincide, hence the estimation of ξ will be more technical.

To complete the proof of the theorem, we need to prove

Lemma 1.3 (coinciding) *Given $\gamma \in (\alpha^{\frac{1}{2}}, 1)$, suppose $\{X_n\}_{n\geqslant 0}, \{Y_n\}_{n\geqslant 0}$ are evolved by the direct coupling scheme with initial values x_0, y_0. There exists $r \in (0, 1)$ and a constant $\kappa > 0$ such that*

$$\max\{\mathbb{E}W(X_n)1_{n<\zeta}, \mathbb{E}W(Y_n)1_{n<\zeta}\} \leqslant \kappa[W(x_0) + W(y_0) + 1]r^n. \tag{1.48}$$

Instead of considering $\{X_n\}_{n\geqslant 0}$ and $\{Y_n\}_{n\geqslant 0}$ separately, view $\{(X_n, Y_n)\}_{n\geqslant 0}$ as a single Markov chain in $\mathbb{R}^d \times \mathbb{R}^d$. For convenience, let $\mathcal{F}_n$ be the filtration at the n-th step. Define the Lyapunov function in $\mathbb{R}^d \times \mathbb{R}^d$ by

$$W'(x, y) = W(x) + W(y), \quad \text{for } x, y \in \mathbb{R}^d, \tag{1.49}$$

then the Lyapunov condition holds

$$\mathbb{E}[W'(X_{n+1}, Y_{n+1})|\mathcal{F}_n] \leqslant \alpha \mathbb{E}[W'(X_n, Y_n)] + 2\beta. \tag{1.50}$$

Choosing $s = 1$ in Lemma 1.1, for any $\gamma \in (\alpha, 1)$ one has

$$\mathbb{E}[W'(X_{n+1}, Y_{n+1})|\mathcal{F}_n] \leqslant \gamma \mathbb{E}[W'(X_n, Y_n)] + 2\beta 1_{C'}(X_n, Y_n), \tag{1.51}$$

where the compact set C' in $\mathbb{R}^d \times \mathbb{R}^d$ is given by

$$C' = \left\{(x, y) : W'(x, y) \leqslant \frac{2\beta}{\gamma - \alpha}\right\}. \tag{1.52}$$

Clearly, $C' \subset C \times C$. Now define the coinciding time in C' by

$$\zeta' = \inf_{n \geqslant 0} \{(X_n, Y_n) \in C', \phi_n = 0\}, \tag{1.53}$$

then $\zeta \leqslant \zeta'$. Hence

$$\max\{\mathbb{E}W(X_n)1_{n<\zeta}, \mathbb{E}W(Y_n)1_{n<\zeta}\} \leqslant \mathbb{E}W'(X_n, Y_n)1_{n<\zeta'}. \tag{1.54}$$

Now we only need to show that

$$\mathbb{E}W'(X_n, Y_n)1_{n<\zeta'} \leqslant \kappa[W(x_0) + W(y_0) + 1]r^n. \tag{1.55}$$

To estimate ζ' in the LHS of (1.55), consider the trajectory of $\{(X_n, Y_n)\}_{n\geqslant 0}$ in the first n steps. Since the compact set C' is recurrent, $\{(X_n, Y_n)\}_{n\geqslant 0}$ shall enter C' frequently. To quantify the frequency staying inside C', let τ_k be the time of the k-th visit to C'. In other words, τ_k is the k-th smallest element of the index set

$$I_{C'} = \{k \geqslant 1 : (X_k, Y_k) \in C'\}. \tag{1.56}$$

For any real s, define $\tau_s = \tau_{\lceil s \rceil}$. Fixing $a \in (0, 1)$, the behavior of the trajectory of $\{(X_n, Y_n)\}_{n\geqslant 0}$ in the first n steps can be classified into two categories:

- **Typical behavior:** $\tau_{an} < n$. Equivalently,

  $$\#\{1 \leqslant k \leqslant n-1 : (X_k, Y_k) \in C'\} \geqslant \lceil an \rceil. \tag{1.57}$$

  The trajectory has at least frequency a to stay inside C', and thus belongs to the typical case.

- **Unusual behavior:** $\tau_{an} \geqslant n$. Equivalently,

  $$\#\{1 \leqslant k \leqslant n-1 : (X_k, Y_k) \in C'\} < \lceil an \rceil. \tag{1.58}$$

  The trajectory has low frequency to stay inside C'. Due to the Lyapunov condition (1.51), this is the unusual case.

