A quantitative Mac Diarmid's inequality for geometrically ergodic Markov chains

D. Belomestny, E. Moulines and S. Samsonov

January 20, 2020

Abstract

We state and prove a quantitative version of the bounded difference inequality for geometrically ergodic Markov chains. Our proof uses the same martingale decomposition as [3] but, compared to this paper, the exact coupling argument is modified to fill a gap between the strongly aperiodic case and the general aperiodic case.

Keywords: Concentration inequalities ; Markov chains ; Geometric ergodicity ; Coupling.
AMS MSC 2010: 60J05; 60E15.

1 Introduction

The purpose of this note is to establish a quantitative version of Mc Diarmid's inequality for geometrically ergodic Markov chains. Let $X_0, \ldots, X_{n-1}$ denote independent random variables taking values in a measurable space $(\mathsf{X}, \mathscr{X})$ and $c = (c_0, \ldots, c_{n-1})$ denote a vector of non-negative real numbers. A function $f : \mathsf{X}^n \to \mathbb{R}$ satisfies the bounded difference inequality if for all $x = (x_0, \ldots, x_{n-1})$ and $y = (y_0, \ldots, y_{n-1}) \in \mathsf{X}^n$, we have

$$|f(x) - f(y)| \leqslant \sum_{i=0}^{n-1} c_i \mathbb{1}_{\{x_i \neq y_i\}} \, . \tag{1}$$

The bounded difference inequality, first established in [8], shows that for all $t > 0$,

$$\mathbb{P}\big(f(X_0, \ldots, X_{n-1}) - \mathbb{E}[f(X_0, \ldots, X_{n-1})] > t\big) \leqslant \mathrm{e}^{-2t^2/\|c\|^2} \, ,$$

where $\|c\|^2 = \sum_{i=0}^{n-1} c_i^2$. Several attempts have been made to extend this result to Markov chains. In [1], the concentration of particular functionals of the form $f(x_0, \ldots, x_{n-1}) = \sup_{g \in \mathscr{F}} \sum_{i=0}^{n-1} g(x_i)$, for centered functions g in a class $\mathscr{F}$ is established. The concentration of general functionals (satisfying (1)) of geometrically ergodic Markov chains was established in [3], where it is also proved that geometric ergodicity is a necessary assumption. However, the result in [3] is not quantitative. It states that for all geometrically recurrent set C, there exists a constant β, depending on C such that for all $x \in C$ and $t > 0$,

$$\mathbb{P}_x\big(f(X_0, \ldots, X_{n-1}) - \mathbb{E}_x[f(X_0, \ldots, X_{n-1})] > t\big) \leqslant \mathrm{e}^{-\beta t^2/\|c\|^2} \, , \tag{2}$$

where for any $x \in \mathscr{X}$, $\mathbb{P}_x$ is the distribution of the Markov chain $\{X_k\}_{k=0}^{\infty}$ starting from x (see the precise definition below). In many applications, it is necessary to get the explicit dependence of the constant β as a function of the set C. In particular, this problem arises when establishing posterior concentration rates of Bayesian non-parametric estimators; see for example [11, 6] for recent accounts on this theory. To extend these results to Markovian settings, the result of [3] cannot be applied directly and a quantitative version of (2) is required, where the dependence of β on constants characterizing the mixing of the Markov chain is needed; see for example [12, 7].

A quantitative version of Mc Diarmid's inequality for Markov chains was established in [9], where the constant β depends here explicitly on the mixing time of the chain. The existence of finite mixing times requires *uniform* ergodicity of the chain, see for example [10, Section 3.3], an assumption that typically fails when the chain takes value in general state spaces. In this note, we prove an extension of Mc Diarmid's inequality to geometrically ergodic Markov chains. Our proof is based on [3], but avoids the use of [3, Lemma 6] which requires the construction of an exact coupling. Exact coupling can actually be built in the strongly aperiodic case but there is a gap in the general aperiodic case.

The remaining of the paper is decomposed as follows, Section 2 introduces formally the notations and the assumptions of the main result, which is stated and proved in Section 3.

2 Notations and assumptions

Let $(\mathsf{X}, \mathscr{X})$ be a measurable space. We denote by d_{TV} the total variation distance between probability measures. For any sequence $x = \{x_n,\ n \in \mathbb{N}\}$ and any non-negative integers a and b, with $a \leqslant b$, let $x_a^b = (x_a, x_{a+1}, \ldots, x_b)$. For any $n \geqslant 0$ and any vector $c = c_0^{n-1} \in \mathbb{R}^n$, let $\|c\|$ denote the Euclidean norm of c and $\|c\|_\infty = \max_{0 \leqslant i \leqslant n-1} |c_i|$ denote its sup-norm.

We denote by $(\mathsf{X}^{\mathbb{Z}_+}, \mathscr{X}^{\otimes \mathbb{Z}_+}, (\mathscr{F}_k)_{k \geqslant 0})$ the canonical filtered space, $\{X_n\}_{n=0}^\infty$ the canonical process and $\theta : \mathsf{X}^{\mathbb{Z}_+} \to \mathsf{X}^{\mathbb{Z}_+}$ the shift operator on the canonical space defined, for any $x = (x_n)_{n \geqslant 0} \in \mathsf{X}^{\mathbb{Z}_+}$ by $\theta(x) \in \mathsf{X}^{\mathbb{Z}_+}$, where, for any $n \geqslant 0$, $\theta(x)_n = x_{n+1}$. Set $\theta_1 = \theta$ and for $n \in \mathbb{N}^*$, define inductively, $\theta_n = \theta_{n-1} \circ \theta$. We also need to define θ_∞. To this aim, fix an arbitrary $x^* \in \mathsf{X}$, we define $\theta_\infty : \mathsf{X}^{\mathbb{N}} \to \mathsf{X}^{\mathbb{N}}$ such that for $z = \{z_k,\ k \in \mathbb{N}\} \in \mathsf{X}^{\mathbb{N}}$, $\theta_\infty z \in \mathsf{X}^{\mathbb{N}}$ is the constant sequence $(\theta_\infty z)_k = x^*$ for all $k \in \mathbb{N}$.

Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. For any probability measure ξ on $(\mathsf{X}, \mathscr{X})$, denote by $\mathbb{P}_\xi$ the unique probability under which $(X_n)_{n \geqslant 0}$ is a Markov chain with Markov kernel P and initial distribution ξ and let $\mathbb{E}_\xi$ denote the expectation under the distribution $\mathbb{P}_\xi$. Recall that $\mathscr{F}_n$ denotes the σ-algebra generated by $X_0, \ldots, X_n$. For any $x \in \mathsf{X}$, let δ_x denote the Dirac mass at point x. With some abuse of notation, we also denote $\mathbb{P}_x$ (resp. $\mathbb{E}_x$) instead of $\mathbb{P}_{\delta_x}$ (resp. $\mathbb{E}_{\delta_x}$).

