Electron. Commun. Probab. **25** (2020), no. 15, 1–11.
https://doi.org/10.1214/20-ECP286
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

A quantitative McDiarmid's inequality for geometrically ergodic Markov chains

Antoine Havet* Matthieu Lerasle† Eric Moulines‡
Elodie Vernet§

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 [2] 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.
Submitted to ECP on July 5, 2019, final version accepted on January 7, 2020.
Supersedes arXiv:1907.02809.

1 Introduction

The purpose of this note is to establish a quantitative version of McDiarmid's inequality for geometrically ergodic Markov chains. Let $X_0, \dots, X_{n-1}$ denote independent random variables taking values in a measurable space $(\mathsf{X}, \mathscr{X})$ and $c = (c_0, \dots, c_{n-1})$ 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 $x = (x_0, \dots, x_{n-1})$ and $y = (y_0, \dots, y_{n-1}) \in \mathsf{X}^n$, we have

$$|f(x) - f(y)| \leqslant \sum_{i=0}^{n-1} c_i \mathbb{1}_{\{x_i \neq y_i\}} . \tag{1.1}$$

The bounded difference inequality, first established in [6], shows that for all $t > 0$,

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

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 $f(x_0, \dots, x_{n-1}) = \sup_{g \in \mathscr{F}} \sum_{i=0}^{n-1} g(x_i)$, for centered functions g in a class $\mathscr{F}$ is established. The concentration of general functionals (satisfying (1.1)) of geometrically ergodic Markov chains was established in [2], where it is also proved that geometric ergodicity is a necessary

*CMAP, Ecole Polytechnique. E-mail: antoine.havetmorel@gmail.com
†CREST, CNRS. E-mail: matthieu.lerasle@ensae.fr
‡CMAP, Ecole Polytechnique. E-mail: eric.moulines@polytechnique.edu
§CMAP, Ecole Polytechnique. E-mail: elodie.vernet@polytechnique.edu

assumption. However, the result in [2] 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 $t > 0$,

$$\mathbb{P}_x\big(f(X_0, \ldots, X_{n-1}) - \mathbb{E}_x[f(X_0, \ldots, X_{n-1})] > t\big) \leqslant \mathrm{e}^{-\beta t^2/\|c\|^2} , \tag{1.2}$$

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 [9, 4] for recent accounts on this theory. To extend these results to Markovian settings, the result of [2] cannot be applied directly and a quantitative version of (1.2) is required, where the dependence of β on constants characterizing the mixing of the Markov chain is needed; see for example [10, 5].

A quantitative version of McDiarmid's inequality for Markov chains was established in [7], 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 [8, 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 McDiarmid's inequality to geometrically ergodic Markov chains. Our proof is based on [2], but avoids the use of [2, 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 There exist a non-empty set $\mathsf{C} \in \mathscr{X}$ and two real numbers $u > 1$ and $M > 0$ such that
$$\sup_{x \in \mathsf{C}} \mathbb{E}_x[u^{\sigma_{\mathsf{C}}}] \leqslant M \; .$$

H3 There exist $r \in (0,1)$ and $L \geqslant 1$ such that, for any x in the set C of **H2** and any $n \geqslant 0$,
$$\mathrm{d}_{\mathrm{TV}}(\delta_x P^n, \pi) \leqslant L r^n \; ,$$
where π is the unique invariant measure granted in **H1**.

