



HAL
open science

Subgaussian concentration inequalities for geometrically ergodic Markov chains

Jérôme Dedecker, Sébastien Gouëzel

► **To cite this version:**

Jérôme Dedecker, Sébastien Gouëzel. Subgaussian concentration inequalities for geometrically ergodic Markov chains. 2014. hal-01091212v1

HAL Id: hal-01091212

<https://hal.science/hal-01091212v1>

Preprint submitted on 4 Dec 2014 (v1), last revised 20 Jul 2015 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SUBGAUSSIAN CONCENTRATION INEQUALITIES FOR GEOMETRICALLY ERGODIC MARKOV CHAINS

JÉRÔME DEDECKER AND SÉBASTIEN GOUËZEL

ABSTRACT. We prove that an irreducible aperiodic Markov chain is geometrically ergodic if and only if any separately bounded functional of the stationary chain satisfies an appropriate subgaussian deviation inequality from its mean.

Let $K(x_0, \dots, x_{n-1})$ be a function of n variables, which is separately bounded in the following sense: there exist constants L_i such that for all $x_0, \dots, x_{n-1}, x'_i$,

$$(1) \quad |K(x_0, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_{n-1}) - K(x_0, \dots, x_{i-1}, x'_i, x_{i+1}, \dots, x_{n-1})| \leq L_i.$$

It is well known that, if the random variables X_0, X_1, \dots are i.i.d., then $K(X_0, \dots, X_{n-1})$ satisfies a subgaussian concentration inequality around its average, of the form

$$(2) \quad \mathbb{P}(|K(X_0, \dots, X_{n-1}) - \mathbb{E}K(X_0, \dots, X_{n-1})| > t) \leq 2e^{-2t/\sum L_i^2},$$

see for instance [McD89].

Such concentration inequalities have also attracted a lot of interest for dependent random variables, due to the wealth of possible applications. For instance, Markov chains with good mixing properties have been considered, as well as weakly dependent sequences.

A particular instance of function K is a sum $\sum f(x_i)$ (also referred to as additive functionals). In this case, one can hope for better estimates than (2), involving for instance the asymptotic variance instead of only L_i (Bernstein-like inequalities). However, for the case of a general functional K , estimates of the form (2) are rather natural.

Under very strong assumptions ensuring that the dependence is uniformly small (say, uniformly ergodic Markov chains, or Φ -mixing dependent sequences), subgaussian concentration inequalities are well known (see [Rio00] for the extension of (2) and [Sam00] for other concentration inequalities). However, there are very few such results under weaker assumptions (say, geometrically ergodic Markov chains, or α -mixing dependent sequences), where other type of exponential bounds are more usual (let us cite [MPR11] for α -mixing sequences and [AB13] for geometrically ergodic Markov chains; see also the references in those two papers for a quite complete picture of the literature). As an exception, let us mention the result of Adamczak, who proves in [Ada08] subgaussian concentration inequalities for geometrically ergodic Markov chains under the additional assumptions that the chain is strongly aperiodic and that the functional K is invariant under permutation of its variables.

Our goal in this note is to prove subgaussian concentration inequalities for aperiodic geometrically ergodic Markov chains, extending the above result of [Ada08]. Such a setting has a wide range of applications, in particular to MCMC (see for instance Section 3.2

Date: December 2, 2014.

in [AB13]). Our proof is mainly a reformulation in probabilistic terms of the proof given in [CG12] for dynamical systems. Similar results can also be proved for Markov chains that mix more slowly (for instance, if the return times to a small set have a polynomial tail, then polynomial concentration inequalities hold). The interested reader is referred to the articles [CG12] and [GM14] where such results are proved for dynamical systems: the proofs given there can be readily adapted to Markov chains using the techniques we describe in the current paper (the only difficulty is to prove an appropriate coupling lemma extending Lemma 5). Since the main case of interest for applications is geometrically ergodic Markov chains, and since the proof is more transparent in this case, we only give details for this situation.

In all the article, we consider an irreducible aperiodic Markov chain on a general state space \mathcal{S} , which we assume as usual to be countably generated.

