Concentration inequality for U-statistics of order two for uniformly ergodic Markov chains

Quentin Duchemin
LAMA, Univ Gustave Eiffel, CNRS, Marne-la-Vallée, France.
quentin.duchemin@univ-eiffel.fr
&
Yohann De Castro
Institut Camille Jordan, École Centrale de Lyon, Lyon, France
yohann.de-castro@ec-lyon.fr
&
Claire Lacour
LAMA, Univ Gustave Eiffel, CNRS, Marne-la-Vallée, France.
claire.lacour@univ-eiffel.fr

May 28, 2021

Abstract

We prove a new concentration inequality for U-statistics of order two for uniformly ergodic Markov chains. Working with bounded and π-*canonical* kernels, we show that we can recover the convergence rate of Arcones and Giné who proved a concentration result for U-statistics of independent random variables and canonical kernels. Our result allows for a dependence of the kernels $h_{i,j}$ with the indexes in the sums, which prevents the use of standard blocking tools. Our proof relies on an inductive analysis where we use martingale techniques, uniform ergodicity, Nummelin splitting and Bernstein's type inequality.

Assuming further that the Markov chain starts from its invariant distribution, we prove a Bernstein-type concentration inequality that provides sharper convergence rate for small variance terms.

1 Introduction

Concentration of measure has been intensely studied during the last decades since it finds application in large span of topics such as model selection (see [33] and [31]), statistical learning (see [9]), online learning (see [44]) or random graphs (see [11] and [10]). Important contributions in this field are those concerning U-statistics. A U-statistic of order m is a sum of the form

$$\sum_{1\leq i_1<\cdots<i_m\leq n} h_{i_1,\dots,i_m}(X_{i_1},\dots,X_{i_m}),$$

where $X_1,\dots,X_n$ are independent random variables taking values in a measurable space (E,Σ) (with E Polish) and with respective laws P_i and where $h_{i_1,\dots,i_m}$ are measurable functions of m variables $h_{i_1,\dots,i_m}: E^m \to \mathbb{R}$.

One important exponential inequality for U-statistics was provided by [4] using a Rademacher chaos approach. Their result holds for bounded and canonical (or degenerate) kernels, namely satisfying for all $i_1,\dots,i_m \in [n] := \{1,\dots,n\}$ with $i_1 < \cdots < i_m$ and for all $x_1,\dots,x_m \in E$,

$$\left\|h_{i_1,\dots,i_m}\right\|_\infty < \infty \quad \text{and} \quad \forall j\in[1,n],\ \mathbb{E}_{X_j}\Big[h_{i_1,\dots,i_m}(x_1,\dots,x_{j-1},X_j,x_{j+1},\dots,x_m)\Big] = 0\,.$$

This work was supported by grants from Région Ile-de-France.

They proved that in the degenerate case, the convergence rates for U statistics are expected to be $n^{m/2}$. Relying on precise moment inequalities of Rosenthal type, Giné, Latala and Zinn in [23] improved the result from [4] by providing the optimal four regimes of the tail, namely Gaussian, exponential, Weibull of orders $2/3$ and $1/2$. In the specific case of order 2 U-statistics, Houdré and Reynaud-Bouret in [26] recovered the result from [23] by replacing the moment estimates by martingales type inequalities, giving as a by-product explicit constants. When the kernels are unbounded, it was shown that some results can be extended provided that the random variables $h_{i_1,\dots,i_m}(X_{i_1},\dots,X_{i_m})$ have sufficiently light tails. One can mention [17, Theorem 3.26] where an exponential inequality for U-statistics with a single Banach-space valued, unbounded and canonical kernel is proved. Their approach is based on a decoupling argument originally obtained by [13] and the tail behavior of the summands is controlled by assuming that the kernel satisfies the so-called weak Cramér condition. It is now well-known that with heavy-tailed distribution for $h_{i_1,\dots,i_m}(X_{i_1},\dots,X_{i_m})$ we cannot expect to get exponential inequalities anymore. Nevertheless working with kernels that have finite p-th moment for some $p \in (1,2]$, Joly and Lugosi in [28] construct an estimator of the mean of the U-process using the median-of-means technique that performs as well as the classical U-statistic with bounded kernels.

All the above mentioned results consider that the random variables $(X_i)_{i\geq 1}$ are independent. This condition can be prohibitive for practical applications since modelization of real phenomena often involves some dependence structure. The simplest and the most widely used tool to incorporate such dependence is Markov chain. One can give the example of Reinforcement Learning (see [42]) or Biology (see [41]). Recent works provide extensions of the classical concentration results to the Markovian settings as [18, 27, 35, 1, 9]. The asymptotic behaviour of U-statistics in the Markovian setup has already been investigated by several papers. We refer to [6] where the authors proved a Strong Law of Large Numbers and a Central Limit Theorem proved for U-statistics of order 2 using the *renewal approach* based on the splitting technique. One can also mention [16] regarding large deviation principles. However, there are only few results for the non-asymptotic behaviour of tails of U-statistics in a dependent framework. The first results were provided in [7] and [25] where exponential inequalities for U-statistics of order $m \geq 2$ of time series under mixing conditions are proved. Those works were improved by [40] where a Hoeffding's type inequality for V and U statistics is provided under conditions on the time dependent process that are easier to check in practice. In Section 1.2, we compare in details our result with the one of [40].

For the first time, we provide in this paper a Bernstein-type concentration inequality for U-statistics of order 2 in a dependent framework with kernels that may depend on the indexes of the sum and that are not assumed to be symmetric or smooth. We work on a general state space with bounded kernels that are π-canonical. This latter notion was first introduced in [20] who proved a variance inequality for U-statistics of ergodic Markov chains. Our Bernstein bound holds for stationary chains but we provide a Hoeffding-type inequality without any assumption on the initial distribution of the Markov chain.

1.1 Main results

We consider a Markov chain $(X_i)_{i\geq 1}$ with transition kernel $P: E\times E \to \mathbb{R}$ taking values in a measurable space (E,Σ), and we introduce bounded functions $h_{i,j}: E^2 \to \mathbb{R}$. Our assumptions on the Markov chain $(X_i)_{i\geq 1}$ are fully-provided in Section 2 and include in particular the uniform ergodicity of the chain and an "upper-bounded" transition kernel P by some probability measure ν (see Assumption 2). The invariant distribution of the chain $(X_i)_{i\geq 1}$ will be denoted π. We further assume that kernel functions are π-canonical, namely

$$\forall i,j \in [n], \quad \forall x,y \in E, \quad \mathbb{E}_\pi h_{i,j}(X,x) = \mathbb{E}_\pi h_{i,j}(X,y) = \mathbb{E}_\pi h_{i,j}(x,X) = \mathbb{E}_\pi h_{i,j}(y,X),$$

and we denote this common expectation $\mathbb{E}_\pi[h_{i,j}]$.

Under those assumptions, Theorem 1 gives an exponential inequality for the U-statistic

$$U_{\mathrm{stat}}(n) = \sum_{1\leq i<j\leq n} \left(h_{i,j}(X_i,X_j) - \mathbb{E}\left[h_{i,j}(X_i,X_j)\right]\right).$$

The proof of Theorem 1 can be found in Section 3.1.

Theorem 1 *Let $n \geq 2$. We suppose Assumptions 1, 2 and 3 described in Section 2. There exist two constants $\beta, \kappa > 0$ such that for any $u > 0$,*

- *if Assumption 4.(i) is satisfied, it holds with probability at least $1 - \beta e^{-u}\log(n)$,*

$$U_{\text{stat}}(n) \leq \kappa \log(n)\Big(\left[\mathbf{A}\sqrt{\mathbf{n}}\right]\sqrt{u} + \left[A + B\sqrt{n}\right]u + \left[2A\sqrt{n}\right]u^{3/2} + A\left[u^2 + n\right]\Big),$$

- *if Assumption 4.(ii) is satisfied, it holds with probability at least $1 - \beta e^{-u}\log(n)$,*

$$U_{\text{stat}}(n) \leq \kappa \log(n)\Big(\mathbf{C}\sqrt{u} + \left[A + B\sqrt{n}\right]u + \left[2A\sqrt{n}\right]u^{3/2} + A\left[u^2 + n\right]\Big),$$

where

$$A := 2\max_{i,j}\|h_{i,j}\|_\infty, \quad C^2 := \sum_{j=2}^{n}\sum_{i=1}^{j-1}\mathbb{E}\left[\mathbb{E}_{X'\sim\nu}[p_{i,j}^2(X_i, X')]\right],$$

$$B^2 := \max\Bigg[\max_{0\leq k\leq t_n}\ \max_i\ \sup_x \sum_{j=i+1}^{n}\mathbb{E}_{X'\sim\nu}\left(\mathbb{E}_{X\sim P^k(X',\cdot)}\, p_{i,j}(x,X)\right)^2,$$
$$\max_{0\leq k\leq t_n}\ \max_j\ \sup_y \sum_{i=1}^{j-1}\mathbb{E}_{\tilde{X}\sim\pi}\left(\mathbb{E}_{X\sim P^k(y,\cdot)} p_{i,j}(\tilde{X},X)\right)^2\Bigg],$$

with

$$\forall i,j, \quad \forall x,y \in E, \quad p_{i,j}(x,y) = h_{i,j}(x,y) - \mathbb{E}_\pi[h_{i,j}],$$

and with the convention that $P^0(y,\cdot)$ is the Dirac measure at point $y \in E$. The constant $\kappa > 0$ only depends on constants related to the Markov chain $(X_i)_{i\geq 1}$, namely δ_M, $\|T_1\|_{\psi_1}$, $\|T_2\|_{\psi_1}$, L, m and ρ. The constant $\beta > 0$ only depends on ρ. $t_n = \lfloor q \log n \rfloor$ with $q > 2(\log(1/\rho))^{-1}$ and ν is a probability measure on (E, Σ) "upper-bounding" the transition kernel P. See Section 2 for details.

Note that the kernels $h_{i,j}$ do not need to be symmetric and that we do not consider any assumption on the initial measure of the Markov chain $(X_i)_{i\geq 1}$ if the kernels $h_{i,j}$ do not depend on i (see Assumption 4). Let us also point out that the uniform ergodicity of the Markov chain ensured by Assumption 1 can allow to bound the constant B since for all $x, y \in E$ and for all $k \geq 0$,

$$\left|\mathbb{E}_{X\sim P^k(y,\cdot)} p_{i,j}(x,X)\right| \leq \sup_z |h_{i,j}(x,z)| \times \|P^k(y,\cdot) - \pi\|_{TV},$$

where for any measure ω on (E,Σ), $\|\omega\|_{TV} := \sup_{A\in\Sigma}|\omega(A)|$ is the total variation norm of ω. Note also that in the specific case where $\nu = \pi$ (which includes the independent setting), we get that

$$C^2 = \sum_{i<j}\mathbb{E}\left\{\mathrm{Var}_{\tilde{X}\sim\pi}\left[h_{i,j}(X_i,\tilde{X})|X_i\right]\right\},$$

and using Jensen inequality that

$$B^2 \leq \max\Big[\sup_{x,i}\sum_{j=i+1}^{n}\mathrm{Var}_{\tilde{X}\sim\pi}[h_{i,j}(x,\tilde{X})]\,,\ \sup_{y,j}\sum_{i=1}^{j-1}\mathrm{Var}_{\tilde{X}\sim\pi}[h_{i,j}(\tilde{X},y)]\Big].$$

Hence, C^2 and B^2 can be understood as variance terms that would tend to be larger as ν moves away from π. By bounding coarsely the constant B in the proof of Theorem 1, we show in Section 3.2 that the following result holds.