When the Markov kernel P is uniformly ergodic, then **H2** and **H3** holds with $\mathsf{C} = \mathsf{X}$. But, as seen below, the assumptions **H2** and **H3** hold under much more general assumptions: these assumptions are indeed equivalent to V-uniform geometric ergodicity. Let $V : \mathsf{X} \to [1,\infty)$ be a measurable function. The Markov kernel P is said to be V-uniformly geometrically ergodic if P admits an invariant probability measure π such that $\pi(V) < \infty$ there exist constants $1 \leq \varsigma < \infty$ and $0 < \rho < 1$ such that, for all $n \in \mathbb{N}$ and $x \in \mathsf{X}$,
$$\mathrm{d}_V(P^n(x,\cdot),\pi) \leq \varsigma \rho^n V(x) \; , \tag{2.1}$$
where
$$\mathrm{d}_V(\xi,\xi') = \frac{1}{2} \sup \left\{ \xi(f) - \xi'(f) \; : \; \sup_{(x,x') \in \mathsf{X}\times\mathsf{X}} \frac{|f(x) - f(x')|}{V(x) + V(x')} \leq 1 \right\} \; .$$

Proposition 2.1. *Assume that **H1** holds.*

(i) Let $V : \mathsf{X} \to [1,\infty)$ be a measurable function. Assume that P is V-uniformly geometrically ergodic, i.e. there exists a probability π such that $\pi P = \pi$, $\pi(V) < \infty$ and (2.1) *is satisfied. Set*
$$m = \inf\left\{k \geq 1 \; : \; \varsigma\rho^k < 1\right\}, \quad \lambda = \varsigma\rho^m, \quad d = \lambda^{-(m-1)/m}\pi(V)\{(1-\lambda^{1/m})/2\}^{-1} \tag{2.2}$$
and $C = \{V \leq d\}$, where ς and ρ are defined in (2.1)*. Then **H2** and **H3** are satisfied with*
$$u = \{(1+\lambda^{1/m})/2\}^{-1}, \quad L = \varsigma d, \quad r = \rho \tag{2.3}$$
$$M = \frac{\varsigma - 1}{1 - \varsigma^{-1/m}} d + \left\{ \frac{1-\lambda^{-1}}{1-\lambda^{-1/m}} + 2(1+\lambda^{1/m})^{-1}\lambda^{-(m-1)/m} \right\} \pi(V) \; . \tag{2.4}$$

*(ii) Conversely, assume that **H2** and **H3** hold. Then the Markov kernel P is V-geometrically ergodic with $V(x) = \mathbb{E}_x[u^{\tau_C}]$, where $\tau_C = \inf\{k \geq 0 : X_k \in C\}$ is the hitting time of the set C.*

Proof. It follows from (2.1) that for all $k \in \mathbb{N}$ and $x \in \mathsf{X}$, $P^k V(x) \leq \varsigma \rho^k V(x) + \pi(V)$. Hence $P^m V(x) \leq \lambda V(x) + \pi(V)$ where by construction $\lambda < 1$. Set
$$V_0(x) = \sum_{\ell=0}^{m-1} \lambda^{-\ell/m} P^\ell V(x) \; . \tag{2.5}$$

By construction, we get
$$\begin{aligned} PV_0(x) = \sum_{\ell=0}^{m-1} \lambda^{-\ell/m} P^{\ell+1} V(x) &\leq \sum_{\ell=1}^{m-1} \lambda^{-(\ell-1)/m} P^\ell V(x) + \lambda^{-(m-1)/m}\{\lambda V(x) + \pi(V)\} \; , \\ &\leq \lambda^{1/m} V_0(x) + \lambda^{-(m-1)/m}\pi(V) \; . \end{aligned} \tag{2.6}$$

Note that, for all $x \in \mathsf{X}$, the following bound holds

$$V_0(x) \le \left\{ \sum_{\ell=0}^{m-1} \lambda^{-\ell/m} \varsigma \rho^{\ell} \right\} V(x) + \pi(V) \sum_{\ell=0}^{m-1} \lambda^{-\ell/m}$$
$$= \varsigma \frac{1-\rho^m \lambda^{-1}}{1-\rho\lambda^{-1/m}} V(x) + \frac{1-\lambda^{-1}}{1-\lambda^{-1/m}} \pi(V) . \tag{2.7}$$