Definition 1. *An irreducible aperiodic Markov chain is geometrically ergodic if the tails of the return time to some small set are exponential. More precisely, there exist a set C , an integer $m > 0$, a probability measure ν , and $\delta \in (0, 1)$, $\kappa > 1$ such that*

- For all $x \in C$, one has

$$(3) \quad P^m(x, \cdot) \geq \delta \nu.$$

- The return time τ_C to C satisfies

$$(4) \quad \sup_{x \in C} \mathbb{E}_x(\kappa^{\tau_C}) < \infty.$$

This is but one of many equivalent definitions, other definitions are for instance given in terms of a drift condition. Essentially, geometrically ergodic Markov chains are those Markov chains that mix exponentially fast, see [MT93, Chapters 15 and 16] for several equivalent characterizations (the above definition is [MT93, Theorem 15.0.2 (ii)]). Such a chain admits a unique stationary distribution π . A set C satisfying (3) is called a *small* set (there is a related notion of *petite* set, these notions coincide in irreducible aperiodic Markov chains, see [MT93, Theorem 5.5.7]).

We prove the following theorem.

Theorem 2. *Let (X_n) be an irreducible aperiodic Markov chain which is geometrically ergodic. Let π be its stationary distribution. There exists a constant M_0 with the following property. Let $n \in \mathbb{N}$. Let $K(x_0, \dots, x_{n-1})$ be a function of n variables on \mathcal{S}^n , which is separately bounded with constants L_i , as in (1). Then, for all $t > 0$,*

$$(5) \quad \mathbb{P}_\pi(|K(X_0, \dots, X_{n-1}) - \mathbb{E}_\pi K(X_0, \dots, X_{n-1})| > t) \leq 2e^{-M_0^{-1}t^2 / \sum L_i^2},$$

and for all x in the small set C ,

$$(6) \quad \mathbb{P}_x(|K(X_0, \dots, X_{n-1}) - \mathbb{E}_x K(X_0, \dots, X_{n-1})| > t) \leq 2e^{-M_0^{-1}t^2 / \sum L_i^2}.$$

As will be clear from the proof, the constant M_0 can be written explicitly in terms of simple numerical properties of the Markov chain, more precisely of its coupling time and of the return time to the small set C . We shall in fact prove (6), and show how it implies (5) (see the first step of the proof of Theorem 2).

Note that there is no strong aperiodicity assumption in our theorems (i.e., we are not requiring $m = 1$). This is due to the fact that we will not rely on the independence of successive excursions. Following the classical strategy of McDiarmid, we will rather decompose K as a sum of martingale increments, and estimate each of them. However, if we try to use the natural filtration given by the time, we have no control on what happens away from C . The only unusual idea in our argument is to use another filtration indexed by the next return to C , the rest is mainly routine.

The following remarks show that the above theorem is sharp: it is not possible to weaken the boundedness assumption (1), nor the assumption of geometric ergodicity.

Remark 3. It is often desirable to have estimates for functions which are unbounded. A typical example in geometrically ergodic Markov chains is the following. Consider an appropriate drift function, i.e., a function $V \geq 1$ which is bounded on a small set C and satisfies $PV(x) \leq \rho V(x) + A\mathbb{1}_C(x)$ for some numbers $\rho < 1$ and $A \geq 0$ (where P is the Markov operator of the chain). One thinks of V as being “large close to infinity”. A natural candidate for stronger concentration inequalities would be functions K satisfying

$$(7) \quad |K(x_0, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_{n-1}) - K(x_0, \dots, x_{i-1}, x'_i, x_{i+1}, \dots, x_{n-1})| \leq L_i f(V(x_i) \vee V(x'_i)),$$

for some positive function f going to infinity at infinity, for instance $f(t) = \log(1 + t)$. Unfortunately, subgaussian concentration inequalities do *not* hold for such functionals in geometrically ergodic Markov chains, whatever the function f , showing that Theorem 2 is rather sharp. Even more, they fail for additive functionals.

Consider for instance the chain on $\{1, 2, \dots\}$ given by $\mathbb{P}(1 \rightarrow s) = 2^{-s}$ for $s \geq 1$ and $\mathbb{P}(s \rightarrow s - 1) = 1$ for $s > 1$. The function $V(s) = 2^{s/2}$ satisfies the drift condition, for the small set $C = \{1\}$, since $PV(s) = 2^{-1/2}V(s)$ for $s > 1$ and $PV(1) = 2^{-1/2}/(1 - 2^{-1/2}) < \infty$. The stationary measure π is given by $\pi(s) = 2^{-s}$. In particular, V is integrable.

Assume by contradiction that a concentration inequality (5) holds for all functionals satisfying the bound (7), for some function f going to infinity. Let \tilde{f} be a nondecreasing function with $\tilde{f}(x) \leq \min(f(x), x)$, tending to infinity at infinity. Define a function $g(s) = \tilde{f}(V(s))$, except for $s = 0$ where $g(0)$ is chosen so that $\int g d\pi = 0$. Let $K(x_0, \dots, x_{n-1}) = \sum g(x_i)$, it satisfies (7) with $L_i = L$ constant and $\mathbb{E}_\pi K = 0$.