Theorem 2 *Let $n \geq 2$. We suppose Assumptions 1, 2, 3 and 4 described in Section 2. Then there exist constants $\beta, \kappa > 0$ such that for any $u \geq 1$, it holds with probability at least $1 - \beta e^{-u}\log n$,*

$$\frac{2}{n(n-1)}U_{\text{stat}}(n) \leq \kappa \max_{i,j}\|h_{i,j}\|_\infty \log n\left\{\frac{u}{n} + \left[\frac{u}{n}\right]^2\right\},$$

where the constant $\kappa > 0$ only depends on constants related to the Markov chain $(X_i)_{i\geq 1}$, namely δ_M, $\|T_1\|_{\psi_1}$, $\|T_2\|_{\psi_1}$, L, m and ρ. The constant $\beta > 0$ only depends on ρ.

Theorem 2 shows

$$\frac{2}{n(n-1)}U_{\mathrm{stat}}(n) = \mathcal{O}_{\mathbb{P}}\left(\frac{\log(n)\log\log n}{n}\right),$$

where $\mathcal{O}_{\mathbb{P}}$ denotes stochastic boundedness. Up to a $\log(n)\log\log n$ multiplicative term, we uncover the optimal rate of Hoeffding's inequality for canonical U-statistics of order 2, see [28]. Taking a close look at the proof of Theorems 1 and 2 (and more specifically at Section 3.1.3), one can remark that the same results hold if the U-statistic is centered with the expectations $\mathbb{E}_\pi[h_{i,j}]$, namely for

$$\sum_{1\le i<j\le n}\left(h_{i,j}(X_i,X_j)-\mathbb{E}_\pi[h_{i,j}]\right).$$

It is well-known that one can expect a better convergence rate when variance terms are small with a Bernstein bound. The main limitation in Theorem 1 that prevents us from taking advantage of small variances is the term at the extreme right on the concentration inequality of Theorem 1, namely $An\log n$. Working with the additional assumption that the Markov chain $(X_i)_{i\ge1}$ is stationary – meaning that the initial distribution of the chain is the invariant distribution π – we are able to prove a Bernstein-type concentration inequality as stated with Theorem 3. Only few updates in the proof of Theorem 1 allow to get Theorem 3 and we provide all the details in Section 3.3.

Theorem 3 *We keep the notations of Theorem 1. We suppose Assumptions 1, 2 and 3 described in Section 2. We further assume that the Markov chain $(X_i)_{i\ge1}$ is stationary. Then there exist two constants $\beta,\kappa>0$ such that for any $u>0$, it holds with probability at least $1-\beta e^{-u}\log n$,*

$$U_{\mathrm{stat}}(n)\le\kappa\log(n)\Big(\left[C+A\log(n)\sqrt{n}\right]\sqrt{u}+\left[A+B\sqrt{n}\right]u+\left[2A\sqrt{n}\right]u^{3/2}+A\left[u^2+\log n\right]\Big).$$

1.2 Connections with the literature

In this section, we describe the concentration inequality obtained in [40] for U-statistics in a dependent framework and we compare it with our results. We consider an integer $n\in\mathbb{N}\backslash\{0\}$ and a geometrically α-mixing sequence $(X_i)_{i\in[n]}$ (see [40, Section 2]) with coefficient

$$\alpha(i)\le\gamma_1\exp(-\gamma_2 i),\quad \text{for all } i\ge1,$$

where γ_1,γ_2 are two positive absolute constants. We consider a kernel $h:\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}$ degenerate, symmetric, continuous, integrable and satisfying for some $q\ge1$, $\int_{\mathbb{R}^{2d}}|\mathscr{F}h(u)|\|u\|_2^q du<\infty$, where $\mathscr{F}h$ denotes the Fourier-transform of h. Then Eq.(2.4) from [40] states that there exists a constant $C>0$ such that for any $u>0$, it holds with probability at least $1-6e^{-u}$

$$\frac{2}{n(n-1)}U_{\mathrm{stat}}(n)\le 4C\|\mathscr{F}h\|_{L^1}\left\{A_n^{1/2}\frac{u}{n}+C\log^4(n)\left[\frac{u}{n}\right]^2\right\},$$

where $A_n^{1/2}=4\left(\frac{64\gamma_1^{1/3}}{1-\exp(-\gamma_2/3)}+\frac{\log^4(n)}{n}\right)$ and $U_{\mathrm{stat}}(n)=\sum_{1\le i<j\le n}\left(h(X_i,X_j)-\mathbb{E}_\pi[h]\right)$.

[40] has the merit of working with geometrically α-mixing stationary sequences which includes in particular geometrically (and hence uniformly) ergodic Markov chains (see [29, p.6]). For the sake of simplicity, we presented the result of [40] for U-statistics of order 2, but their result holds for U-statistics of arbitrary order $m\ge2$. Nevertheless, they only consider state spaces like $\mathbb{R}^d$ with $d\ge1$ and they work with a unique kernel h (i.e. $h_{i_1,\dots,i_m}=h$ for any $i_1,\dots,i_m$) which is assumed to be symmetric continuous, integrable and that satisfies some smoothness assumption. On the contrary, we consider general state spaces and we allow different kernels $h_{i,j}$ that are not assumed to be symmetric or smooth. In a framework where both Theorem 2 and [40, Theorem 2.1] hold, noticing that for any $u\ge1$,

$$\log(n)\frac{u}{n}+\log n\left[\frac{u}{n}\right]^2\le\mathbf{log(n)}\frac{u}{n}+\log^4(n)\left[\frac{u}{n}\right]^2,$$

we deduce that our bound is asymptotically at least as good as the one from [40] (up to a possible log factor marked in bold). However, [40] only provides a Hoeffding-type concentration inequality since the kernel h appears in the constants involved in their bound only trough the L^1 norm of its Fourier-transform. Hence, they cannot benefit from smaller variance terms to get faster convergence rates contrary to our result stated in Theorem 3.

1.3 Outline

In Section 2, we introduce some notations and we present in details the assumptions under which our concentration results from Theorems 1, 2 and 3 hold. The proofs of our results are presented in Section 3. In the Appendix, we provide additional material that is not essential for the understanding of our work but that may be helpful for non-specialist readers. We include in particular a reminder of the useful definitions and properties of Markov chains on a general state space (see Section A), and the presentation of two standard concentration results for sums of function of a Markov chain (see Sections B.3 and B.4).

2 Assumptions and notations

2.1 Uniform ergodicity

Assumption 1 *The Markov chain $(X_i)_{i\geq 1}$ is ψ-irreducible for some maximal irreducibility measure ψ on Σ (see [34, Section 4.2]). Moreover, there exist $\delta_m > 0$ and some integer $m \geq 1$ such that*

$$\forall x \in E,\ \forall A \in \Sigma, \quad \delta_m \mu(A) \leq P^m(x,A).$$

for some probability measure μ.

For the reader familiar with the theory of Markov chains, Assumption 1 states that the whole space E is a small set which is equivalent to the uniform ergodicity of the Markov chain $(X_i)_{i\geq 1}$ (see [34, Theorem 16.0.2]), namely there exist constants $0 < \rho < 1$ and $L > 0$ such that

$$\|P^n(x,\cdot) - \pi\|_{TV} \leq L\rho^n, \qquad \forall n \geq 0,\ \pi\text{—a.e } x \in E,$$

where π is the unique invariant distribution of the chain $(X_i)_{i\geq 1}$. From [19, section 2.3]), we also know that the Markov chain $(X_i)_{i\geq 1}$ admits a spectral gap $1-\lambda > 0$ with $\lambda \in [0,1)$ (thanks to uniform ergodicity). We refer to Section A.1 for a reminder on the spectral gap of Markov chains.

2.2 Upper-bounded Markov kernel

Assumption 2 can be read as a reverse Doeblin's condition and allows us to achieve a change of measure in expectations in our proof to work with i.i.d. random variables with distribution ν. As a result, Assumption 2 is the cornerstone of our approach since it allows to decouple the U-statistic in the proof.

Assumption 2 *There exists $\delta_M > 0$ such that*

$$\forall x \in E,\ \forall A \in \Sigma, \quad P(x,A) \leq \delta_M\, \nu(A),$$

for some probability measure ν.

Assumption 2 has already been used in the literature (see [32, Section 4.2]) and was introduced in [14]. This condition can typically require the state space to be compact as highlighted in [32].

Let us describe another situation where Assumption 2 holds. Consider that $(E, \|\cdot\|)$ is a normed space and that for all $x \in E$, $P(x,dy)$ has density $p(x,\cdot)$ with respect to some measure η on (E,Σ). We further assume that there exists an integrable function $u : E \to \mathbb{R}_+$ such that

$$\forall x,y \in E, \quad p(x,y) \leq u(y).$$

Then considering for ν the probability measure with density $u/\|u\|_1$ with respect to η and $\delta_M = \|u\|_1$, Assumption 2 holds.

2.3 Exponential integrability of the regeneration time

We introduce some additional notations which will be useful to apply Talagrand concentration result from [39]. Note that this section is inspired from [1] and [34, Theorem 17.3.1]. We assume that Assumption 1 is satisfied and we extend the Markov chain $(X_i)_{i\geq 1}$ to a new (so called *split*) chain $(\overline{X}_n, R_n) \in E \times \{0,1\}$ (see Section A.2 for a construction of the split chain), satisfying the following properties.

- $(\overline{X}_n)_n$ is again a Markov chain with transition kernel P with the same initial distribution as $(X_n)_n$. We recall that π is the invariant distribution on the E.

- if we define $T_1 = \inf\{n > 0 : R_{nm} = 1\}$,

$$T_{i+1} = \inf\{n > 0 : R_{(T_1+\cdots+T_i+n)m} = 1\},$$

  then $T_1, T_2, \dots$ are well defined and independent. Moreover $T_2, T_3, \dots$ are i.i.d.