It follows from (2.6) that

$$PV_0(x) \le \left\{ \frac{1+\lambda^{1/m}}{2} \right\} V_0(x) - \left\{ \frac{1-\lambda^{1/m}}{2} \right\} V(x) + \lambda^{-(m-1)/m} \pi(V) ,$$

where we have used that $V(x) \le V_0(x)$ for all $x \in \mathsf{X}$. Using the definition of d in (2.2), we finally get that

$$PV_0(x) \le \left\{ \frac{1+\lambda^{1/m}}{2} \right\} V_0(x) + \lambda^{-(m-1)/m} \pi(V) \mathbb{1}_{\{V \le d\}} . \tag{2.8}$$

Setting $C = \{V \le d\}$ and using [3, Proposition 4.3.3], we get for all $x \in \mathsf{X}$,

$$\mathbb{E}_x[u^{\sigma_C}] \le V_0(x) + u^{-1} \lambda^{-(m-1)/m} \pi(V) .$$

where u is defined in (2.3). Noting that

$$\sup_{x \in C} V_0(x) \le \varsigma \frac{1-\rho^m \lambda^{-1}}{1-\rho\lambda^{-1/m}} d + \frac{1-\lambda^{-1}}{1-\lambda^{-1/m}} \pi(V) ,$$

we get **H2**. **H3** follows immediately from (2.1) using that $\sup_{x \in C} V(x) = d$.

Conversely, assume **H2** and set $V(x) = \mathbb{E}_x[u^{\tau_C}]$. By [3, Proposition 4.3.3], we get that

$$PV \le u^{-1} V + M \mathbb{1}_C . \tag{2.9}$$

Note that $\sup_{x \in C} \mathbb{E}_x[u^{\tau_C}] = 1$. On the other hand, for any $x, x' \in C$, under **H3**,

$$\mathrm{d_{TV}}(\delta_x P^n, \delta_{x'} P^n) \le \mathrm{d_{TV}}(\delta_x P^n, \pi) + \mathrm{d_{TV}}(\delta_{x'} P^n, \pi) \le 2Lr^n .$$

Set $\epsilon \in (0,1)$ and choose m large enough so that $2Lr^m \le 1-\epsilon$. The set C is therefore a (m,ϵ)-Doeblin set (see [3, Definition 18.2.6]. By [3, Lemma 18.2.7], since P is irreducible and aperiodic under **H1**, the set C is small. It follows from (2.9) that

$$PV + (1-u^{-1})/2V \le (1+u^{-1})/2V + M\mathbb{1}_C . \tag{2.10}$$

By [3, Proposition 14.1.2], we get that, setting $\widetilde{\lambda} = (1+u^{-1})/2$ and $\kappa = (1-u^{-1})/2$,

$$\mathbb{E}_x \left[\sum_{k=0}^{\sigma_C - 1} \widetilde{\lambda}^{-k} V(X_k) \right] \le \kappa^{-1} \left\{ \sup_C V + M \widetilde{\lambda}^{-1} \right\} \mathbb{1}_C(x) + \kappa^{-1} V(x) \mathbb{1}_{C^c}(x)$$
$$\le \left\{ \kappa^{-1} \left\{ \sup_C V + M\widetilde{\lambda}^{-1} \right\} + \kappa^{-1} \right\} \varsigma V(x) .$$

We conclude by [3, Theorem 15.2.4] that P is V-uniformly geometrically ergodic. □

We have established in Theorem 2.1 that, under **H1**, assumptions **H2** and **H3** are equivalent to V-uniformly geometric chains. Recall that if the Markov kernel P is irreducible and aperiodic, then P is V-uniformly geometrically ergodic if and only if it satisfies a Foster-Lyapunov drift condition, i.e. there exist a small set C and constants