For any $\mathsf{B} \in \mathscr{X}$ and any integer $i \geqslant 0$, let

$$\tau_{\mathsf{B}}^i = \inf\{n \geqslant i : X_n \in \mathsf{B}\} = i + \tau_{\mathsf{B}}^0 \circ \theta^i \qquad \text{and} \qquad \sigma_{\mathsf{B}} = \tau_{\mathsf{B}}^1 = 1 + \tau_{\mathsf{B}}^0 \circ \theta\,.$$

For $c = c_0^{n-1} \in \mathbb{R}_+^n$, we denote by $\mathbb{BD}(\mathsf{X}^n, c)$ the set of measurable functions $f : \mathsf{X}^n \to \mathbb{R}$ such that for all $x = (x_0, \ldots, x_{n-1})$ and $y = (y_0, \ldots, y_{n-1})$, $|f(x) - f(y)| \leqslant \sum_{i=0}^{n-1} c_i \mathbb{1}_{\{x_i \neq y_i\}}$ The main result is established under the following conditions.

H1 The Markov kernel P is irreducible and aperiodic, with unique invariant probability π.

H2 (drift condition) There exist a measurable function $V : X \to [1; +\infty)$, real numbers $\lambda \in (0,1)$, $b > 0$ and a non-empty set $\mathsf{C} \in \mathscr{X}$ such that for any $x \in X$,

$$PV(x) \leq \lambda V(x) + b\mathbb{1}_{\{x \in \mathsf{C}\}}$$

H3 (V-ergodicity) There exist $r \in (0,1)$, $L \geqslant 1$ and function $V : X \to \mathbb{R}$ such that for any $x \in X$ and any $n \geqslant 0$, it holds

$$d_V(\delta_x P^n, \pi) \leqslant L V(x) r^n \;,$$

where π is the unique invariant measure from **H1**.

When the Markov kernel P is uniformly ergodic, then **H3** holds with $V(x) = 1$.

Note that [4, Proposition 14.3.1] implies that if Markov kernel P satisfies the drift condition **H2** with the set $\mathsf{C} \in \mathscr{X}$, $\mathsf{C} = \{x : V(x) < d\}$, then it holds for any $x \in \mathsf{C}$ that

$$\sup_{x \in \mathsf{C}} \mathbb{E}_x[\lambda^{-\sigma_{\mathsf{C}}}] \leqslant M \tag{3}$$

with some M depending on $V(x)$ and d.

The following Lemma is a coupling result that replaces [3, Lemma 6]. It is instrumental in the sequel.

Lemma 1. *For any probability measures ξ and ξ' on $(\mathsf{X}, \mathscr{X})$, any $n \geqslant 1$, any $c \in \mathbb{R}^n_+$ and any $h \in \mathbb{BD}(\mathsf{X}^n, c)$,*

$$|\mathbb{E}_\xi[h(X_0^{n-1})] - \mathbb{E}_{\xi'}[h(X_0^{n-1})]| \leqslant 2 \sum_{i=0}^{n-1} c_i \mathrm{d}_{\mathrm{TV}}(\xi P^i, \xi' P^i) \leqslant 2 \sum_{i=0}^{n-1} c_i \mathrm{d}_{\mathrm{V}}(\xi P^i, \xi' P^i) \;.$$

Remark 2. *It is possible to avoid the factor* 2 *in* (1) *under additional technical conditions, for example, when there exists a maximal coupling for $(\mathbb{P}_\xi, \mathbb{P}_{\xi'})$, see [4, Lemma 23.2.1].*

Proof. Fix an arbitrary $x^* \in \mathsf{X}$. For $i \in \{1, \dots, n-1\}$, we set $\bar{h}_i(x_i^{n-1}) = h(x^*, \dots, x^*, x_i^{n-1})$. By convention, we set $\bar{h}_n$ the constant function $\bar{h}_n = h(x^*, \dots, x^*)$ and $\bar{h}_0 = h$. With these notations, we have the decomposition

$$h(x_0^{n-1}) = \sum_{i=0}^{n-1} \{\bar{h}_i(x_i^{n-1}) - \bar{h}_{i+1}(x_{i+1}^{n-1})\} + \bar{h}_n \;.$$

For all $i \in \{0, \dots, n-1\}$ and all $x_i \in \mathsf{X}$, let

$$\begin{aligned} \bar{w}_i(x_i) &= \int \left\{\bar{h}_i(x_i^{n-1}) - \bar{h}_{i+1}(x_{i+1}^{n-1})\right\} \prod_{\ell=i+1}^{n-1} P(x_{\ell-1}, \mathrm{d}x_\ell) \;, \\ &= \int \left\{h(x^*, \dots, x^*, x_i^{n-1}) - h(x^*, \dots, x^*, x_{i+1}^{n-1})\right\} \prod_{\ell=i+1}^{n-1} P(x_{\ell-1}, \mathrm{d}x_\ell) \;. \end{aligned} \tag{4}$$

It is easily seen that $\mathbb{E}\left[\{\bar{h}_i(X_i^{n-1}) - \bar{h}_{i+1}(X_{i+1}^{n-1})\} \,\middle|\, \mathscr{F}_i\right] = \bar{w}_i(X_i)$, $\mathbb{P}_\xi - a.s.$, which implies that

$$\mathbb{E}_\xi\left[h(X_0^{n-1})\right] = \sum_{i=0}^{n-1} \xi P^i \bar{w}_i + \bar{h}_n \;.$$

Since $h \in \mathbb{BD}(\mathsf{X}^n, c)$, (4) shows that $|\bar{w}_i|_\infty \leq c_i$. Therefore,

$$|\mathbb{E}_\xi\left[h(X^{n-1})\right] - \mathbb{E}_{\xi'}\left[h(X^{n-1})\right]| \leq \sum_{i=0}^{n-1} |\xi P^i \bar{w}_i - \xi' P^i \bar{w}_i| \leq 2 \sum_{i=0}^{n-1} c_i \mathrm{d}_{\mathrm{TV}}(\xi P^i, \xi' P^i) \;.$$

The result now follows from the trivial bound $\mathrm{d}_{\mathrm{TV}}(\xi, \eta) \leq \mathrm{d}_{\mathrm{V}}(\xi, \eta)$ for arbitrary measures ξ, η given that $V(x) \geq 1$ for any $x \in X$. □

3 Main result

The main result of this paper is the following quantitative version of Mac Diarmid's inequality for $V-$ergodic Markov chains under certain drift condition.