- if we define $S_i = T_1 + \cdots + T_i$, then the "blocks"

$$Y_0 = (\overline{X}_1, \dots, \overline{X}_{mT_1+m-1}), \quad \text{and} \quad Y_i = (\overline{X}_{m(S_i+1)}, \dots, \overline{X}_{m(S_{i+1}+1)-1}),\ i > 0,$$

  form a one-dependent sequence (i.e. for all i, $\sigma((Y_j)_{j<i})$ and $\sigma((Y_j)_{j>i})$ are independent). Moreover, the sequence $Y_1, Y_2, \dots$ is stationary and if $m = 1$ the variables $Y_0, Y_1, \dots$ are independent. In consequence, for any measurable space $(S, \mathscr{B})$ and measurable functions $f : S \to \mathbb{R}$, the variables

$$Z_i = Z_i(f) = \sum_{j=m(S_i+1)}^{m(S_{i+1}+1)-1} f(\overline{X}_j), \quad i \geq 1,$$

  constitute a one-dependent sequence (an i.i.d. sequence if $m = 1$). Additionally, if f is π-integrable (recall that π is the unique stationary measure for the chain), then

$$\mathbb{E}[Z_i] = \delta_m^{-1} m \int f \, d\pi.$$

- the distribution of T_1 depends only on π, P, δ_m, μ, whereas the law of T_2 only on P, δ_m and μ.

Remark Let us highlight that $(\overline{X}_n)_n$ is a Markov chain with transition kernel P and same initial distribution as $(X_n)_n$. Hence for our purposes of estimating the tail probabilities, we will identify $(X_n)_n$ and $(\overline{X}_n)_n$.

To derive a concentration inequality, we use the exponential integrability of the regeneration times which is ensured if the chain is uniformly ergodic as stated by Proposition 1. A proof can be found in Section B.1.

Definition 1 *For $\alpha > 0$, define the function $\psi_\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ with the formula $\psi_\alpha(x) = \exp(x\alpha) - 1$. Then for a random variable X, the α-Orlicz norm is given by*

$$\|X\|_{\psi_\alpha} = \inf\{\gamma > 0 : \mathbb{E}[\psi_\alpha(|X|/\gamma)] \leq 1\}.$$

Proposition 1 *If Assumption 1 holds, then*

$$\|T_1\|_{\psi_1} < \infty \quad \textit{and} \quad \|T_2\|_{\psi_1} < \infty, \tag{1}$$

where $\|\cdot\|_{\psi_1}$ is the 1-Orlicz norm introduced in Definition 1. We denote $\tau := \max(\|T_1\|_{\psi_1}, \|T_2\|_{\psi_1})$.

2.4 π-canonical and bounded kernels

With Assumption 3, we introduce the notion of π-*canonical* kernel which is the counterpart of the canonical property from [21].

Assumption 3 *Let us denote $\mathscr{B}(\mathbb{R})$ the Borel algebra on $\mathbb{R}$. For all $i, j \in [n]$, we assume that $h_{i,j} : (E^2, \Sigma \otimes \Sigma) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ is measurable and is π-canonical, namely*

$$\forall x, y \in E, \quad \mathbb{E}_\pi[h_{i,j}(X, x)] = \mathbb{E}_\pi[h_{i,j}(X, y)] = \mathbb{E}_\pi h_{i,j}(x, X) = \mathbb{E}_\pi h_{i,j}(y, X).$$

This common expectation will be denoted $\mathbb{E}_\pi[h_{i,j}]$.
Moreover, we assume that for all $i, j \in [n]$, $\|h_{i,j}\|_\infty < \infty$.

Remarks:

• A large span of kernels are π-canonical. This is the case of translation-invariant kernels which have been widely studied in the Machine Learning community. Another example of π-canonical kernel is a rotation invariant kernel when $E=\mathbb{S}^{d-1}:=\{x\in\mathbb{R}^d:\|x\|_2=1\}$ with π also rotation invariant (see [11] or [10]).

• The notion of π-canonical kernels is the counterpart of canonical kernels in the i.i.d. framework (see for example [26]). Note that we are not the first to introduce the notion of π-canonical kernels working with Markov chains. In [20], Fort and al. provide a variance inequality for U-statistics whose underlying sequence of random variables is an ergodic Markov Chain. Their results holds for π-canonical kernels as stated with [20, Assumption A2].

• Note that if the kernels $h_{i,j}$ are not π-canonical, the U-statistic decomposes into a linear term and a π-canonical U-statistic. This is called the *Hoeffding decomposition* (see [21, p.176]) and takes the following form

$$\begin{aligned}&\sum_{i\neq j}\left(h_{i,j}(X_i,X_j)-\mathbb{E}_{(X,Y)\sim\pi\otimes\pi}[h_{i,j}(X,Y)]\right)\\ =&\sum_{i\neq j}\widetilde{h}_{i,j}(X_i,X_j)-\mathbb{E}_{\pi}\left[\widetilde{h}_{i,j}\right]+\sum_{i\neq j}\left(\mathbb{E}_{X\sim\pi}\left[h_{i,j}(X,X_j)\right]-\mathbb{E}_{(X,Y)\sim\pi\otimes\pi}\left[h_{i,j}(X,Y)\right]\right)\\ &+\sum_{i\neq j}\left(\mathbb{E}_{X\sim\pi}\left[h_{i,j}(X_i,X)\right]-\mathbb{E}_{(X,Y)\sim\pi\otimes\pi}\left[h_{i,j}(X,Y)\right]\right),\end{aligned}$$

where for all j, the kernel $\widetilde{h}_{i,j}$ is π-canonical with

$$\forall x,y\in E,\quad \widetilde{h}_{i,j}(x,y)=h_{i,j}(x,y)-\mathbb{E}_{X\sim\pi}\left[h_{i,j}(x,X)\right]-\mathbb{E}_{X\sim\pi}\left[h_{i,j}(X,y)\right].$$

2.5 Additional technical assumption

In the case where the kernels $h_{i,j}$ depend on both i and j, we need Assumption 4.(ii) to prove Theorem 1. Assumption 4.(ii) is a mild condition on the initial distribution of the Markov chain that is used when we apply Bernstein's inequality for Markov chains from Proposition 4.

Assumption 4 *At least one of the following conditions holds.*
(i) *For all $i,j\in[n]$, $h_{i,j}\equiv h_{1,j}$, i.e. the kernel function $h_{i,j}$ does not depend on i.*
(ii) *The initial distribution of the Markov chain $(X_i)_{i\geq 1}$, denoted χ, is absolutely continuous with respect to the invariant measure π and its density, denoted by $\frac{d\chi}{d\pi}$, has finite p-moment for some $p\in(1,\infty]$, i.e*

$$\infty>\left\|\frac{d\chi}{d\pi}\right\|_{\pi,p}:=\begin{cases}\left[\int\left|\frac{d\chi}{d\pi}\right|^p d\pi\right]^{1/p} & \text{if } p<\infty,\\ \operatorname{ess\,sup}\left|\frac{d\chi}{d\pi}\right| & \text{if } p=\infty.\end{cases}$$

In the following, we will denote $q=\frac{p}{p-1}\in[1,\infty)$ (with $q=1$ if $p=+\infty$) which satisfies $\frac{1}{p}+\frac{1}{q}=1$.

2.6 Examples of Markov chains satisfying the Assumptions

2.6.1 Example 1: Finite state space.

For Markov chains with finite state space, Assumption 2 holds trivially. Hence, in such framework the result of Theorem 2 holds for any uniformly ergodic Markov chain. In particular, this is true for any aperiodic and irreducible Markov chains using [5, Lemma 7.3.(ii)].

2.6.2 Example 2: AR(1) process.

Let us consider the process $(X_n)_{n\in\mathbb{N}}$ on $\mathbb{R}^k$ defined by

$$X_0\in\mathbb{R}^k \text{ and for all } n\in\mathbb{N},\quad X_{n+1}=H(X_n)+Z_n,$$

where $(Z_n)_{n\in\mathbb{N}}$ are i.i.d random variables in $\mathbb{R}^k$ and $H:\mathbb{R}^k\to\mathbb{R}^k$ is an application. Such a process is called an auto-regressive process of order 1, noted AR(1). We show that under mild conditions, our result can be applied to AR(1) processes. Before providing a result from [15] giving conditions ensuring the uniform ergodicity of an AR(1) process, let us introduce some notations.
We consider $k,d\in\mathbb{N}^*:=\mathbb{N}\backslash\{0\}$ and we denote $\|\cdot\|_k$ (resp. $\|\cdot\|_d$) the euclidean norm on $\mathbb{R}^k$ (resp. on $\mathbb{R}^d$). We need the preliminary notations.

- λ_{Leb} is the Lebesgue measure on $\mathbb{R}^k$ and $L^2(\mathbb{R}^k,\lambda_{Leb})$ is the space of square integrable functions from $\mathbb{R}^k$ to $\mathbb{R}$.
- If v is linear map from $\mathbb{R}^k$ to $\mathbb{R}^d$, we denote $\|v\|$ its norm defined by

$$\|v\|:=\sup_{\|x\|_k=1}\|v(x)\|_d.$$

If some function $G:\mathbb{R}^k\to\mathbb{R}^d$ is differentiable on $\mathbb{R}^k$, we define

$$\|dG\|_2:=\sqrt{\int_{\mathbb{R}^k}\|dG(x)\|^2 d\lambda_{Leb}(x)}.$$

We can now state the following result from [15].

Proposition 2 *Let us consider $\mathbb{R}^k$ endowed with its Borel sigma-algebra. Suppose that the random variables Z_n are i.i.d. with a distribution equivalent to the Lebesgue measure λ_{Leb} on $\mathbb{R}^k$ and with density f_Z with respect λ_{Leb}. Assume that*

- *H is bounded and continuously differentiable with $\|H\|_\infty+\|dH\|_2<\infty$.*
- *f_Z is continuously differentiable and $\|f_Z\|_\infty+\|f_Z\|_2+\|df_Z\|_2<\infty$.*

Then, the Markov chain $(X_n)_n$ satisfying $X_{n+1}=H(X_n)+Z_n$ is uniformly ergodic.

We keep the assumptions of Proposition 2 and we consider some $b>0$ such that $\|H\|_\infty\le b$. Assuming furter that $y\mapsto\sup_{z\in[-b,b]}f_Z(y-z)$ is integrable on E with respect to λ_{Leb}, we get that Assumption 2 holds (see the remark following Assumption 2). The previous condition on f_Z is for example satisfied for Gaussian distributions. We deduce that Theorem 2 can be applied in such settings that can typically be found in nonlinear filtering problem (see [14, Section 4]).