Write $\mathbb{E}W'(X_n, Y_n)1_{n<\zeta'} = I_1 + I_2$, corresponding to the typical and the unusual behavior,

$$I_1 = \mathbb{E}W'(X_n, Y_n)1_{n<\zeta'}1_{\tau_{an}<n}, \quad I_2 = \mathbb{E}W'(X_n, Y_n)1_{n<\zeta'}1_{\tau_{an}\geqslant n}. \tag{1.59}$$

- Estimate I_1: Let $\bar{W} = \sup_{(x,y)\in C'} W'(x,y)$, then

$$
\begin{aligned}
I_1 &\leqslant \mathbb{E} W'(X_n,Y_n) 1_{\tau_{an}<n} 1_{\tau_{an}<\zeta'} \\
&\leqslant \mathbb{E}\Big(1_{\tau_{an}<\zeta'} \mathbb{E}\big(W'(X_n,Y_n)\big|\tau_{an}<n, \mathcal{F}_{\tau_{an}}\big)\Big)\mathbb{P}(\tau_{an}<n).
\end{aligned}
$$

View τ_{an} as the starting time and define $\tilde{X}_k = X_{k-\tau_{an}}$, $\tilde{Y}_k = Y_{k-\tau_{an}}$, then by Lemma 1.2,

$$
\begin{aligned}
\mathbb{E}\big(W'(X_n,Y_n)\big|\tau_{an}<n,\mathcal{F}_{\tau_{an}}\big) &\leqslant \sup_{k>\tau_{an}} \mathbb{E}\big(W'(X_k,Y_k)\big|\mathcal{F}_{\tau_{an}}\big) \\
&\leqslant \sup_{(x,y)\in C'} \sup_{k>0} \mathbb{E}\big(W'(\tilde{X}_k,\tilde{Y}_k)\big|(\tilde{X}_0,\tilde{Y}_0)=(x,y)\big) \\
&\leqslant \sup_{(x,y)\in C'} \kappa[W'(x,y)+1] \\
&\leqslant \kappa(\bar{W}+1).
\end{aligned}
$$

Hence

$$
\begin{aligned}
I_1 &\leqslant \mathbb{E}\Big(1_{\tau_{an}<\zeta'}\kappa(\bar{W}+1)\Big)\mathbb{P}(\tau_{an}<n) \\
&= \kappa(\bar{W}+1)\mathbb{E}1_{\tau_{an}<\zeta'}1_{\tau_{an}<n} \\
&\leqslant \kappa(\bar{W}+1)\mathbb{P}(\tau_{an}<\zeta') \\
&\leqslant \kappa(\bar{W}+1)(1-\eta)^{an}.
\end{aligned}
$$

The last inequality $\mathbb{P}(\tau_{an}<\zeta')$ holds because for each time $(X_n,Y_n)\in C'$, there is ε probability for them to coincide. Finally we obtain $I_1 \leqslant \kappa(\bar{W}+1)(1-\eta)^{an}$.

- Estimate I_2: Let $W_0 = W'(x_0,y_0)$. Write I_2 as

$$
I_2 = \sum_{k=0}^{\lceil an\rceil -1} \mathbb{E}W'(X_n,Y_n)1_{n<\zeta'}1_{\tau_k<n\leqslant\tau_{k+1}}. \tag{1.60}
$$

Note that $\tau_k < n \leqslant \tau_{k+1}$ means there are exactly k visits to C' in the first n steps.

  - For $k=0$, by Lemma 1.2

$$
\mathbb{E}W'(X_n,Y_n)1_{n<\zeta'}1_{\tau_0<n\leqslant\tau_1} \leqslant \mathbb{E}W'(X_n,Y_n)1_{\tau_1\geqslant n} \leqslant \kappa\gamma^n W_0. \tag{1.61}
$$

  - For $k\geqslant 1$, by Lemma 1.2

$$
\begin{aligned}
&\mathbb{E}W'(X_n,Y_n)1_{n<\zeta'}1_{\tau_k<n\leqslant\tau_{k+1}} \\
\leqslant\ &\mathbb{E}W'(X_n,Y_n)1_{\tau_k<\zeta'}1_{\tau_k<n\leqslant\tau_{k+1}} \\
=\ &\mathbb{E}\Big(1_{\tau_k<\zeta'}\mathbb{E}\big(W'(X_n,Y_n)1_{\tau_{k+1}\geqslant n}|\mathcal{F}_{\tau_k},\tau_k<n\big)|\tau_k<n\Big)\mathbb{P}(\tau_k<n) \\
\leqslant\ &\mathbb{E}\Big(1_{\tau_k<n}1_{\tau_k<\zeta'}\kappa\gamma^{n-\tau_k}(\bar{W}+1)\Big) \\
\leqslant\ &\kappa(\bar{W}+1)\Big(\mathbb{E}1_{\tau_k<\zeta'}\mathbb{E}\gamma^{-2\tau_k}\Big)^{\frac{1}{2}} \\
\leqslant\ &\kappa(\bar{W}+1)(1-\eta)^{\frac{k}{2}}\left(\mathbb{E}\gamma^{-2\tau_k}\right)^{\frac{1}{2}}
\end{aligned}
$$

Induction on k gives

$$\mathbb{E}\gamma^{-2\tau_k} \leqslant \kappa^k(\bar{W}+1)^{k-1}(W_0+1). \tag{1.62}$$

For $R = \max\{1, \kappa(\bar{W}+1)\}$, one has

$$\mathbb{E}W'(X_n, Y_n)1_{n<\zeta'}1_{\tau_k<n\leqslant\tau_{k+1}} \leqslant (1-\eta)^{\frac{k}{2}}R^k\sqrt{2}W_0\gamma^n. \tag{1.63}$$

Summation over k gives

$$I_2 \leqslant \kappa\sqrt{2}W_0\gamma^n R^{an}. \tag{1.64}$$

By choosing a sufficiently small so that $\gamma R^a < 1$, we obtain the desired result.