Theorem 3. *Assume* **H1**, **H2** *and* **H3** *holds with drift function* $V(x)$. *Let* M *be a constant from 3,* $\mathsf{C} = \{V(x) < d\}$, $n \geqslant 1$, $c \in \mathbb{R}^n$ *and* $f \in \mathbb{BD}(\mathsf{X}^n, c)$. *Then, for all* $x \in \mathsf{C}$ *and* $t > 0$,

$$\mathbb{P}_x\big(f(X_0^{n-1}) - \mathbb{E}_x[f(X_0^{n-1})] > t\big) \leqslant \exp\left(-\frac{\beta t^2}{\|c\|^2}\right),$$

where β *is given by*

$$\beta = \frac{(1 - r \vee \lambda^{1/4})^2}{16Ld}\left(-\frac{5}{\log \lambda} + 4MLd\right)^{-1}.$$

Proof of Theorem 3. Fix $c \in \mathbb{R}^n$, $x \in \mathsf{X}$ and $f \in \mathbb{BD}(\mathsf{X}^n, c)$. Note that drift condition **H3** implies that

$$\sup_{x \in \mathsf{C}} \mathbb{E}_x[\lambda^{-\sigma_\mathsf{C}}] \leqslant M$$

Following [3], we decompose $f(X_0^{n-1}) - \mathbb{E}_x[f(X_0^{n-1})]$ into martingale increments by conditioning to the stopping times τ_C^i, $i = 0, \ldots, n-1$. For any integer $i \in [0, n-1]$, define

$$G_i = \mathbb{E}_x\left[f(X_0^{n-1}) | \mathscr{F}_{\tau_\mathsf{C}^i}\right].$$

As $\tau_\mathsf{C}^0 = 0$ $\mathbb{P}_x$-a.s., it holds $\mathbb{E}_x[f(X_0^{n-1})] = \mathbb{E}_x[f(X_0^{n-1}) | \mathscr{F}_{\tau_\mathsf{C}^0}] = G_0$. Moreover, as $\tau_\mathsf{C}^{n-1} \geqslant n-1$, it also holds $G_{n-1} = \mathbb{E}_x[f(X_0^{n-1}) | \mathscr{F}_{\tau_\mathsf{C}^{n-1}}] = f(X_0^{n-1})$. Therefore, the difference $f(X_0^{n-1}) - \mathbb{E}_x[f(X_0^{n-1})]$ is decomposed into a sum of the martingale increments $G_{i+1} - G_i$ as follows

$$f(X_0^{n-1}) - \mathbb{E}_x[f(X_0^{n-1})] = G_{n-1} - G_0 = \sum_{i=0}^{n-2} (G_{i+1} - G_i). \tag{5}$$

The proof is now decomposed into three facts that aim at bounding the Laplace transform of $f(X_0^{n-1}) - \mathbb{E}_x[f(X_0^{n-1})]$.
Fact 1. *For any* $i \in \{1, \ldots, n-1\}$,

$$G_i - G_{i-1} = (G_i - G_{i-1}) \mathbb{1}_{\{\tau_\mathsf{C}^{i-1} = i-1\}}. \tag{6}$$

Proof of Fact 1. By definition $\tau_\mathsf{C}^{i-1} \geqslant i-1$ and $\tau_\mathsf{C}^{i-1} > i-1$ if and only if $\tau_\mathsf{C}^{i-1} = \tau_\mathsf{C}^i$. Therefore,

$$G_i - G_{i-1} = (G_i - G_{i-1}) \left(\mathbb{1}_{\{\tau_\mathsf{C}^{i-1} = i-1\}} + \mathbb{1}_{\{\tau_\mathsf{C}^{i-1} = \tau_\mathsf{C}^i\}}\right).$$

To prove that $(G_i - G_{i-1}) \mathbb{1}_{\{\tau_\mathsf{C}^{i-1} = \tau_\mathsf{C}^i\}} = 0$, we decompose according to the values of τ_C^i:

$$(G_i - G_{i-1}) \mathbb{1}_{\{\tau_\mathsf{C}^{i-1} = \tau_\mathsf{C}^i\}} = \sum_{j \geqslant i} (G_i - G_{i-1}) \mathbb{1}_{\{\tau_\mathsf{C}^{i-1} = \tau_\mathsf{C}^i = j\}}.$$

Now, remark that, for any $i \geqslant 0$,

$$G_i \mathbb{1}_{\{\tau_\mathsf{C}^i = j\}} = \begin{cases} \mathbb{E}_x\left[f(X_0^{n-1}) | \mathscr{F}_j\right] & \text{if } j \leqslant n-2, \\ f(X_0^{n-1}) & \text{if } j \geqslant n-1. \end{cases} \tag{7}$$

Then, for any $j \geqslant i$,

$$G_i \mathbb{1}_{\{\tau_{\mathsf{C}}^i = j\}} \mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1} = \tau_{\mathsf{C}}^i\}} = G_{i-1} \mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1} = j\}} \mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1} = \tau_{\mathsf{C}}^i\}} = G_{i-1} \mathbb{1}_{\{\tau_{\mathsf{C}}^i = j\}} \mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1} = \tau_{\mathsf{C}}^i\}} .$$

This proves Fact 1. □

Fact 2. bounds the increments $G_i - G_{i-1}$. The proof relies on the following lemma which is a consequence of the coupling result Lemma 1. Define $g_{n-1} = g_{n-1,\pi} = f$ and, for any $i \in [0, n-2]$, let g_i and $g_{i,\pi}$ denote the functions defined for any $x_0^i \in \mathsf{X}^{i+1}$ by

$$g_i(x_0^i) = \mathbb{E}_{x_i}[f(x_0^i, X_1^{n-1-i})], \qquad g_{i,\pi}(x_0^i) = \mathbb{E}_{\pi}[f(x_0^i, X_1^{n-1-i})] . \tag{8}$$

Lemma 4. *Assume* **H1**, **H2**, **H3** *and let* $\mathsf{C} = \{x : V(x) < d\}$. *For any* $i \in \{0, \dots, n-1\}$ *and* (x_0^{i-1}, x_i) *in* $\mathsf{X}^i \times \mathsf{C}$,

$$|g_i(x_0^i) - g_{i,\pi}(x_0^i)| \leqslant 2LV(x_i) \sum_{j=i+1}^{n-1} c_j r^{j-i} \leqslant 2Ld \sum_{j=i+1}^{n-1} c_j r^{j-i} . \tag{9}$$

Proof. Fix $i \in \{0, \dots, n-1\}$ and $x_0^i \in \mathsf{X}^{i+1}$. As $f \in \mathbb{BD}(\mathsf{X}^n, c)$, the function $\widetilde{f}_i : y_1^{n-1-i} \in \mathsf{X}^{n-1-i} \mapsto f(x_0^i, y_1^{n-1-i}) \in \mathbb{R}$ satisfies

$$|\widetilde{f}_i(y_1^{n-1-i}) - \widetilde{f}_i(z_1^{n-1-i})| \leqslant \sum_{k=1}^{n-1-i} c_{i+k} \mathbb{1}_{\{y_k \neq z_k\}} .$$

Hence, $\widetilde{f}_i \in \mathbb{BD}(\mathsf{X}^{n-1-i}, c_{i+1:n-1})$. Applying Lemma 1 to the function $h = \widetilde{f}_i$ yields

$$\begin{aligned}|g_i(x_0^i) - g_{i,\pi}(x_0^i)| &= |\mathbb{E}_{x_i}[f(x_0^i, X_1^{n-1-i})] - \mathbb{E}_{\pi}[f(x_0^i, X_1^{n-1-i})]| \\ &= |\mathbb{E}_{x_i}[\widetilde{f}_i(X_1^{n-1-i})] - \mathbb{E}_{\pi}[\widetilde{f}_i(X_1^{n-1-i})]| \leqslant 2 \sum_{j=i+1}^{n-1} c_j \mathrm{d_V}(\delta_{x_i} P^{j-i}, \pi) .\end{aligned}$$