2.6.3 Example 3: ARCH process.

Let us consider $E=\mathbb{R}$. The ARCH model is

$$X_{n+1}=H(X_n)+G(X_n)Z_{n+1},$$

where H and G are continuous functions, and $(Z_n)_n$ are i.i.d. centered normal random variables with variance $\sigma^2>0$. Assuming that $\inf_x|G(x)|\ge a>0$, we know that the Markov chain (X_n) is irreducible and aperiodic (see [3, Lemma 1]). Assuming further that $\|H\|_\infty\le b<\infty$ and that $\|G\|_\infty\le c$, we can show that Assumptions 1 and 2 hold. Let us first remark that the transition kernel P of the Markov chain $(X_n)_n$ is such that for any $x\in\mathbb{R}$, $P(x,dy)$ has density $p(x,\cdot)$ with respect to the Lebesgue measure with

$$p(x,y)=\left(2\pi\sigma^2\right)^{-1}\exp\left(-\frac{(y-H(x))^2}{2\sigma^2G(x)^2}\right).$$

We deduce that for any $x,y\in\mathbb{R}$ we have

$$p(x,y)\ge\left(2\pi\sigma^2\right)^{-1}\exp\left(-\frac{(y-H(x))^2}{2\sigma^2a^2}\right)\ge g_m(y):=\frac{1}{2\pi\sigma^2}\times\begin{cases}\exp\left(-\frac{(y-B)^2}{2\sigma^2a^2}\right) & \text{if } y<-b\\ \exp\left(-\frac{2B^2}{\sigma^2a^2}\right) & \text{if } |y|\le b\\ \exp\left(-\frac{(y+B)^2}{2\sigma^2a^2}\right) & \text{if } y>b\end{cases}.$$

With a similar approach, we get

$$p(x,y) \leq \left(2\pi\sigma^2\right)^{-1} \exp\left(-\frac{(y-H(x))^2}{2\sigma^2c^2}\right) \leq g_M(y) := \left(2\pi\sigma^2\right)^{-1} \times \begin{cases} \exp\left(-\frac{(y+b)^2}{2\sigma^2c^2}\right) & \text{if } y < -b \\ 1 & \text{if } |y| \leq b \\ \exp\left(-\frac{(y-b)^2}{2\sigma^2c^2}\right) & \text{if } y > b \end{cases} .$$

We deduce that considering $\delta_m = \|g_m\|_1, \delta_M = \|g_M\|_1$ and μ (resp. ν) with density $g_m/\|g_m\|_1$ (resp. $g_M/\|g_M\|_1$) with respect to the Lebesgue measure on $\mathbb{R}$, Assumptions 1 and 2 hold.

3 Proofs

3.1 Proof of Theorem 1

In Section 3.2, we explain succinctly how to easily obtain the proof of Theorem 2 from the one of Theorem 1.

Let us recall that Theorem 1 requires either a mild condition on the initial distribution of the Markov chain or the fact that the kernels $h_{i,j}$ do not depend on i (see Assumption 4). One only needs to consider different Bernstein concentration inequalities for sums of functions of Markov chains to go from one result to the other. In this section, we give the proof of Theorem 1 in the case where Assumption 4.(i) holds. We specify the part of the proof that should be changed to get the result when $h_{i,j}$ may depend on both i and j and when Assumption 4.(ii) holds. We make this easily identifiable using the symbol ☡.

Our proof is inspired from [21, Section 3.4.3] where a Bernstein-type inequality is shown for U-statistics of order 2 in the independence setting. Their proof relies on the *canonical* property of the kernel functions which endowed the U-statistic with a martingale structure. We want to use a similar argument and we decompose $U_{\text{stat}}(n)$ to recover the martingale property for each term (except for the last one). Considering for any $l \geq 1$ the σ-algebra $G_l = \sigma(X_1, \ldots, X_l)$, the notation $\mathbb{E}_l$ refers to the conditional expectation with respect to G_l. Then we decompose $U_{\text{stat}}(n)$ as follows,

$$\begin{aligned} U_{\text{stat}}(n) = &\sum_{k=1}^{t_n} \sum_{i<j} \left(\mathbb{E}_{j-k+1}[h_{i,j}(X_i, X_j)] - \mathbb{E}_{j-k}[h_{i,j}(X_i, X_j)]\right) \\ &+ \sum_{i<j} \left(\mathbb{E}_{j-t_n}[h_{i,j}(X_i, X_j)] - \mathbb{E}\left[h_{i,j}(X_i, X_j)\right]\right), \end{aligned} \tag{2}$$

where t_n is an integer that scales logarithmically with n and that will be specified latter. By convention, we assume here that for all $k < 1$, $\mathbb{E}_k[\cdot] := \mathbb{E}[\cdot]$. Hence the first term that we will consider is given by

$$U_n = \sum_{1 \leq i < j \leq n} h_{i,j}^{(0)}(X_i, X_{j-1}, X_j),$$

where for all $x, y, z \in E$,

$$h_{i,j}^{(0)}(x,y,z) = h_{i,j}(x,z) - \int_w h_{i,j}(x,w) P(y, dw).$$

We provide a detailed proof of a concentration result for U_n by taking advantage of its martingale structure. Reasoning by induction, we show that the $t_n - 1$ following terms involved in the decomposition (2) of $U_{\text{stat}}(n)$ can be handled using a similar approach. Since the last term of the decomposition (2) has not a martingale property, another argument is required. We deal with the last term exploiting the uniform ergodicity of the Markov chain $(X_i)_{i \geq 1}$ which is guaranteed by Assumption 1 (see [37, Theorem 8]).

3.1.1 Concentration of the first term of the decomposition of the U-statistic

Martingale structure of the U-statistic Defining $Y_j = \sum_{i=1}^{j-1} h_{i,j}^{(0)}(X_i, X_{j-1}, X_j)$, U_n can be written as $U_n = \sum_{j=2}^n Y_j$. Since

$$\mathbb{E}_{j-1}[Y_j] = \mathbb{E}[Y_j \mid X_1, \ldots, X_{j-1}] = 0,$$

we know that $(U_k)_{k\geq 2}$ is a martingale relative to the σ-algebras G_l, $l \geq 2$. This martingale can be extended to $n=0$ and $n=1$ by taking $U_0 = U_1 = 0$, $G_0 = \{\emptyset, E\}$, $G_1 = \sigma(X_1)$. We will use the martingale structure of $(U_n)_n$ through the following Lemma.

Lemma 1 *(see [21, Lemma 3.4.6])*
Let $(U_m, G_m), m \in \mathbb{N}$, be a martingale with respect to a filtration G_m such that $U_0 = U_1 = 0$. For each $m \geq 1$ and $k \geq 2$, define the angle brackets $A_m^k = A_m^k(U)$ of the martingale U by

$$A_m^k = \sum_{i=1}^{m} \mathbb{E}_{i-1}[(U_i - U_{i-1})^k]$$

(and note $A_1^k = 0$ for all k). Suppose that for $\alpha > 0$ and all $i \geq 1$, $\mathbb{E}[e^{\alpha|U_i - U_{i-1}|}] < \infty$. Then

$$\left(\varepsilon_m := e^{\alpha U_m - \sum_{k\geq 2} \alpha^k A_m^k / k!}, G_m\right), \ m \in \mathbb{N},$$

is a supermartingale. In particular, $\mathbb{E}[\varepsilon_m] \leq \mathbb{E}[\varepsilon_1] = 1$, so that, if $A_m^k \leq w_m^k$ for constants $w_m^k \geq 0$; then

$$\mathbb{E}[e^{\alpha U_m}] \leq e^{\sum_{k\geq 2} \alpha^k w_m^k / k!}.$$

We will also use the following convexity result several times.

Lemma 2 *[22, page 179] For all $\theta_1, \theta_2, \varepsilon \geq 0$, and for all integer $k \geq 1$,*

$$(\theta_1 + \theta_2)^k \leq (1+\varepsilon)^{k-1} \theta_1^k + (1+\varepsilon^{-1})^{k-1} \theta_2^k.$$

For all $k \geq 2$ and $n \geq 1$, we have :

$$\begin{aligned}
A_n^k &= \sum_{j=2}^{n} \mathbb{E}_{j-1}\left[\sum_{i=1}^{j-1} h_{i,j}^{(0)}(X_i, X_{j-1}, X_j)\right]^k \\
&\leq V_n^k := \sum_{j=2}^{n} \mathbb{E}_{j-1}\left|\sum_{i=1}^{j-1} h_{i,j}^{(0)}(X_i, X_{j-1}, X_j)\right|^k \\
&= \sum_{j=2}^{n} \mathbb{E}_{j-1}\left|\sum_{i=1}^{j-1}\left(h_{i,j}(X_i, X_j) - \mathbb{E}_{\tilde{X}\sim\pi}[h_{i,j}(X_i, \tilde{X})] + \mathbb{E}_{\tilde{X}\sim\pi}[h_{i,j}(X_i, \tilde{X})] - \mathbb{E}_{j-1}[h_{i,j}(X_i, X_j)]\right)\right|^k \\
&= \sum_{j=2}^{n} \mathbb{E}_{j-1}\left|\sum_{i=1}^{j-1}\big(p_{i,j}(X_i, X_j) + m_{i,j}(X_i, X_{j-1})\big)\right|^k,
\end{aligned}$$

where

$$p_{i,j}(x,z) = h_{i,j}(x,z) - \mathbb{E}_\pi[h_{i,j}] \quad \text{and} \quad m_{i,j}(x,y) = \mathbb{E}_\pi[h_{i,j}] - \int_z h_{i,j}(x,z) P(y, dz).$$

Using Lemma 2 with $\varepsilon = 1/2$, we deduce that

$$\begin{aligned}
V_n^k &\leq \sum_{j=2}^{n} \mathbb{E}_{j-1}\left(\left|\sum_{i=1}^{j-1} p_{i,j}(X_i, X_j)\right| + \left|\sum_{i=1}^{j-1} m_{i,j}(X_i, X_{j-1})\right|\right)^k \\
&\leq \left(\frac{3}{2}\right)^{k-1} \sum_{j=2}^{n} \mathbb{E}_{j-1}\left|\sum_{i=1}^{j-1} p_{i,j}(X_i, X_j)\right|^k + 3^{k-1} \sum_{j=2}^{n} \mathbb{E}_{j-1}\left|\sum_{i=1}^{j-1} m_{i,j}(X_i, X_{j-1})\right|^k.
\end{aligned}$$

Let us remark that

$$\sum_{j=2}^{n}\mathbb{E}_{j-1}\left|\sum_{i=1}^{j-1}m_{i,j}(X_i,X_{j-1})\right|^k=\sum_{j=2}^{n}\mathbb{E}_{j-1}\left|\sum_{i=1}^{j-1}\left(\mathbb{E}_{\tilde{X}\sim\pi}[h_{i,j}(X_i,\tilde{X})]-\mathbb{E}_{j-1}[h_{i,j}(X_i,X_j)]\right)\right|^k$$
$$=\sum_{j=2}^{n}\left|\sum_{i=1}^{j-1}\left(\mathbb{E}_{\tilde{X}\sim\pi}[h_{i,j}(X_i,\tilde{X})]-\mathbb{E}_{j-1}[h_{i,j}(X_i,X_j)]\right)\right|^k$$
$$=\sum_{j=2}^{n}\left|\mathbb{E}_{j-1}\left[\sum_{i=1}^{j-1}\left(\mathbb{E}_{\tilde{X}\sim\pi}[h_{i,j}(X_i,\tilde{X})]-h_{i,j}(X_i,X_j)\right)\right]\right|^k$$
$$\leq\sum_{j=2}^{n}\mathbb{E}_{j-1}\left|\sum_{i=1}^{j-1}\left(\mathbb{E}_{\tilde{X}\sim\pi}[h_{i,j}(X_i,\tilde{X})]-h_{i,j}(X_i,X_j)\right)\right|^k,$$

where the last inequality comes from Jensen's inequality. We obtain the following upper-bound for V_n^k,

$$V_n^k\leq 2\times 3^{k-1}\sum_{j=2}^{n}\mathbb{E}_{j-1}\left|\sum_{i=1}^{j-1}p_{i,j}(X_i,X_j)\right|^k\leq 2\times 3^{k-1}\delta_M\sum_{j=2}^{n}\mathbb{E}_{X'_j}\left|\sum_{i=1}^{j-1}p_{i,j}(X_i,X'_j)\right|^k,$$

where the random variables $(X'_j)_j$ are i.i.d. with distribution ν (see Assumption 2). $\mathbb{E}_{X'_j}$ denotes the expectation on the random variable X'_j.