1.6 Continuous-time dynamics

For continuous-time Markov process $\{x(t)\}_{t\geqslant 0}$, the geometric ergodicity is usually studied via the embedded Markov chain $X_n := x(nT)$, where $T > 0$ is a fixed constant. For the overdamped and the underdamped Langevin dynamics, a sufficient condition [7,8] for the minorization condition is:

$$\text{For any } x, y \in \mathbb{R}^d \text{ and } t, \delta > 0,\ P_t(x, B_\delta(y)) > 0.$$

In other words, Any Region is reachable at Any Time (ARAT). The ARAT condition in fact implies that the local minorization condition (1.16) holds w.r.t. compact sets like

$$C_R := \{x \in \mathbb{R}^d : W(x) \leqslant R\} \tag{1.65}$$

for sufficiently large R. The ARAT condition can be proved via the support theorems [23, 24]. As demonstrated in [7, 8], the verification of the minorization condition for the overdamped and underdamped Langevin dynamics is routine, hence will not be a big issue in the application of the Harris ergodic theorem.

The real difficulty in applicating the Harris theorem for SDEs is the Lyapunov condition. If the SDE possesses the generator $\mathcal{L}$, then the continuous-time Lyapunov condition should be

$$\mathcal{L}W(x) \leqslant -cW(x) + M, \qquad \forall x \in \mathbb{R}^d, \tag{1.66}$$

where $c, M > 0$ are constants. This enables us to deduce

$$\mathbb{E}[W(x(t))] \leqslant e^{-ct}W(x(0)) + \frac{1-e^{-ct}}{c}M, \tag{1.67}$$

that is, the discrete Lyapunov condition (1.17) with $\alpha = e^{-ct} < 1$ and $\beta = (1-e^{-ct})M/c$. For the overdamped and the underdamped Langevin dynamics, finding a Lyapunov function satisfying (1.66) is not a piece of cake. We shall see in next sections various options of the Lyapunov function.

Although for the SDEs verifying the minorization condition is much more easier than the Lyapunov condition, we emphasize that this is not always the case. For some kinetic equations [25], verifying the minorization condition could be a tough thing.

For the continuous-time dynamics, the Harris ergodic theorem shows the geometric ergodicity for the embedded Markov chain $X_n := x(nT)$. With a simple argument, one may also obtain the geometric ergodicity for all $t \geqslant 0$. Suppose $\{x(t)\}_{t\geqslant 0}$ is a Markov process with initial value $x(0) = x$. The local minorization condition (1.5) holds for the transition probability $P_T(x, \cdot)$, where $T > 0$ is a fixed time duration. Also, the continuous Lyapunov condition (1.66) is satisfied. For any $t \geqslant 0$, write $t = nT + \delta$ with $\delta \in [0, T)$. Applying the Harris ergodic theorem with initial value $x(\delta)$, one deduces that $P_T(x, \cdot)$ has a unique invariant distribution $\pi \in \mathcal{P}(\mathbb{R}^d)$ and

$$|\mathbb{E}^x[f(x(t))] - \pi(f)| \leqslant \kappa r^n \mathbb{E}^x[W(x(\delta))], \tag{1.68}$$

where the expectation $\mathbb{E}^x$ denotes the condition $x(0) = x$. Using the Lyapunov condition (1.66),

$$\mathbb{E}^x[W(x(\delta))] \leqslant e^{-c\delta} W(x) + \frac{1 - e^{-c\delta}}{c} M, \tag{1.69}$$

hence

$$|\mathbb{E}^x[f(x(t))] - \pi(f)| \leqslant \kappa r^n W(x) \leqslant \kappa e^{-\lambda t} W(x), \tag{1.70}$$

where $\lambda > 0$ satisfies $e^{-\lambda T} = r < 1$. We obtain the geometric ergodicity for $\{x(t)\}_{t \geqslant 0}$.

1.7 Summary

The Harris ergodic theorem (Theorem 1.2) provides a powerful tool to study the geometric ergodicity of Markov chains and SDEs. The theorem only requires the local minorization condition (1.16) and the Lyapunov condition (1.17) to prove the existence of the invariant distribution and the geometric ergodicity. For diffusion processes including the overdamped and underdamed Langevin dynamics, the verification of the minorization is standard, hence the main difficulty is to find an appropriate Lyapunov function satisfying (1.17). Suppose $\mathcal{L}$ is the generator of the SDE, then our goal is to find a Lyapunov function $W(x)$ satisfying

$$\mathcal{L}W(x) \leqslant -cW(x) + M, \qquad \forall x \in \mathbb{R}^d.$$

Compared to the hypocoercivity and the reflection couplnig, the Harris ergodic theorem can be applied in a wide class of stochastic models, but lacks the ability to quantify the convergence rate. In fact, it is even difficult to determine the explicit bounds for the minorization constant [8]. Nevertheless, there have been some pioneering attempts to combine the Lyapunov condition and the reflection coupling to find the explicit convergence rate [13, 17].