For any $N > 0$ and $n > 0$, the Markov chain has a probability 2^{-n-N} to start from $X_0 = n + N$, and then the next n iterates are $n + N - i \geq N$. In this case, $g(X_0) + \dots + g(X_{n-1}) \geq ng(N)$. Applying (5), we get

$$2^{-n-N} = \pi(n + N) \leq \mathbb{P}_\pi(|K - \mathbb{E}_\pi K| \geq ng(N)) \leq 2e^{-M_0^{-1}(ng(N))^2/(nL^2)} = 2e^{-M_0^{-1}L^{-2}g(N)^2n}.$$

Letting n tend to infinity, we deduce that $M_0^{-1}L^{-2}g(N)^2 \leq \log 2$. This is a contradiction if N is large enough, since g tends to infinity.

For instance, if one takes $f(t) = \sqrt{\ln(t \vee e)}$, then g satisfies the subgaussian condition $\mathbb{E}_\pi(\exp(g(X_0)^2)) < \infty$, but nevertheless the subgaussian inequality for the additive functional $g(X_0) + \dots + g(X_{n-1})$ fails.

Remark 4. One may wonder if the subgaussian concentration inequality (5) can be proved in larger classes of Markov chains. This is not the case: (5) characterizes geometrically ergodic Markov chains, as we now explain.

Consider an irreducible aperiodic Markov chain such that (5) holds for any separately bounded functional. We want to prove that it is geometrically ergodic. By [MT93, Theorem 5.2.2], there exists a small set, i.e., a set C satisfying (3), for some $m \geq 1$. If the original chain satisfies subgaussian concentration inequalities, then the chain at times which are multiples of m (called its m -skeleton) also does. Moreover, an irreducible aperiodic Markov chain is geometrically ergodic if and only if its m -skeleton is, by [MT93, Theorem 15.3.6]. It follows that it suffices to prove the characterization when $m = 1$, which we assume from now on.

The proof uses the *split chain* of Nummelin (see [Num78] and [Num84]). It is a Markov chain Y_n on $\bar{\mathcal{S}} = \mathcal{S} \times [0, 1]$, whose transition probabilities \bar{P} are as follows: if $x \notin C$, then $\bar{P}((x, t), \cdot) = P(x, \cdot) \otimes \lambda$, where λ is the uniform measure on $[0, 1]$. If $x \in C$, then if $t \in [0, \delta]$ one sets $\bar{P}((x, t), \cdot) = \nu \otimes \lambda$, and if $t \in (\delta, 1]$ then $\bar{P}((x, t), \cdot) = (1 - \delta)^{-1}(P(x, \cdot) - \delta\nu) \otimes \lambda$ (where δ and ν are as in (3)). Essentially, the corresponding chain behaves as the chain on \mathcal{S} , except when it enters C where the part of the transition kernel corresponding to $\delta\nu$ is explicitly separated from the rest.

For $x \in \mathcal{S}$, let $\mathbb{P}_{\bar{x}}$ denote the distribution of the Markov chain Y_n started from $\delta_x \otimes \lambda$. The first component of Y_n , living on \mathcal{S} , is then distributed as the original Markov chain started from x . In the same way, the chain Y_n started from $\bar{\pi} = \pi \otimes \lambda$ has a first projection which is distributed as the original Markov chain started from π . For obvious reasons, we still denote by X_n the first component of Y_n .

Let $\bar{C} = C \times [0, \delta]$. This is an atom of the chain Y_n . We will show that the return time $\tau_{\bar{C}}$ to \bar{C} has an exponential moment. Let $C' = C \times [0, 1]$, and let U_n be the second component of Y_n . Each time the chain X_n enters C , i.e., Y_n enters C' , then Y_n enters \bar{C} if and only if $U_n \leq \delta$. Denote by t_k the k -th visit to C' of the chain Y_n , and note that (t_k) is an increasing sequence of stopping times. By the strong Markov property, it follows that (U_{t_k}) is an i.i.d. sequence of random variables with common distribution λ . Let $K(X_1, \dots, X_n) = \sum_{i=1}^n \mathbf{1}_C(X_i)$ denote the number of visits of X_i to C . For any $k \leq n$, $\{K(X_1, \dots, X_n) \geq k\} = \{t_k \leq n\}$. It follows that, for any $k \leq n$,

$$\begin{aligned} \mathbb{P}_{\bar{\pi}}(\tau_{\bar{C}} > n) &\leq \mathbb{P}_{\pi}(K(X_1, \dots, X_n) < k) + \mathbb{P}_{\bar{\pi}}(t_k \leq n, \tau_{\bar{C}} > n) \\ &\leq \mathbb{P}_{\pi}(K(X_1, \dots, X_n) < k) + \mathbb{P}_{\bar{\pi}}(t_k \leq n, U_{t_1} > \delta, \dots, U_{t_k} > \delta) \\ &\leq \mathbb{P}_{\pi}(K(X_1, \dots, X_n) < k) + (1 - \delta)^k. \end{aligned}$$