$b < \infty$ and $\lambda \in [0,1)$ such that $PV \le \lambda V + b\mathbb{1}_C$; see [3, Theorem 15.2.4]. It is possible to relate the constants in assumptions **H2** and **H3** with the constants appearing in the drift condition. Note first that **H2** is satisfied with $u = \lambda - 1$ and $M = \sup_{x\in C} V(x) + b\lambda^{-1}$ (see [3, Proposition 4.3.3-(ii)]) On the other hand, by [3, Corollary 14.1.6], for all $d > 0$, the sets $\{V \le d\}$ are small and there exists $d_0 < \infty$ such that $\{V \le d_0\}$ is accessible (since the set $\{V < \infty\}$ is full and absorbing). Take $d \ge d_0$ satisfying $\lambda + 2b/(1+d) < 1$. Then the set $\{V \le d\}$ is an accessible (m,ϵ)-small set and by [3, Theorem 18.4.3] **H3** is satisfied with $r = \rho^{1/m}$ and $M = \rho^{-1}\beta^{-1}(1+\epsilon)\{\pi(V)+d\}$ for all $\beta \in \big(0, \epsilon(b_m + \lambda^m - 1)^{-1} \wedge 1\big)$ with

$$\rho = \gamma_1(\beta, b_m, \lambda^m, \epsilon) \vee \gamma_2(\beta, b_m, \lambda^m) < 1 \;, \tag{2.11}$$

$$b_m = b(1-\lambda^m)(1-\lambda)^{-1} \;, \tag{2.12}$$

and γ_1 and γ_2 given by

$$\gamma_1(\beta, b, \lambda, \epsilon) = 1 - \epsilon + \beta(b + \lambda - 1) \;, \tag{2.13}$$

$$\gamma_2(\beta, b, \lambda) = 1 - \beta\frac{(1-\lambda)(1+d) - 2b}{2(1-\beta) + \beta(1+d)} \;. \tag{2.14}$$

3 Main result

The main result of this paper is the following quantitative version of McDiarmid's inequality for geometrically ergodic Markov chains.

Theorem 3.1. *Assume **H1**, **H2**, **H3**. Let $n \geqslant 1$, $c \in \mathbb{R}^n$ 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

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

Before proceeding to the proof of our main Theorem, we establish the following Lemma which replaces [2, Lemma 6]. It is instrumental in the sequel.

Lemma 3.2. *For any probability measures ξ and ξ' on $(\mathsf{X}, \mathscr{X})$, any $n \geqslant 1$, any $c \in \mathbb{R}_+^n$ and any $h \in \mathbb{BD}(\mathsf{X}^n, c)$,*

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

Remark 3.3. It is possible to avoid the factor 2 in (3.2) under additional technical conditions, for example, when there exists a maximal coupling for $(\mathbb{P}_\xi, \mathbb{P}_{\xi'})$, see [3, 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

$$\begin{aligned}\bar{w}_i(x_i) &= \int \left\{\bar{h}_i(x_i^{n-1}) - \bar{h}_{i+1}(x_{i+1}^{n-1})\right\} \prod_{\ell=i+1}^{n-1} P(x_{\ell-1}, \mathrm{d}x_\ell)\;, \\ &= \int \left\{h(x^*,\dots,x^*,x_i^{n-1}) - h(x^*,\dots,x^*,x_{i+1}^{n-1})\right\} \prod_{\ell=i+1}^{n-1} P(x_{\ell-1}, \mathrm{d}x_\ell)\;. \end{aligned} \tag{3.1}$$

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 $h \in \mathbb{BD}(\mathsf{X}^n, c)$, (3.1) shows that $|\bar{w}_i|_\infty \le c_i$. Therefore,

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

Proof of Theorem 3.1. Fix $c \in \mathbb{R}^n$, $x \in \mathsf{X}$ and $f \in \mathbb{BD}(\mathsf{X}^n, c)$. Following [2], 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, $i = 0,\dots,n-1$. 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

$$f(X_0^{n-1}) - \mathbb{E}_x[f(X_0^{n-1})] = G_{n-1} - G_0 = \sum_{i=0}^{n-2} (G_{i+1} - G_i)\;. \tag{3.2}$$