2 Ergodicity of SDEs

We study the geometric ergodicity of various SDEs, using the Harris ergodic theorem. Since the minorization condition is easy to check for most diffusion processes, hence this section mainly involves the construction of Lyapunov functions.

2.1 Overdamped Langevin dynamics

The overdamped Langevin dynamics is the simplest kind of SDEs. It is of the first order and the diffusion is nondegenerate. Consider $\{x_t\}_{t \geqslant 0}$ in $\mathbb{R}^d$ evolved by the SDE

$$\mathrm{d}x_t = b(x_t)\mathrm{d}t + \sqrt{\frac{2}{\beta}}\mathrm{d}B_t, \qquad t \geqslant 0, \tag{2.1}$$

where $b(\cdot):\mathbb{R}^d\to\mathbb{R}^d$ is the drift force, $\beta>0$ is the inverse temperature, and B_t is the standard Brwonian motion in $\mathbb{R}^d$. It is clear that the generator of (2.1) is

$$\mathcal{L}u(x)=b(x)\cdot\nabla u(x)+\frac{1}{\beta}\Delta u(x),\qquad u\in C^2(\mathbb{R}^d). \tag{2.2}$$

Due to the simple structure of the SDE, there are various choices of the Lyapunov function $W(x)$.

1. If the drift force $b(x)$ satisfies the dissipation condition

   $$-b(x)\cdot x\geqslant c|x|^2-M,\qquad \forall x\in\mathbb{R}^d, \tag{2.3}$$

   we can simply choose $W(x)=\frac{1}{2}|x|^2$. This is because

   $$\begin{aligned}\mathcal{L}W(x)&=b(x)\cdot\nabla W(x)+\frac{1}{\beta}\Delta W(x)\\&=b(x)\cdot x+\frac{1}{\beta}\\&\leqslant -c|x|^2+M+\frac{1}{\beta}\\&=-2c\cdot W(x)+M+\frac{1}{\beta},\end{aligned}$$

   hence the Lyapunov condition holds true. In this case, the drift force $b(x)$ need not be the gradient. However, $b(x)$ should be regular force otherwise (2.3) fails.

2. If $b(x)=-\nabla U(x)$ for some potential function $U(x)$, and $U(x)$ satisfies

   $$\lim_{x:U(x)\to+\infty}|\nabla U(q)|=+\infty,\qquad \lim_{x:U(x)\to+\infty}\frac{\nabla^2U(x)}{|\nabla U(x)|^2}=0, \tag{2.4}$$

   we can choose $W(x)=e^{bU(x)}$ for some $b\in(0,\beta)$.

   An equivalent condition for (2.4) is, if a sequence $\{x_k\}_{k\geqslant1}$ satisfies $U(x_k)\to+\infty$, then

   $$|\nabla U(x_k)|\to+\infty,\qquad \frac{|\nabla^2U(x_k)|}{|\nabla U(x_k)|^2}=0. \tag{2.5}$$

   To verify this condition, we only need to show $\mathcal{L}W(x)\leqslant -cW(x)$ for sufficiently large $U(x)$. For $W(x)=e^{bU(x)}$, it is easy to see

   $$\nabla W(x)=W(x)\big(b\nabla U(x)\big),\qquad \Delta W(x)=W(x)\big(b^2|\nabla U(x)|^2+b\Delta U(x)\big). \tag{2.6}$$

   hence

   $$\begin{aligned}\mathcal{L}W(x)&=-\nabla U(x)\cdot\nabla W(x)+\frac{1}{\beta}\Delta W(x)\\&=\frac{b}{\beta}W(x)\Big(-(\beta-b)|\nabla U(x)|^2+\Delta U(x)\Big).\end{aligned}$$

   A sufficient condition for $W(x)\leqslant -cW(x)$ is

   $$-(\beta-b)|\nabla U(x)|^2+\Delta U(x)\leqslant -c,\qquad \text{for sufficiently large } U(x). \tag{2.7}$$

   This is true under the condition (2.4). Note that the choice $W(x)=e^{bU(x)}$ is valid even for singular potentials. In particular, (2.4) holds for the Coulomb potential when $d\geqslant3$.