Lemma 3 *(see [21, Ex.1 Section 3.4]) Let Z_j be independent random variables with respective probability laws P_j. Let $k>1$, and consider functions $f_1,\dots,f_N$ where for all $j\in[N]$, $f_j\in L^k(P_j)$. Then the duality of L^p spaces and the independence of the variables Z_j imply that*

$$\left(\sum_{j=1}^{N}\mathbb{E}\left[|f_j(Z_j)|^k\right]\right)^{1/k}=\sup_{\sum_{j=1}^{N}\mathbb{E}|\xi_j(Z_j)|^{k/(k-1)}=1}\sum_{j=1}^{N}\mathbb{E}\left[f_j(Z_j)\xi_j(Z_j)\right],$$

where the sup runs over $\xi_j\in L^{k/(k-1)}(P_j)$.

Then by the duality result of Lemma 3,

$$\left(V_n^k\right)^{1/k}\leq\left(2\delta_M\times 3^{k-1}\sum_{j=2}^{n}\mathbb{E}_{X'_j}\left|\sum_{i=1}^{j-1}p_{i,j}(X_i,X'_j)\right|^k\right)^{1/k}$$
$$\leq(2\delta_M)^{1/k}\sup_{\xi\in\mathscr{L}_k}\sum_{j=2}^{n}\sum_{i=1}^{j-1}\mathbb{E}_{X'_j}\left[p_{i,j}(X_i,X'_j)\xi_j(X'_j)\right]$$

where $\mathscr{L}_k=\left\{\xi=(\xi_2,\dots,\xi_n)\ s.t.\ \forall 2\leq j\leq n,\ \xi_j\in L^{k/(k-1)}(\nu)\ with\ \sum_{j=2}^{n}\mathbb{E}|\xi_j(X'_j)|^{k/(k-1)}=1\right\}.$

$$=(2\delta_M)^{1/k}\sup_{\xi\in\mathscr{L}_k}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\mathbb{E}_{X'_j}\left[p_{i,j}(X_i,X'_j)\xi_j(X'_j)\right]$$

Let us denote F the subset of the set $\mathscr{F}(E,\mathbb{R})$ of all measurable functions from (E,Σ) to $(\mathbb{R},\mathscr{B}(\mathbb{R}))$ that are bounded by A. We set $S:=E\times F^{n-1}$. For all $i\in[n]$, we define W_i by

$$W_i:=\left(X_i,\underbrace{0,\dots,0}_{(i-1)\text{ times}},p_{i,i+1}(X_i,\cdot),p_{i,i+2}(X_i,\cdot),\dots,p_{i,n}(X_i,\cdot)\right)\in S.$$

Hence for all $i\in[n]$, W_i is $\sigma(X_i)$-measurable. We define for any $\xi=(\xi_2,\dots,\xi_n)\in\prod_{i=2}^{n}L^{k/(k-1)}(\nu)$ the function

$$\forall w=(x,p_2,\dots,p_n)\in S,\quad f_\xi(w)=\sum_{j=2}^{n}\int p_j(y)\xi_j(y)d\nu(y).$$

Then setting $\mathscr{F} = \{f_\xi : \sum_{j=2}^n \mathbb{E}|\xi_j(X'_j)|^{k/(k-1)} = 1\}$, we have

$$(V_n^k)^{1/k} \le (2\delta_M)^{1/k} \sup_{f_\xi \in \mathscr{F}} \sum_{i=1}^{n-1} f_\xi(W_i).$$

By the separability of the L^p spaces of finite measures, $\mathscr{F}$ can be replaced by a countable subset $\mathscr{F}_0$. To upper-bound the tail probabilities of U_n, we will bound the variable V_n^k on sets of large probability using Talagrand's inequality. Then we will use Lemma 1 on these sets by means of optional stopping.

Application of Talagrand's inequality for Markov chains The proof of Lemma 4 is provided in Section B.2.

Lemma 4 *Let us denote*

$$Z = \sup_{f_\xi \in \mathscr{F}} \sum_{i=1}^{n-1} f_\xi(W_i), \quad \sigma_k^2 = \mathbb{E}\left[\sum_{i=1}^{n-1} \sup_{f_\xi \in \mathscr{F}} f_\xi(W_i)^2\right] \quad \text{and} \quad b_k = \sup_{w \in S} \sup_{f_\xi \in \mathscr{F}} |f_\xi(w)|.$$

Then it holds for any $t > 0$,

$$\mathbb{P}(Z > \mathbb{E}[Z] + t) \le \exp\left(-\frac{1}{8\|\Gamma\|^2} \min\left(\frac{t^2}{4\sigma_k^2}, \frac{t}{b_k}\right)\right),$$

where Γ is a $n \times n$ matrix defined in Section B.2 which satisfies $\|\Gamma\| \le \frac{2L}{1-\rho}$.

Using Lemma 4, we deduce that for any $t > 0$,

$$\mathbb{P}\left((V_n^k)^{1/k} \ge (2\delta_M)^{1/k}\mathbb{E}[Z] + (2\delta_M)^{1/k} t\right) \le \exp\left(-\frac{1}{8\|\Gamma\|^2} \min\left(\frac{t^2}{4\sigma_k^2}, \frac{t}{b_k}\right)\right),$$

which implies that for any $x \ge 0$,

$$\mathbb{P}\left((V_n^k)^{1/k} \ge (2\delta_M)^{1/k}\mathbb{E}[Z] + (2\delta_M)^{1/k} 2\sigma_k\sqrt{x} + (2\delta_M)^{1/k} b_k x\right) \le \exp\left(-\frac{x}{8\|\Gamma\|^2}\right).$$

Using the change of variable $x = k8\|\Gamma\|^2 u$ with $u \ge 0$ in the previous inequality leads to

$$\mathbb{P}\left(\bigcup_{k=2}^{\infty} (V_n^k)^{1/k} \ge (2\delta_M)^{1/k}\mathbb{E}[Z] + (2\delta_M)^{1/k}\sigma_k 3\|\Gamma\|\sqrt{ku} + (2\delta_M)^{1/k} k8\|\Gamma\|^2 b_k u\right) \le 1.62e^{-u},$$

because

$$1 \wedge \sum_{k=2}^{\infty} \exp(-ku) \le 1 \wedge \frac{1}{e^u(e^u-1)} = \left(e^u \wedge \frac{1}{e^u-1}\right)e^{-u} \le \frac{1+\sqrt{5}}{2}e^{-u} \le 1.62e^{-u}.$$

Bounding b_k. Using Hölder's inequality we have,

$$\begin{aligned}
b_k &= \sup_{w \in S} \sup_{f_\xi \in \mathscr{F}} |f_\xi(w)| \\
&= \sup_{(p_2,\dots,p_n) \in F^{n-1}} \sup_{\xi \in \mathscr{L}_k} \sum_{j=2}^n \mathbb{E}[p_j(X'_j)\xi_j(X'_j)] \\
&\le \sup_{(p_2,\dots,p_n) \in F^{n-1}} \sup_{\sum_{j=2}^n \mathbb{E}|\xi_j(X'_j)|^{k/(k-1)}=1} \sum_{j=2}^n \left(\mathbb{E}\left|p_j(X'_j)\right|^k\right)^{1/k} \left(\mathbb{E}\left|\xi_j(X'_j)\right|^{k/(k-1)}\right)^{(k-1)/k} \\
&\le \sup_{(p_2,\dots,p_n) \in F^{n-1}} \sup_{\sum_{j=2}^n \mathbb{E}|\xi_j(X'_j)|^{k/(k-1)}=1} \left(\sum_{j=2}^n \mathbb{E}\left|p_j(X'_j)\right|^k\right)^{1/k} \left(\sum_{j=2}^n \mathbb{E}\left|\xi_j(X'_j)\right|^{k/(k-1)}\right)^{(k-1)/k} \\
&\le \sup_{(p_2,\dots,p_n) \in F^{n-1}} \left(\sum_{j=2}^n \mathbb{E}\left|p_j(X'_j)\right|^k\right)^{1/k} \\
&\le ((nA^2)A^{k-2})^{1/k},
\end{aligned}$$

where $A := 2\max_{i,j} \|h_{i,j}\|_\infty$ which satisfies $\max_{i,j} \|p_{i,j}\|_\infty \le A$. Here, we used that F is the set of measurable functions from (E,Σ) to $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ bounded by A.