Now inequality (9) follows from **H3**. □

Fact 2. *Let ρ such that $r \leqslant \rho < 1$ and $i \in \{1, \dots, n-1\}$. Then,*

$$|G_i - G_{i-1}| \leqslant C_1 \|c\|_\infty \mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1} = i-1\}} \sigma_{\mathsf{C}} \circ \theta^{i-1} , \tag{10}$$

$$|G_i - G_{i-1}|^2 \leqslant C_2 \mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1} = i-1\}} \frac{1}{\rho^{2\sigma_{\mathsf{C}} \circ \theta^{i-1}}} \sum_{k=i}^{n-1} c_k^2 \rho^{k-i} . \tag{11}$$

where, $C_1 = 5Ld/(1-r)$ and $C_2 = 16L^2d^2/(1-\rho)$.

Proof of Fact 2. For any integer $i \in \{1, \dots, n\}$, let

$$G_{i,1} = \mathbb{E}_x[f(X_0^{n-1}) | \mathscr{F}_{\tau_{\mathsf{C}}^{i-1}}] \mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1} = i-1\}}, \qquad G_{i,2} = \mathbb{E}_x[f(X_0^{n-1}) | \mathscr{F}_{\tau_{\mathsf{C}}^i}] \mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1} = i-1\}} .$$

From Fact 1., $G_i - G_{i-1} = G_{i,2} - G_{i,1}$. By Markov's property, for any $i \in \{0, \dots, n-1\}$ and $x \in \mathsf{X}$,

$$\mathbb{E}_x[f(X_0^{n-1}) | \mathscr{F}_i] = g_i(X_{0:i}), \qquad \mathbb{P}_x - \text{a.s.} .$$

Now, let $R_{i,1} = g_{i-1}(X_0^{i-1})\mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1}=i-1\}} - g_{i-1,\pi}(X_0^{i-1})\mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1}=i-1\}}$. We have

$$\begin{aligned}G_{i,1} &= \mathbb{E}_x[f(X_0^{n-1})|\mathscr{F}_{\tau_{\mathsf{C}}^{i-1}}]\mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1}=i-1\}} = \mathbb{E}_x[f(X_0^{n-1})|\mathscr{F}_{i-1}]\mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1}=i-1\}} \\ &= g_{i-1}(X_0^{i-1})\mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1}=i-1\}} = g_{i-1,\pi}(X_0^{i-1})\mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1}=i-1\}} + R_{i,1}\ .\end{aligned} \tag{12}$$

Moreover, as $\tau_{\mathsf{C}}^i \geqslant i$, by (7),

$$\begin{aligned}G_{i,2} &= \sum_{j\geqslant i}\mathbb{E}_x[f(X_0^{n-1})|\mathscr{F}_{\tau_{\mathsf{C}}^i}]\mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1}=i-1\}}\mathbb{1}_{\{\tau_{\mathsf{C}}^i=j\}} \\ &= \sum_{j=i}^{n-2} g_j(X_0^j)\mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1}=i-1,\tau_{\mathsf{C}}^i=j\}} + f(X_0^{n-1})\mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1}=i-1,\tau_{\mathsf{C}}^i\geqslant n-1\}}\ .\end{aligned} \tag{13}$$

Let $R_{i,2} = \sum_{j=i}^{n-2}(g_j(X_0^j) - g_{j,\pi}(X_0^j))\mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1}=i-1,\tau_{\mathsf{C}}^i=j\}}$. From (12) and (13),

$$\begin{aligned}|G_{i,2} - G_{i,1}| = |R_{i,2} - R_{i,1} &+ \sum_{j=i}^{n-2}(g_{j,\pi}(X_0^j) - g_{i-1,\pi}(X_0^{i-1}))\mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1}=i-1,\tau_{\mathsf{C}}^i=j\}} \\ &+ (f(X_0^{n-1}) - g_{i-1,\pi}(X_0^{i-1}))\mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1}=i-1,\tau_{\mathsf{C}}^i\geqslant n-1\}}|\ .\end{aligned} \tag{14}$$

We bound separately all the terms in this decomposition. First, as π is invariant and $f \in \mathbb{BD}(\mathsf{X}^n, c)$, for any $j \in \{i+1,\ldots,n-1\}$ and any $x_0^j \in \mathsf{X}^{j+1}$,

$$|g_{j,\pi}(x_0^j) - g_{i-1,\pi}(x_0^{i-1})| = \mathbb{E}_\pi[f(x_0^j, X_{j+1}^{n-1}) - f(x_0^{i-1}, X_i^{n-1})] \leqslant \sum_{k=i}^{j} c_k\ .$$

Hence,

$$\begin{aligned}&\sum_{j=i}^{n-2}|(g_{j,\pi}(X_0^j) - g_{i-1,\pi}(X_0^{i-1}))|\mathbb{1}_{\{\tau_{\mathsf{C}}^i=j\}} \leqslant \sum_{j=i}^{n-2}\mathbb{1}_{\{\tau_{\mathsf{C}}^i=j\}}\sum_{k=i}^{j} c_k = \mathbb{1}_{\{\tau_{\mathsf{C}}^i\leqslant n-2\}}\sum_{k=i}^{\tau_{\mathsf{C}}^i} c_k\ , \\ &|f(X_0^{n-1}) - g_{i-1,\pi}(X_0^{i-1})|\mathbb{1}_{\{\tau_{\mathsf{C}}^i\geqslant n-1\}} \leqslant \mathbb{1}_{\{\tau_{\mathsf{C}}^i\geqslant n-1\}}\sum_{k=i}^{n-1} c_k\ .\end{aligned} \tag{15}$$

To bound $|R_{i,1}|$ and $|R_{i,2}|$ in (14), we use Lemma 4. First, (9) directly yields

$$|R_{i,1}| \leqslant 2\mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1}=i-1\}} Ld \sum_{j=i+1}^{n-1} c_j r^{j-i}\ . \tag{16}$$

Moreover, as $\{\tau_{\mathsf{C}}^i = j\} \subset \{X_j \in \mathsf{C}\}$, (9) also yields

$$(g_j(X_0^j) - g_{j,\pi}(X_0^j))\mathbb{1}_{\{\tau_{\mathsf{C}}^i=j\}} \leqslant 2L\sum_{k=j+1}^{n-1} c_k r^{k-j}\mathbb{1}_{\{\tau_{\mathsf{C}}^i=j\}} \leqslant 2Ld\mathbb{1}_{\{\tau_{\mathsf{C}}^i=j\}}\sum_{k=\tau_{\mathsf{C}}^i+1}^{n-1} c_k r^{k-\tau_{\mathsf{C}}^i}\ .$$