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 $i \in \{1,\dots,n-1\}$,

$$G_i - G_{i-1} = (G_i - G_{i-1})\mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1} = i-1\}}\;. \tag{3.3}$$

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 $i \geqslant 0$,

$$G_i \mathbb{1}_{\{\tau_{\mathsf{C}}^i = j\}} = \begin{cases} \mathbb{E}_x\left[f(X_0^{n-1})|\mathscr{F}_j\right] & \text{if } j \leqslant n-2\;, \\ f(X_0^{n-1}) & \text{if } j \geqslant n-1\;. \end{cases} \tag{3.4}$$

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. $\square$

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 3.2. 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

$$g_i(x_0^i) = \mathbb{E}_{x_i}[f(x_0^i, X_1^{n-1-i})], \qquad g_{i,\pi}(x_0^i) = \mathbb{E}_\pi[f(x_0^i, X_1^{n-1-i})] \; . \tag{3.5}$$

Lemma 3.4. *Assume **H1**, **H2**, **H3**. For any $i \in \{0, \dots, n-1\}$ and (x_0^{i-1}, x_i) in $\mathsf{X}^i \times \mathsf{C}$,*

$$|g_i(x_0^i) - g_{i,\pi}(x_0^i)| \leqslant 2L \sum_{j=i+1}^{n-1} c_j r^{j-i} \; . \tag{3.6}$$

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 3.2 to the function $h = \widetilde{f}_i$ yields

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

Inequality (3.6) follows from **H3**. □

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

$$|G_i - G_{i-1}| \leqslant C_1 \|c\|_\infty \mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1} = i-1\}} \sigma_{\mathsf{C}} \circ \theta^{i-1} \; , \tag{3.7}$$

$$|G_i - G_{i-1}|^2 \leqslant C_2 \mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1} = i-1\}} \frac{1}{\rho^{2\sigma_{\mathsf{C}} \circ \theta^{i-1}}} \sum_{k=i}^{n-1} c_k^2 \rho^{k-i} \; . \tag{3.8}$$

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

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

$$\begin{aligned} G_{i,1} &= \mathbb{E}_x[f(X_0^{n-1}) | \mathscr{F}_{\tau_{\mathsf{C}}^{i-1}}] \mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1} = i-1\}} = \mathbb{E}_x[f(X_0^{n-1}) | \mathscr{F}_{i-1}] \mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1} = i-1\}} \\ &= g_{i-1}(X_0^{i-1}) \mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1} = i-1\}} = g_{i-1,\pi}(X_0^{i-1}) \mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1} = i-1\}} + R_{i,1} \; . \end{aligned} \tag{3.9}$$

Moreover, as $\tau_{\mathsf{C}}^i \geqslant i$, by (3.4),

$$\begin{aligned} G_{i,2} &= \sum_{j \geqslant i} \mathbb{E}_x[f(X_0^{n-1}) | \mathscr{F}_{\tau_{\mathsf{C}}^i}] \mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1} = i-1\}} \mathbb{1}_{\{\tau_{\mathsf{C}}^i = j\}} \\ &= \sum_{j=i}^{n-2} g_j(X_0^j) \mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1} = i-1, \tau_{\mathsf{C}}^i = j\}} + f(X_0^{n-1}) \mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1} = i-1, \tau_{\mathsf{C}}^i \geqslant n-1\}} \; . \end{aligned} \tag{3.10}$$

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 (3.9) and (3.10),

$$|G_{i,2} - G_{i,1}| = |R_{i,2} - R_{i,1} + \sum_{j=i}^{n-2}(g_{j,\pi}(X_0^j) - g_{i-1,\pi}(X_0^{i-1}))\mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1}=i-1,\tau_{\mathsf{C}}^i=j\}} \\ + (f(X_0^{n-1}) - g_{i-1,\pi}(X_0^{i-1}))\mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1}=i-1,\tau_{\mathsf{C}}^i\geqslant n-1\}}| \; . \tag{3.11}$$