Take $k = \varepsilon n$ for $\varepsilon = \pi(C)/2 < \pi(C)$. The subgaussian concentration inequality (5) applied to K gives, for some $c > 0$, the inequality $\mathbb{P}_{\pi}(K(X_1, \dots, X_n) \leq \varepsilon n) \leq 2e^{-cn}$. We deduce that $\tau_{\bar{C}}$ has an exponential moment, as desired, first for $\bar{\pi}$, then for its restriction to \bar{C} since $\bar{\pi}(\bar{C}) > 0$, and then for any point in \bar{C} since it is an atom (i.e., all starting points in \bar{C} give rise to a chain with the same distribution after time 1). Hence, for some $\kappa > 1$,

$$\sup_{y \in \bar{C}} \mathbb{E}_y(\kappa^{\tau_{\bar{C}}}) < \infty.$$

By definition, this shows that the extended chain Y_n is geometrically ergodic. It is then easy to deduce that X_n also is. For instance, [MT93, Theorem 15.0.2] shows the existence

of a measurable function \bar{V} on $\bar{\mathcal{S}}$ and $\rho < 1$ such that $\bar{V}(y) := \sup_{n \geq 1} \rho^{-n} \|\bar{P}^n(y, \cdot) - \bar{\pi}\|$ is everywhere finite (where $\|\cdot\|$ is the total variation norm). For $x \notin C$, the function \bar{V} is constant on $\{x\} \times [0, 1]$ since the chains Y_n starting from (x, t) or (x, t') have the same distribution after time 1. In the same way, for $x \in C$, the function \bar{V} is constant on $\{x\} \times [0, \delta]$ and on $\{x\} \times (\delta, 1]$. In particular, \bar{V} is bounded, hence integrable, on every fiber $\{x\} \times [0, 1]$. Letting $V(x) = \int \bar{V}(x, t) dt$, we get $\|(\delta_x \otimes \lambda)\bar{P}^n - \bar{\pi}\| \leq V(x)\rho^n$. Since the first marginal of the chain Y_n started from $\delta_x \otimes \lambda$ is X_n started from x , this yields $\|P^n(x, \cdot) - \pi\| \leq V(x)\rho^n$ on \mathcal{S} . This is one of the characterizations of geometric ergodicity for X_n , see [MT93, Theorem 15.0.2].

For the proof of Theorem 2, we will use the following coupling lemma. It says that the chains starting from any point in C or from the stationary distribution can be coupled in such a way that the coupling time has an exponential moment.

Lemma 5. *Consider an irreducible aperiodic geometrically ergodic Markov chain as in Definition 1. There exist constants $M_1 > 0$ and $\kappa > 1$ with the following property. Fix $x \in C$. Consider the Markov chain X_n started from x , and another copy X'_n started from the stationary distribution π . There exists a coupling between X_n and X'_n , with coupling time τ (i.e., $X_n = X'_n$ for $n \geq \tau$) such that*

$$\mathbb{E}(\kappa^\tau) \leq M_1.$$

While this lemma has a very classical flavor, we have not been able to locate a precise reference in the literature. We stress that the constants κ and M_1 are uniform, i.e., they do not depend on $x \in C$.

Proof. We will use the m -skeleton of the original Markov chain, and do our coupling at times which are multiples of m . By [MT93, Theorem 15.3.6], this m -skeleton is still geometrically ergodic, and the return times to C have a uniformly bounded exponential moment. Hence, we may without loss of generality assume that $m = 1$.