Therefore,

$$|R_{i,2}| \leqslant 2Ld\mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1}=i-1\}}\sum_{k=\tau_{\mathsf{C}}^i+1}^{n-1} c_k r^{k-\tau_{\mathsf{C}}^i}\ . \tag{17}$$

Plugging (15), (16) and (17) in (14) yields

$$
(18) \qquad |G_{i,2}-G_{i,1}| \leqslant 2Ld\left(\sum_{j=i+1}^{n-1} c_j r^{j-i} + \sum_{k=\tau_{\mathsf{C}}^i+1}^{n-1} c_k r^{k-\tau_{\mathsf{C}}^i} + \frac{1}{2Ld}\sum_{k=i}^{\tau_{\mathsf{C}}^i\wedge(n-1)} c_k\right)\mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1}=i-1\}} \,.
$$

Both (10) and (11) follow from (18) by bounding separately the 3 terms in the right-hand side of this inequality. Let us first establish (10). Since $r<1$,

$$
\sum_{j=i+1}^{n-1} c_j r^{j-i} \leqslant \frac{\|c\|_\infty r}{1-r}, \qquad \sum_{k=\tau_{\mathsf{C}}^i+1}^{n-1} c_k r^{k-\tau_{\mathsf{C}}^i} \leqslant \frac{\|c\|_\infty r}{1-r} \,.
$$

Moreover,

$$
\sum_{k=i}^{\tau_{\mathsf{C}}^i\wedge(n-1)} c_k \leqslant \|c\|_\infty[1-i+\tau_{\mathsf{C}}^i\wedge(n-1)] \leqslant \|c\|_\infty[1+\tau_{\mathsf{C}}^0\circ\theta^i] = \|c\|_\infty \sigma_{\mathsf{C}}\circ\theta^{i-1} \,.
$$

As $r<1\leqslant \sigma_{\mathsf{C}}\circ\theta^{i-1}$, plugging these upper bounds in (18) shows

$$
|G_i-G_{i-1}| = |G_{i,2}-G_{i,1}| \leqslant \frac{5Ld\|c\|_\infty}{1-r}\sigma_{\mathsf{C}}\circ\theta^{i-1}\mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1}=i-1\}} \,.
$$

This proves (10). We use slightly different controls to prove (11) from (18). As $r\leqslant\rho<1$, $\rho^{-\sigma_{\mathsf{C}}\circ\theta^{i-1}}\geqslant 1$, and

$$
(19) \qquad \sum_{j=i+1}^{n-1} c_j r^{j-i} \leqslant \sum_{j=i}^{n-1} c_j\rho^{j-i} \leqslant \rho^{-\sigma_{\mathsf{C}}\circ\theta^{i-1}}\sum_{j=i}^{n-1} c_j\rho^{j-i} \,.
$$

Moreover,

$$
\sum_{k=\tau_{\mathsf{C}}^i+1}^{n-1} c_k r^{k-\tau_{\mathsf{C}}^i} \leqslant \rho^{i-\tau_{\mathsf{C}}^i}\sum_{k=\tau_{\mathsf{C}}^i+1}^{n-1} c_k\rho^{k-i} \,.
$$

As $\tau_{\mathsf{C}}^i\geqslant i$ and $i-\tau_{\mathsf{C}}^i = 1-\sigma_{\mathsf{C}}\circ\theta^{i-1}$,

$$
(20) \qquad \sum_{k=\tau_{\mathsf{C}}^i+1}^{n-1} c_k r^{k-\tau_{\mathsf{C}}^i} \leqslant \rho^{1-\sigma_{\mathsf{C}}\circ\theta^{i-1}}\sum_{j=\tau_{\mathsf{C}}^i+1}^{n-1} c_j\rho^{j-i} \leqslant \rho^{-\sigma_{\mathsf{C}}\circ\theta^{i-1}}\sum_{j=\tau_{\mathsf{C}}^i+1}^{n-1} c_j\rho^{j-i} \,.
$$

In addition,

$$
(21) \qquad \sum_{k=i}^{\tau_{\mathsf{C}}^i\wedge(n-1)} c_k \leqslant \sum_{k=i}^{\tau_{\mathsf{C}}^i\wedge(n-1)} c_k\rho^{k-\tau_{\mathsf{C}}^i} = \sum_{k=i}^{\tau_{\mathsf{C}}^i\wedge(n-1)} c_k\rho^{k-i-\sigma_{\mathsf{C}}\circ\theta^{i-1}+1} \leqslant \rho^{-\sigma_{\mathsf{C}}\circ\theta^{i-1}}\sum_{k=i}^{\tau_{\mathsf{C}}^i\wedge(n-1)} c_k\rho^{k-i} \,.
$$

Plugging (19), (20) and (21) in (18) and applying Cauchy-Schwarz inequality shows

$$
\begin{aligned}
|G_i-G_{i-1}|^2 = & |G_{i,2}-G_{i,1}|^2 \leqslant 16L^2d^2\rho^{-2\sigma_{\mathsf{C}}\circ\theta^{i-1}}\left(\sum_{k=i}^{n-1} c_k\rho^{k-i}\right)^2\mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1}=i-1\}} \\
\leqslant & \frac{16L^2d^2}{1-\rho}\rho^{-2\sigma_{\mathsf{C}}\circ\theta^{i-1}}\sum_{k=i}^{n-1} c_k^2\rho^{k-i}\mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1}=i-1\}} \,.
\end{aligned}
$$

This proves (11) and thus **Fact 2.** □

Fact 3. *Assume* **H1**, **H2**, **H3**. *Let* $\mathsf{C} = \{x : V(x) < d\}$. *Then for any* $x \in \mathsf{C}$,

$$\mathbb{E}_x\left[\mathrm{e}^{f(X_0^{n-1})-\mathbb{E}_x[f(X_0^{n-1})]}\right] \leqslant \mathrm{e}^{C_3\|c\|^2} \ . \tag{22}$$

where $C_3 = 4Ld\left(-5/\log\lambda + 4MLd\right)/(1 - r \vee \lambda^{1/4})^2$.