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,

$$\begin{aligned} \sum_{j=i}^{n-2} |(g_{j,\pi}(X_0^j) - g_{i-1,\pi}(X_0^{i-1}))|\mathbb{1}_{\{\tau_{\mathsf{C}}^i=j\}} &\leqslant \sum_{j=i}^{n-2}\mathbb{1}_{\{\tau_{\mathsf{C}}^i=j\}}\sum_{k=i}^{j} c_k = \mathbb{1}_{\{\tau_{\mathsf{C}}^i\leqslant n-2\}}\sum_{k=i}^{\tau_{\mathsf{C}}^i} c_k \; , \\ |f(X_0^{n-1}) - g_{i-1,\pi}(X_0^{i-1})|\mathbb{1}_{\{\tau_{\mathsf{C}}^i\geqslant n-1\}} &\leqslant \mathbb{1}_{\{\tau_{\mathsf{C}}^i\geqslant n-1\}}\sum_{k=i}^{n-1} c_k \; . \end{aligned} \tag{3.12}$$

To bound $|R_{i,1}|$ and $|R_{i,2}|$ in (3.11), we use Lemma 3.4. First, (3.6) directly yields

$$|R_{i,1}| \leqslant 2\mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1}=i-1\}} L \sum_{j=i+1}^{n-1} c_j r^{j-i} \; . \tag{3.13}$$

Moreover, as $\{\tau_{\mathsf{C}}^i = j\} \subset \{X_j \in \mathsf{C}\}$, (3.6) also yields

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

Therefore,

$$|R_{i,2}| \leqslant 2L\mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1}=i-1\}}\sum_{k=\tau_{\mathsf{C}}^i+1}^{n-1} c_k r^{k-\tau_{\mathsf{C}}^i} \; . \tag{3.14}$$

Plugging (3.12), (3.13) and (3.14) in (3.11) yields

$$|G_{i,2} - G_{i,1}| \leqslant 2L\left(\sum_{j=i+1}^{n-1} c_j r^{j-i} + \sum_{k=\tau_{\mathsf{C}}^i+1}^{n-1} c_k r^{k-\tau_{\mathsf{C}}^i} + \frac{1}{2L}\sum_{k=i}^{\tau_{\mathsf{C}}^i\wedge(n-1)} c_k\right)\mathbb{1}_{\{\tau_{\mathsf{C}}^{i-1}=i-1\}} \; . \tag{3.15}$$

Both (3.7) and (3.8) follow from (3.15) by bounding separately the 3 terms in the right-hand side of this inequality. Let us first establish (3.7). 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 (3.15) shows

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

This proves (3.7). We use slightly different controls to prove (3.8) from (3.15). As $r \leqslant \rho < 1$, $\rho^{-\sigma_{\mathsf{C}}\circ\theta^{i-1}} \geqslant 1$, and

$$\sum_{j=i+1}^{n-1} c_j r^{j-i} \leqslant \sum_{j=i}^{n-1} c_j \rho^{j-i} \leqslant \rho^{-\sigma_{\mathsf{C}}\circ\theta^{i-1}} \sum_{j=i}^{n-1} c_j \rho^{j-i} . \tag{3.16}$$

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

$$\sum_{k=\tau_{\mathsf{C}}^i+1}^{n-1} c_k r^{k-\tau_{\mathsf{C}}^i} \leqslant \rho^{1-\sigma_{\mathsf{C}}\circ\theta^{i-1}} \sum_{j=\tau_{\mathsf{C}}^i+1}^{n-1} c_j \rho^{j-i} \leqslant \rho^{-\sigma_{\mathsf{C}}\circ\theta^{i-1}} \sum_{j=\tau_{\mathsf{C}}^i+1}^{n-1} c_j \rho^{j-i} . \tag{3.17}$$