As in Remark 4, we use the split chain Y_n on $\bar{\mathcal{S}} = \mathcal{S} \times [0, 1]$. We will use the notations of this remark, in particular $\bar{C} = C \times [0, \delta]$ and $\bar{\pi} = \pi \otimes \lambda$ is the stationary distribution for Y_n . Every time the Markov chain X_n on \mathcal{S} returns to C , there is by definition a probability δ that the lifted chain Y_n enters \bar{C} . Hence, it follows from (4) that, for some $\kappa_1 > 1$,

$$(8) \quad \sup_{(x,s) \in C \times [0,1]} \mathbb{E}_{(x,s)}(\kappa_1^{\tau_{\bar{C}}}) < \infty.$$

In the same way, the entrance time to C starting from π has an exponential moment, by Theorem 2.5 (i) in [NT82]. It follows that, for some $\kappa_2 > 1$,

$$(9) \quad \mathbb{E}_{\bar{\pi}}(\kappa_2^{\tau_{\bar{C}}}) < \infty.$$

Define $T_0 = \inf\{n > 0 : Y_n \in \bar{C}\}$ and the return times

$$T_0 + \dots + T_{i+1} = \inf\{n > T_0 + \dots + T_i : Y_n \in \bar{C}\}.$$

Then T_0 is independent of $(T_i)_{i>0}$ and T_1, T_2, \dots are i.i.d. Denote by $\mathbb{P}_{\bar{\pi}}$ the probability measure on the underlying space starting from the invariant distribution $\bar{\pi}$, and by $\mathbb{P}_{\bar{x}}$ the probability measure starting from $\delta_x \otimes \lambda$ for $x \in \mathcal{S}$: the corresponding Markov chains lift

the Markov chains on \mathcal{S} starting from π and x respectively. We infer from (8) and (9) that there exists $\kappa_3 > 1$ such that

$$(10) \quad \sup_{x \in C} \mathbb{E}_x(\kappa_3^{T_0}) < \infty, \quad \mathbb{E}_\pi(\kappa_3^{T_0}) < \infty \quad \text{and} \quad \mathbb{E}(\kappa_3^{T_1}) < \infty.$$

Let now Y_n be the chain starting from $\delta_x \otimes \lambda$ with $x \in C$, and let Y'_n be a copy of the extended chain for which $Y'_0 \sim \bar{\pi}$. We apply the results in [Lin79] (see especially the computation on top of Page 66) to the renewal process $S_k = T_0 + \dots + T_k$. Since (10) is satisfied, it follows that we can couple the two chains so that the coupling time has an exponential moment, uniformly in x . Considering their first marginals, this yields the desired coupling between X_n (the Markov chain on \mathcal{S} started from x) and X'_n (the Markov chain on \mathcal{S} started from π). \square

The following lemma readily follows.

Lemma 6. *Under the assumptions of Lemma 5, let $K(x_0, \dots)$ be a function of finitely or infinitely many variables, satisfying the boundedness condition (1) for some constants L_i . Then, for all $x \in C$,*

$$|\mathbb{E}_x(K(X_0, X_1, \dots)) - \mathbb{E}_\pi(K(X_0, X_1, \dots))| \leq M_1 \sum_{i \geq 0} L_i \rho^i,$$

where $M_1 > 0$ and $\rho < 1$ do not depend on x or K .

Proof. Consider the coupling given by the previous lemma, between the Markov chain X_n started from x and the Markov chain X'_n started from the stationary distribution π . Replacing successively X_i with X'_i for $i < \tau$, we get

$$|K(X_0, X_1, \dots) - K(X'_0, X'_1, \dots)| \leq \sum_{i < \tau} L_i.$$

Taking the expectation, we obtain

$$\begin{aligned} |\mathbb{E}(K(X_0, X_1, \dots)) - \mathbb{E}(K(X'_0, X'_1, \dots))| &\leq \mathbb{E}\left(\sum_{i < \tau} L_i\right) = \sum_i L_i \mathbb{P}(\tau > i) \\ &\leq \sum_i L_i \kappa^{-i} \mathbb{E}(\kappa^\tau) \leq M_1 \sum_i L_i \kappa^{-i}. \quad \square \end{aligned}$$

We start the proof of Theorem 2. To simplify the notations, consider K as a function of infinitely many variables, with $L_i = 0$ for $i \geq n$. We start with several simple reductions in the first steps, before giving the real argument in Step 5.

First step: It suffices to prove (6), i.e., the estimate starting from a point $x_0 \in C$.