Proof of Fact 3. For any $t \in \mathbb{R}$, $\mathrm{e}^t \leqslant 1 + t + t^2\mathrm{e}^{|t|}$. Hence, as $\mathbb{E}_x[G_{i+1} - G_i|\mathscr{F}_{\tau_{\mathsf{C}}^i}] = 0$, for any $i \geqslant 0$, we have

$$\mathbb{E}_x[\mathrm{e}^{G_{i+1}-G_i}|\mathscr{F}_{\tau_{\mathsf{C}}^i}] \leqslant 1 + \mathbb{E}_x[(G_{i+1} - G_i)^2\mathrm{e}^{|G_{i+1}-G_i|}|\mathscr{F}_{\tau_{\mathsf{C}}^i}] \ .$$

By Fact 2.,

$$\mathbb{E}_x[\mathrm{e}^{G_{i+1}-G_i}|\mathscr{F}_{\tau_{\mathsf{C}}^i}] \leqslant 1 + C_2\sum_{k=i+1}^{n-1} c_k^2\rho^{k-i-1}\mathbb{1}_{\{\tau_{\mathsf{C}}^i=i\}}\mathbb{E}_x[\rho^{-2\sigma_{\mathsf{C}}\circ\theta^i}\mathrm{e}^{C_1\|c\|_\infty\sigma_{\mathsf{C}}\circ\theta^i}|\mathscr{F}_{\tau_{\mathsf{C}}^i}] \ .$$

Now by Markov's property,

$$\begin{aligned}\mathbb{1}_{\{\tau_{\mathsf{C}}^i=i\}}\mathbb{E}_x[\rho^{-2\sigma_{\mathsf{C}}\circ\theta^i}\mathrm{e}^{C_1\|c\|_\infty\sigma_{\mathsf{C}}\circ\theta^i}|\mathscr{F}_{\tau_{\mathsf{C}}^i}] &= \mathbb{1}_{\{\tau_{\mathsf{C}}^i=i\}}\mathbb{E}_x[\rho^{-2\sigma_{\mathsf{C}}\circ\theta^i}\mathrm{e}^{C_1\|c\|_\infty\sigma_{\mathsf{C}}\circ\theta^i}|\mathscr{F}_i] \\ &= \mathbb{1}_{\{\tau_{\mathsf{C}}^i=i\}}\mathbb{E}_{X_i}[\rho^{-2\sigma_{\mathsf{C}}}\mathrm{e}^{C_1\|c\|_\infty\sigma_{\mathsf{C}}}] \ .\end{aligned}$$

Hence,

$$\mathbb{E}_x[\mathrm{e}^{G_{i+1}-G_i}|\mathscr{F}_{\tau_{\mathsf{C}}^i}] = 1 + C_2\sum_{k=i+1}^{n-1} c_k^2\rho^{k-i-1}\mathbb{1}_{\{\tau_{\mathsf{C}}^i=i\}}\mathbb{E}_{X_i}[\rho^{-2\sigma_{\mathsf{C}}}\mathrm{e}^{C_1\|c\|_\infty\sigma_{\mathsf{C}}}] \ .$$

Let $\rho = r \vee u^{-1/4}$, $\varepsilon = \log u/(2C_1)$ and assume first that $\|c\|_\infty \leqslant \varepsilon$. By **H**2,

$$\mathbb{1}_{\{\tau_{\mathsf{C}}^i=i\}}\mathbb{E}_{X_i}[\rho^{-2\sigma_{\mathsf{C}}}\mathrm{e}^{C_1\|c\|_\infty\sigma_{\mathsf{C}}}] \leqslant \mathbb{1}_{\{\tau_{\mathsf{C}}^i=i\}}\sup_{x\in\mathsf{C}}\mathbb{E}_x[\rho^{-2\sigma_{\mathsf{C}}}\mathrm{e}^{C_1\|c\|_\infty\sigma_{\mathsf{C}}}] \leqslant \sup_{x\in\mathsf{C}}\mathbb{E}_x[u^{\sigma_{\mathsf{C}}}] \leqslant M \ .$$

Hence,

$$\mathbb{E}_x[\mathrm{e}^{G_{i+1}-G_i}|\mathscr{F}_{\tau_{\mathsf{C}}^i}] \leqslant 1 + C_2M\sum_{k=i+1}^{n-1} c_k^2\rho^{k-i-1} \leqslant \mathrm{e}^{C_2M\sum_{k=i+1}^{n-1}c_k^2\rho^{k-i-1}} \ .$$

By recurrence, it follows that

$$\begin{aligned}\mathbb{E}_x\left[\mathrm{e}^{f(X_0^{n-1})-\mathbb{E}_x[f(X_0^{n-1})]}\right] &\leqslant \mathrm{e}^{C_2M\sum_{i=0}^{n-2}\sum_{k=i+1}^{n-1}c_k^2\rho^{k-i-1}} \\ &= \mathrm{e}^{C_2M\sum_{k=1}^{n-1}c_k^2\sum_{i=0}^{k-1}\rho^{k-i-1}} \leqslant \mathrm{e}^{\frac{C_2M}{1-\rho}\|c\|^2} \ .\end{aligned}$$

Fix $\widetilde{x}$ in X and let $\widetilde{f} : \mathsf{X}^n \to \mathbb{R}$ be defined, for any $x_{0:n-1}$ in X^n, by

$$\widetilde{f}(X_0^{n-1}) = f(x_0\mathbb{1}_{\{c_0\leq\varepsilon\}} + \widetilde{x}\mathbb{1}_{\{c_0>\varepsilon\}}, \ldots, x_{n-1}\mathbb{1}_{\{c_{n-1}\leq\varepsilon\}} + \widetilde{x}\mathbb{1}_{\{c_{n-1}>\varepsilon\}}) \ .$$

As f belongs to $\mathbb{BD}\left(\mathsf{X}^n, c\right)$, $\widetilde{f}$ belongs to $\mathbb{BD}\left(\mathsf{X}^n, \widetilde{c}\right)$, where

$$\widetilde{c} = \left(c_0\mathbb{1}_{\{c_0\leq\varepsilon\}}, \ldots, c_{n-1}\mathbb{1}_{\{c_{n-1}\leq\varepsilon\}}\right) \ .$$

Since $\|\widetilde{c}\|_\infty < \varepsilon$ and $\|\widetilde{c}\| \leqslant \|c\|$, $\widetilde{f}$ satisfies

$$\mathbb{E}_x\left[\mathrm{e}^{\widetilde{f}(X_0^{n-1})-\mathbb{E}_x[\widetilde{f}(X_0^{n-1})]}\right] \leqslant \mathrm{e}^{\frac{MC_2}{1-\rho}\|\widetilde{c}\|^2} \leqslant \mathrm{e}^{\frac{MC_2}{1-\rho}\|c\|^2}\ . \tag{23}$$

Furthermore, by definition of $\widetilde{f}$ and since f is in $\mathbb{BD}(\mathsf{X}^n, c)$, for any $x \in \mathsf{X}^n$,

$$|f(x) - \widetilde{f}(x)| = \sum_{i=0}^{n-1} c_i \mathbb{1}_{\{c_i > \varepsilon\}} \leq \sum_{i=0}^{n-1} c_i \frac{c_i}{\varepsilon} \leq \frac{\|c\|^2}{\varepsilon}. \tag{24}$$

This implies

$$\mathbb{E}_x\left[\mathrm{e}^{f(X_0^{n-1})-\mathbb{E}_x[f(X_0^{n-1})]}\right] \leqslant \mathrm{e}^{\frac{2\|c\|^2}{\varepsilon}} \mathbb{E}_x\left[\mathrm{e}^{\widetilde{f}(X_0^{n-1})-\mathbb{E}_x[\widetilde{f}(X_0^{n-1})]}\right] \leqslant \mathrm{e}^{\left(\frac{2}{\varepsilon}+\frac{MC_2}{1-\rho}\right)\|c\|^2}\ .$$