Bounding the variance.

$$\begin{aligned}
\sigma_k^2 &= \mathbb{E}\left[\sum_{i=1}^{n-1} \sup_{f_\xi \in \mathscr{F}} f_\xi(W_i)^2\right] = \sum_{i=1}^{n-1} \mathbb{E}\left[\sup_{\xi \in \mathscr{L}_k} \left(\sum_{j=i+1}^{n} \mathbb{E}_{X'_j}\left[p_{i,j}(X_i, X'_j)\xi_j(X'_j)\right]\right)^2\right] \\
&= \sum_{i=1}^{n-1} \mathbb{E}\left[\left(\sup_{\xi \in \mathscr{L}_k} \left|\sum_{j=i+1}^{n} \mathbb{E}_{X'_j}\left[p_{i,j}(X_i, X'_j)\xi_j(X'_j)\right]\right|\right)^2\right] \\
&\le n\left(B_0^2 A^{k-2}\right)^{2/k},
\end{aligned}$$

where the last inequality comes from the following (where we use twice Holder's inequality),

$$\begin{aligned}
&\sup_{\xi \in \mathscr{L}_k} \left|\sum_{j=i+1}^{n} \mathbb{E}_{X'_j}\left[p_{i,j}(X_i, X'_j)\xi_j(X'_j)\right]\right| \\
\le\ & \sup_{\xi \in \mathscr{L}_k} \sum_{j=i+1}^{n} \left(\mathbb{E}_{X'_j}\left|p_{i,j}(X_i, X'_j)\right|^k\right)^{1/k} \left(\mathbb{E}\left|\xi_j(X'_j)\right|^{k/(k-1)}\right)^{(k-1)/k} \\
\le\ & \sup_{\sum_{j=2}^{n} \mathbb{E}|\xi_j(X'_j)|^{k/(k-1)}=1} \left(\sum_{j=i+1}^{n} \mathbb{E}_{X'_j}\left|p_{i,j}(X_i, X'_j)\right|^k\right)^{1/k} \left(\sum_{j=i+1}^{n} \mathbb{E}_{X'_j}\left|\xi_j(X'_j)\right|^{k/(k-1)}\right)^{(k-1)/k} \\
\le\ & \left(\sum_{j=i+1}^{n} \mathbb{E}_{X'_j}\left|p_{i,j}(X_i, X'_j)\right|^k\right)^{1/k} \\
\le\ & \left(B_0^2 A^{k-2}\right)^{1/k},
\end{aligned}$$

$$\text{where } B_0^2 := \max\left[\max_i \left\|\sum_{j=i+1}^{n} \mathbb{E}_{X\sim\nu}\left[p_{i,j}^2(\cdot, X)\right]\right\|_\infty, \max_j \left\|\sum_{i=1}^{j-1} \mathbb{E}_{X\sim\pi}[p_{i,j}^2(X,\cdot)]\right\|_\infty\right] \le B^2. \tag{3}$$

Using Lemma 2 twice and the bounds obtained on b_k and σ_k^2 gives for $u > 0$,

$$\begin{aligned}
&\left[(2\delta_M)^{1/k}\mathbb{E}[Z] + (2\delta_M)^{1/k}\sigma_k 3\|\Gamma\|\sqrt{ku} + (2\delta_M)^{1/k} k 8\|\Gamma\|^2 b_k u\right]^k \\
\le\ & \left[(2\delta_M)^{1/k}\mathbb{E}[Z] + (2\delta_M)^{1/k} 3\|\Gamma\|(B_0^2 A^{k-2})^{1/k}\sqrt{nku} + (2\delta_M)^{1/k} 8\|\Gamma\|^2((nA^2)A^{k-2})^{1/k} ku\right]^k \\
\le\ & (1+\varepsilon)^{k-1} 2\delta_M \left(\mathbb{E}[Z]\right)^k + (1+\varepsilon^{-1})^{k-1}\Big[(2\delta_M)^{1/k} 8\|\Gamma\|^2((nA^2)A^{k-2})^{1/k} ku \\
&\quad + (2\delta_M)^{1/k} 3\|\Gamma\|(B_0^2 A^{k-2})^{1/k}\sqrt{nku}\Big]^k \\
\le\ & (1+\varepsilon)^{k-1} 2\delta_M \left(\mathbb{E}[Z]\right)^k + 2\delta_M (1+\varepsilon^{-1})^{2k-2}\left(8\|\Gamma\|^2\right)^k (nA^2)A^{k-2}(ku)^k \\
&\quad + (1+\varepsilon)^{k-1}(1+\varepsilon^{-1})^{k-1} 2\delta_M \left(3\|\Gamma\|\right)^k B_0^2 A^{k-2}(nku)^{k/2}.
\end{aligned}$$

So, setting

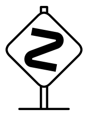
$$\begin{aligned}
w_n^k :=\ & ((1+\varepsilon)^{k-1} 2\delta_M \left(\mathbb{E}[Z]\right)^k + 2\delta_M (1+\varepsilon^{-1})^{2k-2}\left(8\|\Gamma\|^2\right)^k (nA^2)A^{k-2}(ku)^k \\
& + (1+\varepsilon)^{k-1}(1+\varepsilon^{-1})^{k-1} 2\delta_M \left(3\|\Gamma\|\right)^k B_0^2 A^{k-2}(nku)^{k/2},
\end{aligned}$$

we have

$$\mathbb{P}\left(V_n^k \leq w_n^k \quad \forall k \geq 2\right) \geq 1 - 1.62e^{-u}, \tag{4}$$

where the dependence in u of w_n^k is leaved implicit.

Bounding $(\mathbb{E}[Z])^k$.



The way we bound $(\mathbb{E}[Z])^k$ is the only part of the proof that needs to be modified to get the concentration result when Assumption 4.(i) or Assumption 4.(ii) holds. This is where we can use different Bernstein concentration inequalities. Here we present the approach when $h_{i,j} \equiv h_{1,j}$, $\forall i,j$ (i.e. when Assumption 4.(i) is satisfied). We refer to Section B.5 for the details regarding the way we bound $(\mathbb{E}Z)^k$ when Assumption 4.(ii) holds.

Using Jensen inequality and Lemma 3, we obtain

$$\begin{aligned}
(\mathbb{E}[Z])^k \leq \mathbb{E}[Z^k] &= \mathbb{E}\left[\left(\sup_{\xi \in \mathcal{L}_k} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbb{E}_{X'_j}[p_{i,j}(X_i, X'_j)\xi_j(X'_j)]\right)^k\right] \\
&= \mathbb{E}\left[\sum_{j=2}^{n} \mathbb{E}_{X'_j}\left[\left|\sum_{i=1}^{j-1} p_{i,j}(X_i, X'_j)\right|^k\right]\right] = \sum_{j=2}^{n} \mathbb{E}\left[\left|\sum_{i=1}^{j-1} p_{i,j}(X_i, X'_j)\right|^k\right],
\end{aligned}$$

where we recall that $\mathbb{E}_{X'_j}$ denotes the expectation on the random variable X'_j. One can remark that conditionally to X'_j, the quantity $\sum_{i=1}^{j-1} p_{i,j}(X_i, X'_j)$ is a sum of function of the Markov chain $(X_i)_{i\geq 1}$. Hence to control this term, we apply a Bernstein inequality for Markov chains.

Let us consider some $j \in [n]$ and some $x \in E$. We define

$$\forall l \in \{0, \dots, n\}, \quad Z_l^j(x) = \sum_{i=m(S_l+1)}^{m(S_{l+1}+1)-1} p_{i,j}(X_i, x).$$

By convention, we set $p_{i,j} \equiv 0$ for any $i \geq j$. Let us consider $N_j = \sup\{i \in \mathbb{N} \;:\; mS_{i+1} + m - 1 \geq j - 1\}$. Then using twice Lemma 2, we have

$$\begin{aligned}
\left|\sum_{i=1}^{j-1} p_{i,j}(X_i, X'_j)\right|^k &= \left|\sum_{l=0}^{N_j} Z_l^j(x) + \sum_{i=m(S_{N_j}+1)}^{j-1} p_{i,j}(X_i, x)\right|^k \\
&\leq \left(\frac{3}{2}\right)^{k-1} \left|\sum_{l=1}^{N_j} Z_l^j(x)\right|^k + 3^{k-1} \left|\sum_{i=m(S_{N_j}+1)}^{j-1} p_{i,j}(X_i, x)\right|^k \\
&\leq \left(\frac{9}{4}\right)^{k-1} \left|\sum_{l=0}^{\lfloor N_j/2 \rfloor} Z_{2l}^j(x)\right|^k + \left(\frac{9}{2}\right)^{k-1} \left|\sum_{l=0}^{\lfloor (N_j-1)/2 \rfloor} Z_{2l+1}^j(x)\right|^k + 3^{k-1} \left|\sum_{i=m(S_{N_j}+1)}^{j-1} p_{i,j}(X_i, x)\right|^k.
\end{aligned} \tag{5}$$

We have $|\sum_{i=m(S_{N_j}+1)}^{j-1} p_{i,j}(X_i, x)| \leq AmT_{N_j+1}$. So using the definition of the Orlicz norm and the fact that the random variables $(T_i)_{i\geq 2}$ are i.i.d., it holds for any $t \geq 0$,

$$\begin{aligned}
\mathbb{P}\left(\left|\sum_{i=m(S_{N_j}+1)}^{j-1} p_{i,j}(X_i, x)\right| \geq t\right) &\leq \mathbb{P}(T_{N_j+1} \geq \frac{t}{Am}) \leq \mathbb{P}(\max(T_1, T_2) \geq \frac{t}{Am}) \\
&\leq \mathbb{P}(T_1 \geq \frac{t}{Am}) + \mathbb{P}(T_2 \geq \frac{t}{Am}) \leq 4\exp(-\frac{t}{Am\tau}).
\end{aligned}$$

Hence,

$$\mathbb{E}\left[\left|\sum_{i=m(S_{N_j}+1)}^{j-1} p_{i,j}(X_i,x)\right|^k\right] = 4\int_0^{+\infty} \mathbb{P}\left(\left|\sum_{i=m(S_{N_j}+1)}^{j-1} p_{i,j}(X_i,x)\right|^k \geq t\right)dt$$
$$\leq 4\int_0^{+\infty} \exp(-\frac{t^{1/k}}{Am\tau})$$
$$\leq 4(Am\tau)^k \int_0^{+\infty} \exp(-v)kv^{k-1}dv$$
$$= 4(Am\tau)^k k!,$$

where we used that if G is an exponential random variable with parameter 1, then for any $p \in \mathbb{N}$, $\mathbb{E}[G^p] = p!$.

The random variable $Z_{2l}^j(x)$ is $\sigma(X_{m(S_{2l}+1)},\dots,X_{m(S_{2l+1}+1)-1})$-measurable. Let us insist that this holds because we consider that $h_{i,j} \equiv h_{1,j},\ \forall i,j$ which implies that $p_{i,j} \equiv p_{1,j},\ \forall i,j$. Hence for any $x \in E$, the random variables $(Z_{2l}^j(x))_l$ are independent (see Section 2.3). Moreover, one has that for any l, $\mathbb{E}[Z_{2l}^j(x)] = 0$. This is due to [34, Eq.(17.23) Theorem 17.3.1] together with Assumption 3 which gives that $\forall x' \in E,\quad \mathbb{E}_{X\sim\pi}[p_{i,j}(X,x')] = 0$. Let us finally notice that for any $x \in E$ and any $l \geq 0$, $|Z_{2l}^j(x)| \leq AmT_{2l+1}$, so $\|Z_{2l}^j(x)\|_{\psi_1} \leq Am\max(\|T_1\|_{\psi_1},\|T_2\|_{\psi_1}) \leq Am\tau$. First, we use Lemma 5 to obtain that

$$\mathbb{E}\left|\sum_{l=0}^{\lfloor N_j/2\rfloor} Z_{2l}^j(x)\right|^k \leq \mathbb{E}\max_{0\leq s\leq n-1}\left|\sum_{l=0}^{s} Z_{2l}^j(x)\right|^k \leq 2\times 4^k\mathbb{E}\left|\sum_{l=0}^{n-1} Z_{2l}^j(x)\right|^k,$$

where for the last inequality we gathered (7) with the left hand side of (6) from Lemma 5.