In addition,

$$\sum_{k=i}^{\tau_{\mathsf{C}}^i\wedge(n-1)} c_k \leqslant \sum_{k=i}^{\tau_{\mathsf{C}}^i\wedge(n-1)} c_k \rho^{k-\tau_{\mathsf{C}}^i} = \sum_{k=i}^{\tau_{\mathsf{C}}^i\wedge(n-1)} c_k \rho^{k-i-\sigma_{\mathsf{C}}\circ\theta^{i-1}+1} \leqslant \rho^{-\sigma_{\mathsf{C}}\circ\theta^{i-1}} \sum_{k=i}^{\tau_{\mathsf{C}}^i\wedge(n-1)} c_k \rho^{k-i} . \tag{3.18}$$

Plugging (3.16), (3.17) and (3.18) in (3.15) and applying Cauchy-Schwarz inequality shows

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

This proves (3.8) and thus **Fact 2.** □

Fact 3. Assume **H1**, **H2**, **H3**. For any $x \in \mathsf{C}$,

$$\mathbb{E}_x\left[\mathrm{e}^{f(X_0^{n-1})-\mathbb{E}_x[f(X_0^{n-1})]}\right] \leqslant \mathrm{e}^{C_3\|c\|^2} . \tag{3.19}$$

where $C_3 = 4L\left(5/\log u + 4ML\right)/(1 - r \vee u^{-1/4})^2$.

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,

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

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

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

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

$$\mathbb{E}_x\left[\mathrm{e}^{\widetilde{f}(X_0^{n-1})-\mathbb{E}_x[\widetilde{f}(X_0^{n-1})]}\right] \leqslant \mathrm{e}^{\frac{MC_2}{1-\rho}\|\widetilde{c}\|^2} \leqslant \mathrm{e}^{\frac{MC_2}{1-\rho}\|c\|^2} \ . \tag{3.20}$$

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

$$|f(x)-\widetilde{f}(x)| = \sum_{i=0}^{n-1}c_i\mathbb{1}_{\{c_i>\varepsilon\}} \leq \sum_{i=0}^{n-1}c_i\frac{c_i}{\varepsilon} \leq \frac{\|c\|^2}{\varepsilon}. \tag{3.21}$$

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

$$\frac{2}{\varepsilon} + \frac{MC_2}{1-\rho} \leqslant \frac{4L}{(1-r\vee u^{-1/4})^2}\left(\frac{5}{\log u} + 4ML\right) \ . \qquad \square$$

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 $x \in \mathsf{C}$,

$$\mathbb{E}_x\left[\mathrm{e}^{f(X_0^{n-1})-\mathbb{E}_x[f(X_0^{n-1})]}\right] \leqslant \mathrm{e}^{C\|c\|^2/2} . \tag{3.22}$$

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 (3.22), 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 3.1 with

$$\beta = \frac{1}{2C} = \frac{1}{4C_3} = \frac{(1 - r \vee u^{-1/4})^2}{16L}\left(\frac{5}{\log u} + 4ML\right)^{-1} . \qquad \square$$

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. MR-2424985

[2] J. Dedecker and S. Gouëzel. Subgaussian concentration inequalities for geometrically ergodic Markov chains. *Electron. Commun. Probab.*, 20:no. 64, 12, 2015. MR-3407208

[3] R. Douc, E. Moulines, P. Priouret, and P. Soulier. *Markov chains*. Springer, 2018. MR-3889011

[4] S. Ghosal and A. van der Vaart. *Fundamentals of Nonparametric Bayesian Inference*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2017. MR-3587782

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

[6] 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. MR-1036755

[7] D. Paulin. Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electronic Journal of Probability*, 20:1–32, 2015. MR-3383563

[8] G. O. Roberts and J. S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probab. Surv.*, 1:20–71, 2004. MR-2095565

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

[10] E. Vernet. Posterior consistency for nonparametric hidden Markov models with finite state space. *Electronic Journal of Statistics*, 9(1):717–752, 2015. MR-3331855