Indeed, fix some large $N > 0$, and consider the function $K_N(x_0, \dots, x_{n+N-1}) = K(x_N, \dots, x_{N+n-1})$. It satisfies $L_i(K_N) = 0$ for $i < N$ and $L_i(K_N) = L_{i-N}(K)$ for $N \leq i < n+N$. In particular, $\sum L_i(K_N)^2 = \sum L_i(K)^2$. Applying the inequality (6) to K_N , we get

$$\mathbb{P}_{x_0}(|K(X_N, \dots, X_{N+n-1}) - \mathbb{E}_{x_0}K(X_N, \dots, X_{N+n-1})| > t) \leq e^{-M_0^{-1}t^2 / \sum_{i \geq 0} L_i^2}.$$

When N tends to infinity, the Markov chain starting from x_0 equidistributes towards π . Hence, the left-hand term of the above equation converges to $\mathbb{P}_\pi(|K(X_0, \dots, X_{n-1}) - \mathbb{E}_\pi K(X_0, \dots, X_{n-1})| > t)$, giving the desired inequality.

Second step: It suffices to prove that, for $x_0 \in C$,

$$(11) \quad \mathbb{E}_{x_0}(e^{K - \mathbb{E}_{x_0} K}) \leq e^{M_2 \sum_{i \geq 0} L_i^2},$$

for some constant M_2 independent of K .

Indeed, assume that this holds. Then, for any $\lambda > 0$,

$$\mathbb{P}_{x_0}(K - \mathbb{E}_{x_0} K > t) \leq \mathbb{E}_{x_0}(e^{\lambda K - \lambda \mathbb{E}_{x_0} K - \lambda t}) \leq e^{-\lambda t} e^{\lambda^2 M_2 \sum_{i \geq 0} L_i^2},$$

by (11). Taking $\lambda = t/(2M_2 \sum L_i^2)$, we get a bound $e^{-t^2/(4M_2 \sum L_i^2)}$. Applying also the same bound to $-K$, we obtain

$$\mathbb{P}_{x_0}(|K - \mathbb{E}_{x_0} K| > t) \leq 2e^{-\frac{t^2}{4M_2 \sum L_i^2}},$$

as desired.

Third step: Fix some $\varepsilon_0 > 0$. It suffices to prove (11) assuming moreover that each L_i satisfies $L_i \leq \varepsilon_0$.

Indeed, assume that (11) is proved whenever $L_i(K) \leq \varepsilon_0$ for all i . Consider now a general function K . Take an arbitrary point $x_* \in \mathcal{S}$. Define a new function \tilde{K} by

$$\tilde{K}(x_0, \dots, x_{n-1}) = K(y_0, \dots, y_{n-1}),$$

where $y_i = x_i$ if $L_i(K) \leq \varepsilon_0$, and $y_i = x_*$ if $L_i(K) > \varepsilon_0$. This new function \tilde{K} satisfies $L_i(\tilde{K}) = L_i(K) \mathbb{1}(L_i(K) \leq \varepsilon_0) \leq \varepsilon_0$. Therefore, it satisfies (11). Moreover, $|K - \tilde{K}| \leq \sum_{L_i(K) > \varepsilon_0} L_i(K) \leq \sum L_i(K)^2 / \varepsilon_0$. Hence,

$$\mathbb{E}_{x_0}(e^{K - \mathbb{E}_{x_0} K}) \leq e^{2 \sum L_i(K)^2 / \varepsilon_0} \mathbb{E}_{x_0}(e^{\tilde{K} - \mathbb{E}_{x_0} \tilde{K}}) \leq e^{2 \sum L_i(K)^2 / \varepsilon_0} e^{M_2 \sum L_i(\tilde{K})^2}.$$

This is the desired inequality.

Let us now start the proof of (11) for a function K with $L_i \leq \varepsilon_0$ for all i . We consider the Markov chain X_0, X_1, \dots starting from a fixed point $x_0 \in C$. We define a stopping time $\tau_i = \inf\{n \geq i : X_n \in C\}$. Let \mathcal{F}_i be the σ -field corresponding to this stopping time: an event A is \mathcal{F}_i -measurable if, for all n , $A \cap \{\tau_i = n\}$ is measurable with respect to $\sigma(X_0, \dots, X_n)$. Let

$$D_i = \mathbb{E}(K | \mathcal{F}_i) - \mathbb{E}(K | \mathcal{F}_{i-1}).$$

It is \mathcal{F}_i -measurable. Moreover, $\sum_{i=0}^n D_i = K(X_0, \dots) - \mathbb{E}_{x_0}(K(X_0, \dots))$ is the quantity to understand.

Fourth step: It suffices to prove that

$$(12) \quad \mathbb{E}(e^{D_i} | \mathcal{F}_{i-1}) \leq e^{M_3 \sum_{k \geq i} L_k^2 \rho^{k-i}},$$

for some $M_3 > 0$ and some $\rho < 1$, both independent of K .