Lemma 5 *(see [12, Lemma 1.2.6])*
Let us consider some separable Banach space B endowed with the norm $\|\cdot\|$. Let X_i, $i \leq n$, be independent centered B-valued random variables with norms L_p for some $p \geq 1$ and let ε_i be independent Rademacher random variables independent of the variables X_i. Then

$$2^{-p}\mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i X_i\right\|^p \leq \mathbb{E}\left\|\sum_{i=1}^n X_i\right\|^p \leq 2^p\mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i X_i\right\|^p, \tag{6}$$

and

$$\mathbb{E}\max_{k\leq n}\left\|\sum_{i=1}^k X_i\right\|^p \leq 2^{p+1}\mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i X_i\right\|^p \tag{7}$$

Similarly, the random variables $(Z_{2l+1}^j(x))_l$ are independent and satisfy for any l, $\mathbb{E}[Z_{2l+1}^j(x)] = 0$. With an analogous approach, we get that

$$\mathbb{E}\left|\sum_{l=0}^{\lfloor (N_j-1)/2\rfloor} Z_{2l+1}^j(x)\right|^k \leq \mathbb{E}\max_{0\leq s\leq n-1}\left|\sum_{l=0}^{s} Z_{2l+1}^j(x)\right|^k \leq 2\times 4^k\mathbb{E}\left|\sum_{l=0}^{n-1} Z_{2l+1}^j(x)\right|^k.$$

Let us denote for any $j \in [n]$, $\mathbb{E}_{|X'_j}$ the conditional expectation with respect to the σ-algebra $\sigma(X'_j)$. Coming back to (5), we proved that

$$\begin{aligned}
&\mathbb{E}_{|X'_j}\left|\sum_{i=1}^{j-1} p_{i,j}(X_i,X'_j)\right|^k \\
&\leq \left(\frac{9}{4}\right)^{k-1}\mathbb{E}_{|X'_j}\left|\sum_{l=0}^{\lfloor N_j/2\rfloor} Z_{2l}^j(X'_j)\right|^k + \left(\frac{9}{2}\right)^{k-1}\mathbb{E}_{|X'_j}\left|\sum_{l=0}^{\lfloor (N_j-1)/2\rfloor} Z_{2l+1}^j(X'_j)\right|^k + 3^{k-1}\mathbb{E}_{|X'_j}\left|\sum_{i=m(S_{N_j}+1)}^{j-1} p_{i,j}(X_i,X'_j)\right|^k \\
&\leq 2\times 9^k\mathbb{E}_{|X'_j}\left|\sum_{l=0}^{n-1} Z_{2l+1}^j(X'_j)\right|^k + 2\times 18^k\mathbb{E}_{|X'_j}\left|\sum_{l=0}^{n-1} Z_{2l}^j(X'_j)\right|^k + 4(3Am\tau)^k k!.
\end{aligned} \tag{8}$$

It remains to bound the two expectations in (8). The two latter expectations will be control similarly and we give the details for the first one. We use the following Bernstein's inequality with the sequence of random variables $(Z^j_{2l+1}(x))_l$.

Lemma 6 *(Bernstein's ψ_1 inequality, [43, Lemma 2.2.11] and the subsequent remark).*
If $Y_1,\dots,Y_n$ are independent random variables such that $\mathbb{E}Y_i=0$ and $\|Y_i\|_{\psi_1}\le\tau$, then for every $t>0$,

$$\mathbb{P}\left(\left|\sum_{i=1}^n Y_i\right|>t\right)\le 2\exp\left(-\frac{1}{K}\min\left(\frac{t^2}{n\tau^2},\frac{t}{\tau}\right)\right),$$

for some universal constant $K>0$ ($K=8$ fits).

We obtain

$$\mathbb{P}\left(\left|\sum_{l=0}^{n-1} Z^j_{2l+1}(x)\right|>t\right)\le 2\exp\left(-\frac{1}{K}\min\left(\frac{t^2}{nA^2m^2\tau^2},\frac{t}{Am\tau}\right)\right).$$

We deduce that for any $x\in E$, any $j\in[n]$ and any $t\ge 0$,

$$\mathbb{E}\left[\left|\sum_{l=0}^{n-1} Z^j_{2l+1}(x)\right|^k\right]=\int_0^\infty \mathbb{P}\left(\left|\sum_{l=0}^{n-1} Z^j_{2l+1}(x)\right|^k>t\right)dt=2\int_0^\infty \exp\left(-\frac{1}{K}\min\left(\frac{t^{2/k}}{nA^2m^2\tau^2},\frac{t^{1/k}}{Am\tau}\right)\right)dt.$$

Let us remark that

$$\frac{t^{2/k}}{A^2m^2n\tau^2}\le\frac{t^{1/k}}{Am\tau}\Leftrightarrow t\le(nA\tau m)^k.$$

Hence for any $j\in[n]$,

$$\begin{aligned}
&\mathbb{E}\left[\left|\sum_{l=0}^{n-1} Z^j_{2l+1}(X'_j)\right|^k\right]\\
&\le 2\int_0^{(nA\tau m)^k}\exp\left(-\frac{t^{2/k}}{KnA^2m^2\tau^2}\right)dt+2\int_0^\infty\exp\left(-\frac{t^{1/k}}{KAm\tau}\right)dt.\\
&\le 2\int_0^{n/K}\exp(-v)\,\frac{k}{2}v^{k/2-1}\left(\sqrt{K}n^{1/2}A\tau m\right)^k dv+2\int_0^\infty\exp(-v)\,kv^{k-1}(KAm\tau)^k dv.\\
&\le 2\int_0^{n/K}\exp(-v)\,\frac{k}{2}v^{k/2-1}\left(\sqrt{K}n^{1/2}A\tau m\right)^k dv+2k\times(k-1)!(KAm\tau)^k\\
&\le k\left(\sqrt{K}n^{1/2}A\tau m\right)^k\int_0^{n/K}\exp(-v)\,v^{k/2-1}dv+2k!(KAm\tau)^k,
\end{aligned}$$

where we used again that if G is an exponential random variable with parameter 1, then for any $p\in\mathbb{N}$, $\mathbb{E}[G^p]=p!$. Since for any real $l\ge 1$,

$$\begin{aligned}
\int_0^{n/K}\exp(-v)\,v^{l-1}dv&=\sum_{r=0}^{+\infty}\frac{(-1)^r}{r!}\int_0^{n/K}v^{r+l-1}dv=\sum_{r=0}^{+\infty}\frac{(-1)^r}{r!}\frac{1}{r+l}(n/K)^{r+l}\\
&\le(n/K)^l\sum_{r=0}^{+\infty}\frac{(-1)^r}{r!}\frac{1}{l}(n/K)^r\le\frac{(n/K)^l}{l}e^{-\frac{n}{K}},
\end{aligned}$$

we get that

$$k\left(\sqrt{K}n^{1/2}A\tau m\right)^k\int_0^{n/K}\exp(-v)\,v^{k/2-1}dv\le 2\left(\sqrt{K}n^{1/2}A\tau m\right)^k e^{-n/K}(n/K)^{k/2}=2\,(KnA\tau m)^k\,e^{-n/K}.$$

Hence we proved that for some universal constant $K > 1$,

$$\mathbb{E}\left[\left|\sum_{l=0}^{n-1} Z_{2l+1}^j(x)\right|^k\right] \le 2\,(KnA\tau m)^k\, e^{-n/K} + 2k!(KAm\tau)^k \le 4k!(K^2Am\tau)^k,$$

since for all $k \ge 2$, $e^{-n/K}(n/K)^k/(k!) \le 1$. Using a similar approach, one can show the same bound for the second expectation in (8). We proved that for some universal constant $K > 1$,

$$\begin{aligned}
(\mathbb{E}[Z])^k &\le \sum_{j=2}^n \mathbb{E}\left[\mathbb{E}_{|X_j'}\left|\sum_{i=1}^{j-1} p_{i,j}(X_i, X_j')\right|^k\right] \\
&\le 2\times 9^k \sum_{j=2}^n \mathbb{E}\left[\mathbb{E}_{|X_j'}\left|\sum_{l=0}^{n-1} Z_{2l+1}^j(X_j')\right|^k\right] + 2\times 18^k \sum_{j=2}^n \mathbb{E}\left[\mathbb{E}_{|X_j'}\left|\sum_{l=0}^{n-1} Z_{2l}^j(X_j')\right|^k\right] + 4\sum_{j=2}^n (3Am\tau)^k k! \\
&\le 2n \times 18^k \times 4k!(K^2Am\tau)^k + 4n(3Am\tau)^k k! \\
&= 16n \times k!(KAm\tau)^k,
\end{aligned}$$

where in the last inequality, we still call K the universal constant defined by $18K^2$.

Upper-bounding U_n using the martingale structure Let

$$T + 1 := \inf\{l \in \mathbb{N} : V_l^k \ge w_n^k \text{ for some } k \ge 2\}.$$

Then, the event $\{T \le l\}$ depends only on $X_1, \ldots, X_l$ for all $l \ge 1$. Hence, T is a stopping time for the filtration $(\mathcal{G}_l)_l$ where $\mathcal{G}_l = \sigma((X_i)_{i\in[l]})$ and we deduce that $U_l^T := U_{l\wedge T}$ for $l = 0, \ldots, n$ is a martingale with respect to $(\mathcal{G}_l)_l$ with $U_0^T = U_0 = 0$ and $U_1^T = U_1 = 0$. We remark that $U_j^T - U_{j-1}^T = U_j - U_{j-1}$ if $T \ge j$ and zero otherwise, and that $\{T \ge j\}$ is $\mathcal{G}_{j-1}$ measurable. Then, the angle brackets of this martingale admit the following bound:

$$\begin{aligned}
A_n^k(U^T) &= \sum_{j=2}^n \mathbb{E}_{j-1}[(U_j^T - U_{j-1}^T)^k] \\
&= \sum_{j=2}^n \mathbb{E}_{j-1}|U_j - U_{j-1}|^k \mathbb{1}_{T\ge j} = \sum_{j=2}^n \mathbb{E}_{j-1}\left|\sum_{i=1}^{j-1} h(X_i, X_{j-1}, X_j)\right|^k \mathbb{1}_{T\ge j} \\
&= \sum_{j=2}^{n-1} V_j^k \mathbb{1}_{T=j} + V_n^k \mathbb{1}_{T\ge n} \le w_n^k\left(\sum_{j=2}^{n-1} \mathbb{1}_{T=j} + \mathbb{1}_{T\ge n}\right) \le w_n^k,
\end{aligned}$$

since, by definition of T, $V_j^k \le w_n^k$ for all k on $\{T \ge j\}$. Hence, Lemma 1 applied to the martingale U_n^T implies

$$\mathbb{E}e^{\alpha U_n^T} \le \exp\left(\sum_{k\ge 2} \frac{\alpha^k}{k!} w_n^k\right).$$