2.2 Stochastic gradient Langevin dynamics

The Harris ergodic theorem is naturally suitable for the stochastic gradient Langevin dynamics (SGLD). Let $\tau > 0$ be the time step and $t_n := n\tau$. Let $\{b_i(x)\}_{i=1}^N$ be a family of drift forces and the SGLD in the time interval $[t_n, t_{n+1})$ is given by

$$\mathrm{d}x_t = b_i(x_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t, \qquad t \in [t_n, t_{n+1}), \tag{2.8}$$

where $i \in \{1, \cdots, N\}$ is a batch index randomly generated at time t_n. Choose the Lyapunov function $W(x) = \frac{1}{2}|x|^2$, then we can verify the embedded Markov chain $X_n := x_{n\tau}$ satisfies

$$\mathbb{E}^i[W(X_{n+1})|X_n] \leqslant \alpha \cdot W(X_n) + \beta, \tag{2.9}$$

where i denotes the index i is given. Taking the expectation over i, we obtain

$$\mathbb{E}[W(X_{n+1})|X_n] \leqslant \alpha \cdot W(X_n) + \beta, \tag{2.10}$$

establishing the Lyapunov condition. Similar discussion can also be found in [10].

2.3 Second-order Langevin dynamics

Consider the second-order Langevin dynamics in $\mathbb{R}^d \times \mathbb{R}^d$ given by

$$\begin{cases} \mathrm{d}q_t = p_t \mathrm{d}t \\ \mathrm{d}p_t = -\gamma p_t \mathrm{d}t - \nabla U(q_t)\mathrm{d}t + \sqrt{\dfrac{2\gamma}{\beta}}\mathrm{d}B_t \end{cases} \tag{2.11}$$

where $U(\cdot) : \mathbb{R}^d \to \mathbb{R}$ is the potential function, and the corresponding Hamiltonian is

$$H(q, p) = \frac{1}{2}|p|^2 + U(q). \tag{2.12}$$

Assume $U(q)$ satisfies the dissipation condition: there exists $c_0, M > 0$ such that

$$\nabla U(q) \cdot q \geqslant c_0 U(q) + c_0|q|^2 - M. \tag{2.13}$$

The invariant distribution of (2.11) is $\pi(q, p) \propto \exp(-\beta H(q, p))$. The generator $\mathcal{L}$ is given by

$$\mathcal{L} = p \cdot \nabla_q - \nabla_q U(q) \cdot \nabla_p - \gamma p \cdot \nabla_p + \gamma\beta^{-1}\Delta_p. \tag{2.14}$$

Now we aim to construct the Lyapunov function. Since $H(q, p)$ is conserved under the Hamiltonian dynamics, it is natural to include $H(q, p)$ in the expression of the Lyapunov function. By simple calculation, for $T = \beta^{-1}$ we have

$$\mathcal{L}H(q, p) = -\gamma|p|^2 + \gamma T. \tag{2.15}$$

Here, $|p|^2$ term is not enough to bound the Hamiltonian $H(q, p)$. A simple trick is to add the $p \cdot q$ cross term into $H(q, p)$. Note that

$$\mathcal{L}(p \cdot q) = |p|^2 - \nabla U(q) \cdot q - \gamma p \cdot q \tag{2.16}$$

Using the assumption we obtain

$$\begin{aligned}\mathcal{L}(p\cdot q) &\leqslant |p|^2 - \frac{c_0}{2}U(q) - \frac{c_0}{2}|q|^2 + \gamma pq + C\\ &\leqslant C(|p|^2+1) - \frac{c_0}{2}U(q)\end{aligned}$$

Now choose c sufficiently small and define the Lyapunov function

$$W(q,p) = H(q,p) + cp\cdot q \tag{2.17}$$

then for any real number $k > 0$,

$$\begin{aligned}\mathcal{L}W + kW &= \mathcal{L}H + c\mathcal{L}(p\cdot q) + k\Big(U(q) + \frac{1}{2}|p|^2\Big)\\ &\leqslant -\gamma|p|^2 + \gamma T + c\bigg(C(|p|^2+1) - \frac{c_1}{2}U(q)\bigg) + k\Big(U(q) + \frac{1}{2}|p|^2\Big)\\ &\leqslant -(\gamma - cC - k)|p|^2 + \Big(k - \frac{cc_0}{2}\Big)U(q) + C.\end{aligned}$$

If we choose $k = cc_0/2$, this reduces to

$$\mathcal{L}W + \frac{cc_0}{2}W \leqslant -\Big(\gamma - cC - \frac{cc_0}{2}\Big)|p|^2 + C. \tag{2.18}$$

Therefore, by choosing c sufficient small we are able to obtain

$$\mathcal{L}W(q,p) \leqslant -\frac{cc_0}{2}W(q,p) + C, \tag{2.19}$$

which is the desired result. A detailed discussion of the second-order Langevin dynamics can be seen in [8]. Also note the the dissipation condition (2.13) only holds for regular potentials $U(q)$. The design of the Lyapunov function for singular potentials will be more technical.