This shows **Fact 3** since

$$\frac{2}{\varepsilon} + \frac{MC_2}{1-\rho} \leqslant \frac{4Ld}{(1-r\vee\lambda^{1/4})^2}\left(-\frac{5}{\log\lambda} + 4MLd\right)\ .$$

□

Fact 3 proves that there exists a constant $C = 2C_3$ such that, for any $c \in \mathbb{R}^n$, $f \in \mathbb{BD}(\mathsf{X}^n, c)$ and $x \in \mathsf{C}$,

$$\mathbb{E}_x\left[\mathrm{e}^{f(X_0^{n-1})-\mathbb{E}_x[f(X_0^{n-1})]}\right] \leqslant \mathrm{e}^{C\|c\|^2/2}\ . \tag{25}$$

Let $f \in \mathbb{BD}(\mathsf{X}^n, c)$ and $x \in \mathsf{C}$. For any $s > 0$, $sf \in \mathbb{BD}(\mathsf{X}^n, c)$. Hence, from (25), for any $s, t > 0$,

$$\begin{aligned}\mathbb{P}\big(f(X_0^{n-1}) - \mathbb{E}_x[f(X_0^{n-1})] > t\big) &\leqslant \mathrm{e}^{-st+\log\mathbb{E}_x\left[\mathrm{e}^{sf(X_0^{n-1})-\mathbb{E}_x[sf(X_0^{n-1})]}\right]}\\ &\leqslant \mathrm{e}^{-st+s^2C\|c\|^2/2}\ .\end{aligned}$$

Choosing $s = t/(C\|c\|^2)$ proves Theorem **??** with

$$\beta = \frac{1}{2C} = \frac{1}{4C_3} = \frac{(1-r\vee\lambda^{1/4})^2}{16Ld}\left(-\frac{5}{\log\lambda} + 4MLd\right)^{-1}\ .$$

□

4 Applications to Unadjusted Langevin Algorithm

We illustrate the applicability of the results by the example on one of the MCMC methods. Suppose that we aim at sampling from the distribution with density $\pi(x) = \frac{e^{-U(x)}}{\int_{\mathbb{R}^d} e^{-U(y)}\,dy}$. One of the popular algorithms for solving this problem is the Unadjusted Langevin Algorithm, which suggests constructing the Markov Chain

$$X_{k+1} = h_{k+1}(X_k) + \sqrt{2\gamma_{k+1}}Z_{k+1}, h_{k+1}(X_k) := X_k - \gamma_{k+1}\nabla U(X_k)$$

with some non-increasing sequence $\gamma_k \geq 0$, where $Z_k \sim \mathcal{N}(0, I)$ are i.i.d. standard normal d-dimensional random variables. Let us denote ULA kernel with parameter γ as P_γ, that is,

$$P_\gamma(x, dy) = \frac{1}{(4\pi\gamma)^{\frac{d}{2}}} \exp\left(-\frac{\|y - x + \gamma \nabla U(x)\|^2}{4\gamma}\right) dy \tag{26}$$

Let us also introduce the multiple steps kernel

$$Q^{p,n} = P_{\gamma_p} \dots P_{\gamma_n}, Q^n := Q^{1,n}$$

Standard assumptions on the sequence $\{\gamma_k\}$ are $\sum\limits_{k=0}^{\infty} \gamma_k = \infty$ and $\sum\limits_{k=0}^{\infty} \gamma_k^2 < \infty$. Thus it is impossible to analyze properties of additive functionals of $\{X_k\}$ by standard tools, since the assumptions **H2** and **H3** fails in case $\gamma_k \to 0$. Yet we might establish some alternative bounds on the speed of convergence to stationary distribution. We need to impose condition

[**C1**] Potential $U(x)$ satisfies for all $x, y \in \mathbb{R}^d$ the condition

$$\|U(x) - U(y)\| \leq L\|x - y\|$$

This condition obviously implies

$$\|h_k(X) - h_k(Y)\| \leq (1 + L\gamma_k)\|X - Y\|$$

Hence, by [5, Theorem 19], we might obtain the following minorization condition

Lemma 5. *Assume that the condition* **C1** *holds. Then for all* $x, y \in \mathbb{R}^d, n \geq 1$ *it holds*

$$\|\delta_x Q^n - \delta_y Q^n\|_{TV} \leq \mathbb{I}(x \neq y)\left(1 - 2\Phi\left(-\frac{\|x - y\|}{2\sqrt{b_n}}\right)\right)$$

where $b_n := \sum_{i=1}^{k} \frac{\gamma_i}{\prod_{j=1}^{i}(1 + L\gamma_j)^2}$

Now we aim at establishing Foster-Lyapunov drift conditions for the kernel P_{γ_n}.We will say that function $V(x) : \mathbb{R}^d \to [1, +\infty)$ satisfies a Foster-Lyapunov drift condition for the Markov kernel P_γ if. there are such $\overline{\gamma} > 0, \lambda \in [0, 1), C > 0$ such that for any $\gamma \in (0, \overline{\gamma}]$ it holds

$$P_\gamma V \leq \lambda^\gamma V + \gamma C \tag{27}$$

Given that this condition holds, we may prove drift condition for. the kernel Q^n.

Lemma 6. *Assume that* P_γ *satisfies condition 4 for any* $0 < \gamma < \overline{\gamma}$. *Then it holds*

$$Q^{n,p} V \leq \lambda^{\Gamma_{p,n}} V + C \frac{\lambda^{\Gamma_{1,p}} - \lambda^{\Gamma_{1,n}}}{\log \frac{1}{\lambda}}$$

Proof. For notation simplicity let us consider only $p = 1$. Then by induction

$$Q^n V \leq \lambda^{\Gamma_{1,n}} V + C \sum_{k=1}^{n} \gamma_k \lambda^{\Gamma_{1,k-1}} \leq \lambda^{\Gamma_{1,n}} V + C \int\limits_0^{\Gamma_{1,n}} \lambda^x \, dx \leq$$

$$\leq \lambda^{\Gamma_{1,n}} V + C \frac{1 - \lambda^{\Gamma_{1,n}}}{\log \frac{1}{\lambda}}$$

□

Let us check that ULA kernel **??** satisfies the drift condition . We will need additional assumption

[**C2**] There exists such constants $K_1 > 0$ and $m > 0$, such that for any $x \notin B(0, K_1)$ it holds

$$\langle \nabla U(x), x \rangle \geq \frac{m}{2} \|x\|^2$$

Another popular condition on the potential is strong convexity outside some euclidean ball:

[**C3**] There exist constants $K_1 > 0$ and $m > 0$ such that for any $x \notin B(0, K_1)$ and $y \in \mathbb{R}^d$, it holds $\langle D^2 U(x) y, y \rangle \geq m \|y\|^2$. Moreover, there exists $M \geq 0$ such that for any $x \in \mathbb{R}^d$, $\|D^3 U(x)\| \leq M$.