Indeed, assume that this inequality holds. Conditioning successively with respect to \mathcal{F}_n , then \mathcal{F}_{n-1} , and so on, we get

$$\mathbb{E}(e^{K - \mathbb{E} K}) = \mathbb{E}(e^{\sum D_i}) \leq e^{M_3 \sum_{i=0}^n \sum_{k \geq i} L_k^2 \rho^{k-i}} \leq e^{M_3 / (1-\rho) \cdot \sum_i L_i^2}.$$

This is the desired inequality.

Fifth step: Proof of (12).

Note first that on the set $\{\tau_{i-1} > i - 1\}$ one has $\tau_{i-1} = \tau_i$, and consequently $D_i = 0$. Hence, the following decomposition holds:

$$(13) \quad \begin{aligned} D_i &= \sum_{j=i}^{\infty} (\mathbb{E}(K \mid \mathcal{F}_j) - \mathbb{E}(K \mid \mathcal{F}_{i-1})) \mathbf{1}_{\tau_i=j, \tau_{i-1}=i-1} \\ &= \sum_{j=i}^{\infty} (g_j(X_0, \dots, X_j) - g_{i-1}(X_0, \dots, X_{i-1})) \mathbf{1}_{\tau_i=j, \tau_{i-1}=i-1}, \end{aligned}$$

where

$$g_j(x_0, \dots, x_j) = \mathbb{E}_{X_0=x_j} K(x_0, \dots, x_j, X_1, \dots, X_{n-j-1}).$$

Let now

$$g_{j,\pi}(x_0, \dots, x_j) = \mathbb{E}_{X_0 \sim \pi} K(x_0, \dots, x_j, X_1, \dots, X_{n-j-1}).$$

By Lemma 6, for any $x_j \in C$,

$$(14) \quad |g_j(x_0, \dots, x_j) - g_{j,\pi}(x_0, \dots, x_j)| \leq M_1 \sum_{k \geq j+1} L_k \rho^{k-j}.$$

From (13) and (14), we infer that

$$(15) \quad \begin{aligned} D_i &= \sum_{j=i}^{\infty} (g_{j,\pi}(X_0, \dots, X_j) - g_{i-1,\pi}(X_0, \dots, X_{i-1})) \mathbf{1}_{\tau_i=j, \tau_{i-1}=i-1} \\ &\quad + O\left(\sum_{k \geq \tau_i+1} L_k \rho^{k-\tau_i}\right) + O\left(\sum_{k \geq i} L_k \rho^{k-i}\right). \end{aligned}$$

Since π is the stationary measure, $g_{j,\pi}$ can also be written as

$$g_{j,\pi}(x_0, \dots, x_j) = \mathbb{E}_{X_0 \sim \pi} K(x_0, \dots, x_j, X_{j-i+2}, \dots, X_{n-i}).$$

It follows that

$$(16) \quad |g_{j,\pi}(x_0, \dots, x_j) - g_{i-1,\pi}(x_0, \dots, x_{i-1})| \leq \sum_{k=i}^j L_k.$$

Write $\tau = \tau_i - (i - 1)$ for the return time to C of X_{i-1} . From (15) and (16) we get

$$(17) \quad |D_i| \leq M_4 \sum_{k=i}^{i+\tau-1} L_k + M_4 \sum_{k \geq i+\tau} L_k \rho^{k-i-\tau}.$$

As all the L_k are bounded by ε_0 , we obtain

$$(18) \quad |D_i| \leq M_4 \varepsilon_0 (\tau + 1/(1 - \rho)) \leq M_5 \varepsilon_0 \tau.$$

Choose $\sigma \in [\rho, 1)$. The equation (17) also gives

$$|D_i| \leq M_4 \sum_{k \geq i} L_k \sigma^{k-i} \sigma^{-\tau}.$$

With Cauchy-Schwarz inequality, this yields

$$(19) \quad |D_i|^2 \leq M_4^2 \sigma^{-2\tau} \left(\sum_{k \geq i} L_k^2 \sigma^{k-i} \right) \left(\sum_{k \geq i} \sigma^{k-i} \right) \leq M_6 \sigma^{-2\tau} \sum_{k \geq i} L_k^2 \sigma^{k-i}.$$

We have $e^t \leq 1 + t + t^2 e^{|t|}$ for all real t . Applying this inequality to D_i , taking the conditional expectation with respect to \mathcal{F}_{i-1} and using $\mathbb{E}(D_i | \mathcal{F}_{i-1}) = 0$, this gives

$$\mathbb{E}(e^{D_i} | \mathcal{F}_{i-1}) \leq 1 + \mathbb{E}(D_i^2 e^{|D_i|} | \mathcal{F}_{i-1}).$$