2.4 Singular potentials

2.4.1 Methodology: exponential form

We follow [7] to construct the Lyapunov function $W(q,p)$ in the case of singular potentials. The underlying dynamics is still the second-order Langevin dyanmics (2.11), but the potential $U(q)$ might be singular. Assume the potential satisfies the admissible assumption

Assumption 2.1 (admissible) *A function $U : \mathbb{R}^d \to [0, +\infty]$ is an admissible potential if satisfies the normalzation condition and the following regularity and growth conditions:*

- *$U \in C^\infty(\mathscr{O})$ in its domian $\mathscr{O}$;*
- *$\mathscr{O}$ is an open and path-connected set. Moreover, for each $R > 0$, the level set*

$$\{q \in \mathbb{R}^d : U(q) < R\}$$

  is open and precompact.

- *For any sequence $\{q_k\} \subset \mathscr{O}$ with $U(q_k) \to \infty$, we have the following asymptotic properties:*

$$\nabla U(q_k) \to \infty, \quad \text{and} \quad \frac{|\nabla^2 U(q_k)|}{|\nabla U(q_k)|^2} \to 0$$

The LJ potential and the Coulomb potential with $d \geqslant 3$ are admissible, but the Coulomb potential with $d = 2$ is not admissible. Note that [8] constructs a Lyapunov function solving the case $d = 2$, but the approach is almost the same with [7], hence we only consider the construction in [7] here.

Recall that our goal is to construct a Lyapunov function $W(q,p)$ so that

$$\mathcal{L}W \leqslant -cW, \quad \text{when } W \geqslant R, \tag{2.20}$$

where $R > 0$ is a constant. Consider the Lyapunov function in the exponential form:

$$W = e^{bV}, \quad \text{where } V \text{ is a perturbation of } H, \tag{2.21}$$

then

$$\mathcal{L}W = bW[\mathcal{L}V + b\gamma T|\nabla_p V|^2] \tag{2.22}$$

Hence we only need to show

$$\mathcal{L}V + b\gamma T|\nabla_p V|^2 \leqslant -C, \quad \text{when } H \geqslant R. \tag{2.23}$$

2.4.2 Construction of V_0: intuition

To build an ituition how to construct V, we first consider the simplified problem: finding the Lypanov function V_0 such that

$$\mathcal{L}V_0 \leqslant -C, \quad \text{when } H \geqslant R. \tag{2.24}$$

Consider V_0 to be a perturbation of H, i.e., $V_0 = H + \psi$. Recall that

$$\mathcal{L}H = -\gamma p^2 + \gamma T. \tag{2.25}$$

When $|p|$ is large, $\mathcal{L}H$ is negative and (2.24) automatically holds true. In this case, we expect $\psi(q,p)$ to be close to 0. When $|p|$ is small, under the condition $H \geqslant R$, the potential $U(q)$ is large. In this case, we expect $\mathcal{L}\psi$ to produce negative terms, which requires to choose $\psi(q,p)$ carefully. Recall that the genertor $\mathcal{L}$ is given by

$$\mathcal{L} = p \cdot \nabla_q - \nabla_q U(q) \cdot \nabla_p - \gamma p \cdot \nabla_p + \gamma\beta^{-1}\Delta_p, \tag{2.26}$$

when $U(q)$ is very large, the operator

$$\mathcal{A} = -\nabla_q U(q) \cdot \nabla_p \tag{2.27}$$

dominates $\mathcal{L}$. Hence we choose $\psi(q,p)$ according to

$$\mathcal{A}\psi(q,p) = -\kappa, \quad \text{when } U \geqslant R, \tag{2.28}$$

whose solution is

$$\psi(q,p) = \kappa\frac{p \cdot \nabla U(q)}{|\nabla U(q)|^2} \tag{2.29}$$

Clearly, when $U(q)$ is large, $|\nabla U(q)|$ is also large, which implies $\psi(q,p)$ is a small perturbation.

For the case when $U(q)$ is small, we use a trunction scheme. For $R_2 > R_1 > 0$, introduce $\alpha \in C^\infty([0,+\infty)$ satisfying

$$\alpha(x) = \begin{cases} 1, & x \geqslant R_2 \\ 0, & x \leqslant R_1 \end{cases} \tag{2.30}$$

and $|\alpha'| \leqslant 2/(R_2 - R_1)$. Finally, $\psi(q,p)$ is constructed as

$$\psi(q,p) = \begin{cases} \kappa\alpha(U(q))\dfrac{p \cdot \nabla U(q)}{|\nabla U(q)|^2}, & \text{if } U(q) \geqslant R_1/2 \\ 0, & \text{otherwise} \end{cases} \tag{2.31}$$

Since $\psi(q,p)$ is close to 0 if $U(q) < R_1$ or $U(q)$ is sufficiently large, $\psi(q,p)$ can be viewed as a small perturbation of the zero function.