Yet it is possible to show that assumption **C3** and **C1** imply **C2** (see [2, Lemma 14]), so in the sequel we are going to work with the condition **C2**. It allows us to prove the following result

Lemma 7. *Assume that the potential $U(x)$ satisfies conditions* **C1** *and* **C2** *and without loss of generality consider $\nabla U(0) = 0$. Then the kernel* **??** *satisfies drift condition 4 for any $0 < \gamma < \overline{\gamma} = \frac{m}{4L^2}$ with drift function $V(x) = 1 + \|x\|^2$, constants $\lambda = \exp -\left(\frac{m}{2}\right)$, $C = 3K_1^2 + 2d + C_1$ for some absolute constant $C_1 > 0$.*

Proof. Consider $V(x) = 1 + \|x\|^2$, then

$$P_\gamma V(x) = \int_{\mathbb{R}^d} V(y) P_\gamma(x, dy) = \int_{\mathbb{R}^d} (1 + \|y\|^2) \frac{1}{(4\pi\gamma)^{\frac{d}{2}}} \exp\left(-\frac{\|y - x + \gamma \nabla U(x)\|^2}{4\gamma}\right) dy =$$

$$= 1 + \frac{1}{(4\pi\gamma)^{\frac{d}{2}}} \int_{\mathbb{R}^d} \|z + x - \gamma \nabla U(x)\|^2 \exp\left(-\frac{\|z\|^2}{4\gamma}\right) dz$$

Let us first consider the case $x \notin B(0, K_1)$. Note that $\|z + x - \gamma \nabla U(x)\|^2 = \|z\|^2 + 2\langle z, x - \gamma \nabla U(x)\rangle + \gamma^2 \|x - \gamma \nabla U(x)\|^2$ and the linear term vanishes after integration, moreover, for $Z \sim \mathcal{N}(0, 2\gamma I_d$, it holds $\mathbb{E}\|Z\|^2 = 2\gamma d$. It remains to notice that due to **C1** and **C2**

$$\|x - \gamma \nabla U(x)\|^2 = \|x\|^2 - 2\gamma \langle \nabla U(x), x \rangle + \gamma^2 \|\nabla U(x)\|^2 \leq$$

$$\leq (1 - \gamma m)\|x\|^2 + 2\gamma^2 L^2 \|x\|^2$$

Thus, plugging everything into expression for $P_\gamma V$ and using $\gamma < \frac{m}{4L}$, we obtain

$$P_\gamma V(x) \leq (1 - \gamma m + 2\gamma^2 L^2) V(x) + 2\gamma d + (\gamma m - 2\gamma^2 L^2) \leq \exp^{-\frac{\gamma m}{2}} V(x) + 2\gamma d$$

Now let $x \in B(0, K_1)$. Then simply using $\|x - \gamma \nabla U(x)\|^2 \leq 2(1 + L\gamma)^2 \|x\|^2$, we obtain

$$P_\gamma V(x) \leq (1 - \gamma m + 2\gamma^2 L^2) V(x) + \gamma \left((m - 2\gamma L^2)(1 + \|x\|^2) + 2d + 2(1 + L\gamma)^2 \|x\|^d\right) \leq$$

$$\leq \exp^{-\frac{\gamma m}{2}} V(x) + \gamma(3K_1^2 + 2d + C_1)$$

for some absolute constant C_1, which does not depend on γ, d, x. $\square$

Now the geometric ergodicity result readily follows from [4, Theorem 19.4.1], namely, the following lemma holds:

Lemma 8. *Let the potential $U(x), x \in \mathbb{R}^D$ satisfy conditions* **C1** *and* **C2**. *Then for $0 < \gamma < \overline{\gamma} = \frac{m}{4L^2}$, for any $x \in X$ it holds*

$$d_V(\delta_x Q_\gamma^n, \pi_\gamma) \leq C\rho^n \left(V(x) + \pi_\gamma(V)\right)$$

with $V(x) = 1 + \|x\|^2$ and constants

$$b = \gamma(3K_1^2 + 2D + C_1); \quad \overline{b} = b\exp^{-\frac{m\gamma}{2}} + d; \quad \varepsilon = 2\Phi\left(-\frac{\sqrt{d}(1+L\gamma)}{2\sqrt{\gamma}}\right);$$

$$C = \left(1 + \exp^{-\frac{m\gamma}{2}}\right)\left(1 + \frac{\overline{b}}{(1-\varepsilon)(1-\exp^{-\frac{m\gamma}{2}} - \frac{2b}{1+d})}\right);$$

$$\rho = \exp^{-\gamma\left(\frac{m}{2} - 2\frac{3K_1^2+2D+C_1}{d}\right)\frac{\log(1-\varepsilon)}{\log(1-\varepsilon)+\log(\exp^{-\frac{m\gamma}{2}}+\frac{2b}{d+1})}}$$

for some $C_1 > 0$, which does not depend on d, γ and n.

Proof. Note that the condition **C1** implies that the Markov kernel Q_γ^n satisfies $(1,\varepsilon)$-Doeblin condition with $\varepsilon = 2\Phi\left(-\frac{\sqrt{d}(1+L\gamma)}{2\sqrt{\gamma}}\right)$. Together with drift condition **C2** it allows to apply [4, Theorem 19.4.1] with appropriate constants. □

References

[1] R. Adamczak. A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.*, 13:no. 34, 1000–1034, 2008.

[2] N. Brosse, A. Durmus, S. Meyn, E. Moulines and A. Radhakrishnan. Diffusion approximations and control variates for MCMC. arXiv:1808.01665, 2018.

[3] J. Dedecker and S. Gouëzel. Subgaussian concentration inequalities for geometrically ergodic Markov chains. *Electron. Commun. Probab.*, 20:no. 64, 12, 2015.

[4] R. Douc, E. Moulines, P. Priouret, and P. Soulier. *Markov chains.* Springer, 2018.

[5] A. Durmus and E. Moulines. High-dimensional Bayesian inference via the Unadjusted Langevin Algorithm. arXiv:1605.01559, 2018

[6] S. Ghosal and A. van der Vaart. *Fundamentals of Nonparametric Bayesian Inference.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2017.

[7] S. Le Corff, M. Lerasle, and E. Vernet. A Bayesian nonparametric approach for generalized Bradley-Terry models in random environment. arXiv:1808.08104, 2018.

[8] C. McDiarmid. On the method of bounded differences. In *Surveys in combinatorics, 1989 (Norwich, 1989)*, volume 141 of *London Math. Soc. Lecture Note Ser.*, pages 148–188. Cambridge Univ. Press, Cambridge, 1989.

[9] D. Paulin. Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electronic Journal of Probability*, 20:1–32, 2015.

[10] G. O. Roberts and J. S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probab. Surv.*, 1:20–71, 2004.

[11] J. Rousseau. On the frequentist properties of Bayesian nonparametric methods. *Annual Review of Statistics and Its Application*, 3(1):211–231, 2016.

[12] E. Vernet. Posterior consistency for nonparametric hidden Markov models with finite state space. *Electronic Journal of Statistics*, 9(1):717–752, 2015.