Combining this estimate with (18) and (19), we get

$$\mathbb{E}(e^{D_i} | \mathcal{F}_{i-1}) \leq 1 + \mathbb{E} \left(M_6 e^{M_5 \varepsilon_0 \tau} \sigma^{-2\tau} \sum_{k \geq i} L_k^2 \sigma^{k-i} | \mathcal{F}_{i-1} \right).$$

If ε_0 is small enough and σ is close enough to 1, then $e^{M_5 \varepsilon_0 \tau} \sigma^{-2\tau}$ has uniformly bounded expectation, by definition of geometric ergodicity (see Definition 1). We obtain

$$\mathbb{E}(e^{D_i} | \mathcal{F}_{i-1}) \leq 1 + M_7 \sum_{k \geq i} L_k^2 \sigma^{k-i} \leq e^{M_7 \sum_{k \geq i} L_k^2 \sigma^{k-i}}.$$

This concludes the proof of (12), and of Theorem 2. \square

REFERENCES

- [AB13] Radosław Adamczak and Witold Bednorz, *Exponential concentration inequalities for additive functionals of markov chains*, arXiv:1201.3569v2, 2013. Cited pages 1 and 2.
- [Ada08] Radosław Adamczak, *A tail inequality for suprema of unbounded empirical processes with applications to Markov chains*, Electron. J. Probab. **13** (2008), no. 34, 1000–1034. MR2424985. Cited page 1.
- [CG12] Jean-René Chazottes and Sébastien Gouëzel, *Optimal concentration inequalities for dynamical systems*, Comm. Math. Phys. **316** (2012), 843–889. MR2993935. Cited page 2.
- [GM14] Sébastien Gouëzel and Ian Mebourne, *Moment bounds and concentration inequalities for slowly mixing dynamical systems*, Electron. J. Probab. **19** (2014), no. 93, 30. Cited page 2.
- [Lin79] Torgny Lindvall, *On coupling of discrete renewal processes*, Z. Wahrsch. Verw. Gebiete **48** (1979), no. 1, 57–70. MR533006. Cited page 6.
- [McD89] Colin McDiarmid, *On the method of bounded differences*, Surveys in combinatorics, 1989 (Norwich, 1989), London Math. Soc. Lecture Note Ser., vol. 141, Cambridge Univ. Press, Cambridge, 1989, pp. 148–188. MR1036755. Cited page 1.
- [MPR11] Florence Merlevède, Magda Peligrad, and Emmanuel Rio, *A Bernstein type inequality and moderate deviations for weakly dependent sequences*, Probab. Theory Related Fields **151** (2011), 435–474. MR2851689. Cited page 1.
- [MT93] Sean P. Meyn and Richard L. Tweedie, *Markov chains and stochastic stability*, Communications and Control Engineering Series, Springer-Verlag London Ltd., London, 1993. MR1287609. Cited pages 2, 4, and 5.
- [NT82] Esa Nummelin and Pekka Tuominen, *Geometric ergodicity of Harris recurrent markov chains with applications to renewal theory*, Stochastic Process. Appl. **12** (1982), no. 2, 187–202. MR651903. Cited page 5.
- [Num78] Esa Nummelin, *A splitting technique for Harris recurrent Markov chains*, Z. Wahrsch. Verw. Gebiete **43** (1978), no. 4, 309–318. MR0501353. Cited page 4.
- [Num84] ———, *General irreducible Markov chains and nonnegative operators*, Cambridge Tracts in Mathematics, Cambridge University Press, Cambridge, 1984. MR776608. Cited page 4.

- [Rio00] Emmanuel Rio, *Inégalités de Hoeffding pour les fonctions lipschitziennes de suites dépendantes*, C. R. Acad. Sci. Paris Sér. I Math. **330** (2000), 905–908. MR1771956. Cited page 1.
- [Sam00] Paul-Marie Samson, *Concentration of measure inequalities for Markov chains and Φ -mixing processes*, Ann. Probab. **28** (2000), 416–461. MR1756011. Cited page 1.

LABORATOIRE MAP5 UMR CNRS 8145, UNIVERSITÉ PARIS DESCARTES, SORBONNE PARIS CITÉ
E-mail address: `jerome.dedecker@parisdescartes.fr`

IRMAR, CNRS UMR 6625, UNIVERSITÉ DE RENNES 1, 35042 RENNES, FRANCE
E-mail address: `sebastien.gouezel@univ-rennes1.fr`