2.4.3 Construction of V: rigorous proof

We show that under appropriate choice of the constants R_1, R_2, κ, the Lyapunov condition holds.

$$W(q,p) = \exp(bH(q,p) + \psi(q,p)), \tag{2.32}$$

then by direct calculation,

$$\begin{aligned}\frac{\mathcal{L}W(q,p)}{W(q,p)} = &-b\gamma(1-bT)|p|^2 - \kappa\alpha(U(q)) + p \cdot \nabla_q \psi(q,p) + \\ &(2b-1)\psi(q,p) + \frac{\kappa^2\gamma T\alpha^2(U(q))}{|\nabla U(q)|^2} + \gamma bTNd\end{aligned} \tag{2.33}$$

Our goal is to prove the RHS of (2.33) is negative when $H(q,p)$ is large enough. Now we analyze the terms in (2.33) one by one:

- $-b\gamma(1-bT)|p|^2$: coming from the $-\gamma|p|^2$ term in $\mathcal{L}H = -\gamma|p|^2 + \gamma T$.
- $-\kappa\alpha(U(q))$: constant varying in $[-\kappa, 0]$. Close to $-\kappa$ when $U(q)$ is large enough.
- $p \cdot \nabla_q \psi(q,p)$: Since $\psi(q,p)$ is a small perturbation, this term should be small. In fact,

$$p \cdot \nabla_q \psi(q,p) \leqslant \kappa\alpha(U(q))|\nabla G(q)||p|^2 + \kappa|\alpha'(U(q))||p|^2, \tag{2.34}$$

  where $G(q) = \nabla U/|\nabla U|^2$.
- $\psi(q,p)$: This term is small and bounded by

$$|\psi(q,p)| \leqslant \frac{\kappa}{2C}|p|^2 + \frac{\kappa C}{2}\frac{\alpha^2(U(q))}{|\nabla U(q)|^2} \tag{2.35}$$

  We aim to choose C sufficiently large so that $|\psi(q,p)|$ produces small coefficient on $|p|^2$.

Concluding the estimates above, we obtain

$$\begin{aligned}\frac{\mathcal{L}W(q,p)}{W(q,p)} \leqslant &-\left\{b\gamma(1-bT) - \kappa\alpha(U(q))|\nabla G(q)| - \kappa\alpha'(U(q)) - \frac{|2bT-1|\gamma\kappa}{2C}\right\}|p|^2 \\ &- \kappa\alpha(U(q)) + \left(\frac{\kappa C}{2}|2bT-1|\gamma + \kappa^2\gamma T\right)\alpha^2(U(q))|G(q)|^2 + \gamma bTNd\end{aligned} \tag{2.36}$$

Now we choose the parameters κ, C, R_1, R_2 sequentially:

1. Pick $\kappa > 3\gamma N d$, so that $\kappa\alpha(U(q))$ can be larger than $\gamma b T N d$.
2. Pick $C > \dfrac{4|2bT-1|\kappa}{b(1-bT)}$, so that the coefficient of $|p|^2$ can be negative.
3. Choose R_1 sufficiently large so that

$$|\nabla G(q)| \leqslant \frac{b\gamma(1-bT)}{8\kappa}, \qquad \left(\frac{\kappa C}{2}|2bT-1|\gamma + \kappa^2\gamma T\right)|G(q)|^2 \leqslant \gamma b T N d, \quad \text{when } U(q) \geqslant R_1/2$$

   With the definition of $U(q)$, we have for any $q \in \mathbb{R}^d$,

$$\alpha(U(q))|\nabla G(q)| \leqslant \frac{b\gamma(1-bT)}{8\kappa}, \qquad \alpha(U(q))\left(\frac{\kappa C}{2}|2bT-1|\gamma + \kappa^2\gamma T\right)|G(q)|^2 \leqslant \gamma b T N d,$$

4. Pick $R_2 > R_1$ such that

$$|\alpha'(U(q))| \leqslant \frac{b\gamma(1-bT)}{8\kappa}.$$

Using these estimates, we obtain the global estimate

$$\frac{\mathcal{L}W(q,p)}{W(q,p)} \leqslant -\frac{b\gamma(1-bT)}{2}|p|^2 - \kappa\alpha(U(q)) + 2\gamma b T N d \tag{2.37}$$

for any p, q. Therefore, as long as

$$|p|^2 > \frac{6\gamma N d}{b\gamma(1-bT)} \quad \text{or} \quad U(q) > R_2$$

one is able to obtain

$$\frac{\mathcal{L}W(q,p)}{W(q,p)} \leqslant -\gamma N d, \tag{2.38}$$

which is the desired result.

Finally, note that the Lyapunov function for the Coulomb potential with $d = 2$ suggested in [8] is given by

$$W(q,p) = \exp(bH(q,p) + \psi(q,p)), \tag{2.39}$$

where

$$\psi(q,p) = -\frac{b}{N}\sum_{1\leqslant i\neq j\leqslant N} \frac{(p_i - p_j)\cdot(q_i - q_j)}{|q_i - q_j|} + cp\cdot q. \tag{2.40}$$

References

[1] TE Harris. The existence of stationary measures for certain. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability: Held at the Statistical Laboratory, University of California, December, 1954, July and August, 1955*, volume 2, page 113. Univ of California Press, 1956.

[2] Sean P Meyn and Richard L Tweedie. *Markov chains and stochastic stability*. Springer Science & Business Media, 2012.

[3] Jonathan C Mattingly, Andrew M Stuart, and Desmond J Higham. Ergodicity for sdes and approximations: locally lipschitz vector fields and degenerate noise. *Stochastic processes and their applications*, 101(2):185–232, 2002.

[4] Cédric Villani. Hypocoercivity. *arXiv preprint math/0609050*, 2006.

[5] Tony Lelievre and Gabriel Stoltz. Partial differential equations and stochastic methods in molecular dynamics. *Acta Numerica*, 25:681–880, 2016.

[6] Yu Cao, Jianfeng Lu, and Lihan Wang. On explicit l^2-convergence rate estimate for underdamped langevin dynamics. *arXiv preprint arXiv:1908.04746*, 2019.

[7] David P Herzog and Jonathan C Mattingly. Ergodicity and lyapunov functions for langevin dynamics with singular potentials. *Communications on Pure and Applied Mathematics*, 72(10):2231–2255, 2019.

[8] Yulong Lu and Jonathan C Mattingly. Geometric ergodicity of langevin dynamics with coulomb interactions. *Nonlinearity*, 33(2):675, 2019.

[9] Ben Cooke, David P Herzog, Jonathan C Mattingly, Scott A McKinley, and Scott C Schmidler. Geometric ergodicity of two–dimensional hamiltonian systems with a lennard–jones–like repulsive potential. *arXiv preprint arXiv:1104.3842*, 2011.

[10] Sebastian J Vollmer, Konstantinos C Zygalakis, and Yee Whye Teh. Exploration of the (non-) asymptotic bias and variance of stochastic gradient langevin dynamics. *The Journal of Machine Learning Research*, 17(1):5504–5548, 2016.

[11] Andreas Eberle. Reflection coupling and wasserstein contractivity without convexity. *Comptes Rendus Mathematique*, 349(19-20):1101–1104, 2011.

[12] Andreas Eberle. Reflection couplings and contraction rates for diffusions. *Probability theory and related fields*, 166(3):851–886, 2016.

[13] Andreas Eberle, Arnaud Guillin, and Raphael Zimmer. Couplings and quantitative contraction rates for langevin dynamics. *The Annals of Probability*, 47(4):1982–2010, 2019.

[14] Nawaf Bou-Rabee, Andreas Eberle, and Raphael Zimmer. Coupling and convergence for hamiltonian monte carlo. *The Annals of applied probability*, 30(3):1209–1250, 2020.

[15] Nawaf Bou-Rabee and Andreas Eberle. Two-scale coupling for preconditioned hamiltonian monte carlo in infinite dimensions. *Stochastics and Partial Differential Equations: Analysis and Computations*, 9(1):207–242, 2021.

[16] Nawaf Bou-Rabee and Andreas Eberle. Couplings for andersen dynamics. *arXiv preprint arXiv:2009.14239*, 2020.

[17] Andreas Eberle, Arnaud Guillin, and Raphael Zimmer. Quantitative harris-type theorems for diffusions and mckean–vlasov processes. *Transactions of the American Mathematical Society*, 371(10):7135–7173, 2019.

[18] Wei Liu, Liming Wu, and Chaoen Zhang. Long-time behaviors of mean-field interacting particle systems related to mckean–vlasov equations. *Communications in Mathematical Physics*, 387(1):179–214, 2021.

[19] Shi Jin, Lei Li, Xuda Ye, and Zhennan Zhou. Ergodicity and long-time behavior of the random batch method for interacting particle systems. *arXiv preprint arXiv:2202.04952*, 2022.

[20] Ingo Lütkebohle. Topics in Stochastic Processes. `http://probability.ca/jeff/teaching/1718/sta4502/`, 2018. [Online; accessed 18-Feb-2022].

[21] Martin Hairer and Jonathan C Mattingly. Yet another look at harris' ergodic theorem for markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications VI*, pages 109–117. Springer, 2011.

[22] MIT OpenCourseWare. Topics in Stochastic Processes. `https://ocw.mit.edu/courses/sloan-school-of-management/15-070j-advanced-stochastic-processes-fall-2013/lecture-notes/`, 2013. [Online; accessed 19-Feb-2022].

[23] Daniel W Stroock and SRS Varadhan. On the support of diffusion processes with applications to. *Probability Theory*, 3:333, 1972.

[24] Daniel W Stroock and SRS Varadhan. Probability theory and the strong maximum principle. In *Partial differential equations (Proc. Sympos. Pure Math., Vol. XXIII, Univ. California, Berkeley, Calif., 1971)*, pages 215–220, 1973.

[25] José A Cañizo, Chuqi Cao, Josephine Evans, and Havva Yoldaş. Hypocoercivity of linear kinetic equations via harris's theorem. *arXiv preprint arXiv:1902.10588*, 2019.