Also, since V_n^k is nondecreasing in n for each k, inequality (4) implies that

$$\mathbb{P}(T < n) \le \mathbb{P}\left(V_n^k \ge w_n^k \quad \text{for some } k \ge 2\right) \le 1.62e^{-u}.$$

Thus we deduce that for all $s \ge 0$,

$$\mathbb{P}(U_n \ge s) \le \mathbb{P}(U_n^T \ge s, T \ge n) + \mathbb{P}(T < n) \le e^{-\alpha s} \exp\left(\sum_{k\ge 2} \frac{\alpha^k}{k!} w_n^k\right) + 1.62e^{-u}. \tag{9}$$

The final step of the proof consists in simplifying $\exp\left(\sum_{k\ge 2} \frac{\alpha^k}{k!} w_n^k\right)$.

$$
\begin{aligned}
\sum_{k\geq 2}\frac{\alpha^k}{k!}w_n^k &= 2\delta_M\sum_{k\geq 2}\frac{\alpha^k}{k!}(1+\varepsilon)^{k-1}(\mathbb{E}[Z])^k\\
&\quad + 2\delta_M\sum_{k\geq 2}\frac{\alpha^k}{k!}(2+\varepsilon+\varepsilon^{-1})^{k-1}\,(3\|\Gamma\|)^k\,B_0^2A^{k-2}(nku)^{k/2}\\
&\quad + 2\delta_M\sum_{k\geq 2}\frac{\alpha^k}{k!}(1+\varepsilon^{-1})^{2k-2}\left(8\|\Gamma\|^2\right)^k(nA^2)A^{k-2}(ku)^k\\
&:= a_1+a_2+a_3.
\end{aligned}
$$

Bounding a_3. Using the inequality $k!\geq (k/e)^k$, we have,

$$
\begin{aligned}
a_3 &\leq 2\delta_M\sum_{k\geq 2}\alpha^k(1+\varepsilon^{-1})^{2k-2}\left(8\|\Gamma\|^2\right)^k(nA^2)A^{k-2}(eu)^k\\
&= 2\delta_M\alpha^2\left[\sqrt{n}A(1+\varepsilon^{-1})8\|\Gamma\|^2eu\right]^2\sum_{k\geq 2}\alpha^{k-2}(1+\varepsilon^{-1})^{2(k-2)}\left(8\|\Gamma\|^2\right)^{k-2}A^{k-2}(eu)^{k-2}\\
&= \frac{2\delta_M\alpha^2\left[\sqrt{n}A(1+\varepsilon^{-1})8\|\Gamma\|^2eu\right]^2}{1-\alpha(1+\varepsilon^{-1})^2\,(8\|\Gamma\|^2)\,Aeu}, \quad \text{for } \alpha < ((1+\varepsilon^{-1})^2\left(8\|\Gamma\|^2\right)Aeu)^{-1}.
\end{aligned}
$$

Bounding a_2. We use the inequality $k!\geq k^{k/2}$ because $(k/e)^k > k^{k/2}$ for $k\geq e^2$ and for k smaller, the inequality follows by direct verification. Hence,

$$
\begin{aligned}
a_2 &\leq 2\delta_M\sum_{k\geq 2}\alpha^k(2+\varepsilon+\varepsilon^{-1})^{k-1}\,(3\|\Gamma\|)^k\,B_0^2A^{k-2}(nu)^{k/2}\\
&= 2\delta_M(2+\varepsilon+\varepsilon^{-1})\alpha^2\left[3\|\Gamma\|B_0\sqrt{nu}\right]^2\sum_{k\geq 2}\alpha^{k-2}(2+\varepsilon+\varepsilon^{-1})^{k-2}\,(3\|\Gamma\|)^{k-2}A^{k-2}(nu)^{(k-2)/2}\\
&= \frac{2\delta_M(2+\varepsilon+\varepsilon^{-1})\alpha^2\left[3\|\Gamma\|B_0\sqrt{nu}\right]^2}{1-\alpha(2+\varepsilon+\varepsilon^{-1})\,(3\|\Gamma\|)A(nu)^{1/2}}, \quad \text{for } \alpha < ((2+\varepsilon+\varepsilon^{-1})\,(3\|\Gamma\|)A(nu)^{1/2})^{-1}.
\end{aligned}
$$

Bounding a_1. Using the bound previously obtained for $(\mathbb{E}[Z])^k$ we get,

$$
\begin{aligned}
a_1 &= 2\delta_M\sum_{k\geq 2}\frac{\alpha^k}{k!}(1+\varepsilon)^{k-1}(\mathbb{E}[Z])^k\\
&\leq 32\delta_M n\sum_{k\geq 2}\alpha^k(1+\varepsilon)^{k-1}(KAm\tau)^k\\
&\leq 32\delta_M n\alpha^2(1+\varepsilon)[KAm\tau]^2\sum_{k\geq 2}\alpha^{k-2}(1+\varepsilon)^{k-2}(KAm\tau)^{k-2}\\
&\leq \frac{32\delta_M n\alpha^2(1+\varepsilon)[KAm\tau]^2}{1-\alpha(1+\varepsilon)KAm\tau}, \quad \text{for } 0<\alpha<((1+\varepsilon)KAm\tau)^{-1}.
\end{aligned}
$$

Putting altogether we obtain

$$
\exp\left(\sum_{k\geq 2}\frac{\alpha^k}{k!}w_n^k\right)\leq \exp\left(\frac{\alpha^2W^2}{1-\alpha c}\right),
$$

where

$$
\begin{aligned}
W &= 6\sqrt{\delta_M}(1+\varepsilon)^{1/2}n^{1/2}KA\tau m\\
&\quad + \sqrt{2\delta_M}(2+\varepsilon+\varepsilon^{-1})^{1/2}3\|\Gamma\|B_0\sqrt{nu} + \sqrt{2\delta_M}A(1+\varepsilon^{-1})8\|\Gamma\|^2\sqrt{n}eu,
\end{aligned}
$$

and

$$c = \max\left[(1+\varepsilon)KA\tau m\ ,\ (2+\varepsilon+\varepsilon^{-1})(3\|\Gamma\|)A(nu)^{1/2}\ ,\ (1+\varepsilon^{-1})^2\left(8\|\Gamma\|^2\right)Aeu\right].$$

Using this estimate in (9) and taking $s = 2W\sqrt{u}+cu$ and $\alpha = \sqrt{u}/(W+c\sqrt{u})$ in this inequality yields

$$\mathbb{P}\left(U_n \geq 2W\sqrt{u}+cu\right) \leq e^{-u} + 1.62e^{-u} \leq (1+e)e^{-u}.$$

By taking $\varepsilon = 1/2$, we deduce that for any $u \geq 0$, it holds with probability at least $1-(1+e)e^{-u}$

$$\begin{aligned}\sum_{i<j} h_j^{(0)}(X_i,X_{j-1},X_j) \leq\quad & 12\sqrt{\delta_M}KA\tau m\sqrt{nu} + 18\sqrt{\delta_M}\|\Gamma\|B_0\sqrt{n}u + 100\sqrt{\delta_M}\|\Gamma\|^2A\sqrt{n}eu^{3/2}\\ & + 3KA\tau mu + 27A\|\Gamma\|\sqrt{n}u^{3/2} + 72A\|\Gamma\|^2eu^2,\end{aligned}$$

Denoting $\kappa := \max\left(12\sqrt{\delta_M}K\tau m\ ,\ 18\sqrt{\delta_M}\|\Gamma\|, 100\sqrt{\delta_M}\|\Gamma\|^2e, 3K\tau m, 72\|\Gamma\|^2e\right)$, we have with probability at least $1-(1+e)e^{-u}$

$$\sum_{i<j} h_j^{(0)}(X_i,X_{j-1},X_j) \leq \kappa\left(A\sqrt{n}\sqrt{u} + (A+B_0\sqrt{n})u + 2A\sqrt{n}u^{3/2} + Au^2\right).$$

3.1.2 Reasoning by descending induction with a logarithmic depth

As previously explained, we apply a proof similar to the one of the previous subsection on the $t_n := \lfloor q\log n\rfloor$ first terms of the decomposition (2), with $q > (\log(1/\rho))^{-1}$. Let us give the key elements to justify such approach by considering the second term of the decomposition (2), namely

$$\begin{aligned}&\sum_{i<j}\left(\mathbb{E}_{j-1}\left[h_{i,j}(X_i,X_j)\right] - \mathbb{E}_{j-2}\left[h_{i,j}(X_i,X_j)\right]\right)\\ =\ & \sum_{i=1}^{n-2}\sum_{j=i+2}^{n} h_{i,j}^{(1)}(X_i,X_{j-2},X_{j-1}) + \sum_{i=1}^{n-1}\left\{\mathbb{E}_i\left[h_{i,i+1}(X_i,X_{i+1})\right] - \mathbb{E}_{i-1}\left[h_{i,i+1}(X_i,X_{i+1})\right]\right\}\\ =\ & \sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1} h_{i,j}^{(1)}(X_i,X_{j-1},X_j) + \sum_{i=1}^{n-1}u_i(X_{i-1},X_i),\end{aligned}\tag{10}$$

where

$$h_{i,j}^{(1)}(x,y,z) = \int_w h_{i,j}(x,w)P(z,dw) - \int_w h_{i,j}(x,w)P^2(y,dw)$$

and

$$u_i(x,y) = \int_w h_{i,i+1}(x,w)P(x,dw) - \int_a\int_w P(y,da)P(a,dw)h_{i,i+1}(a,w).$$

We can upper-bound directly $\left|\sum_{i=1}^{n-1}u_i(X_{i-1},X_i)\right|$ by $2n\max_{i,j}\|h_{i,j}\|_\infty$ and we aim at proving a concentration result for the term

$$U_{n-1}^{(1)} := \sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1} h_{i,j}^{(1)}(X_i,X_{j-1},X_j) = \sum_{j=2}^{n-1}\sum_{i=1}^{j-1} h_{i,j}^{(1)}(X_i,X_{j-1},X_j),$$

using an approach similar to the one of the previous subsection. One can use exactly the same sketch of proof.

- Martingale structure
  Using the notation $Y_j^{(1)} = \sum_{i=1}^{j-1}h_{i,j}^{(1)}(X_i,X_{j-1},X_j)$, we have $U_{n-1}^{(1)} = \sum_{j=2}^{n-1}Y_j^{(1)}$ which shows that $(U_n^{(1)})_n$ is a martingale with respect to the σ-algebras $(G_l)_l$. Indeed, we have $\mathbb{E}_{j-1}[Y_j^{(1)}] = 0$.

- Talagrand's inequality
To upper-bound $(V_n^k)_n$, we split it as previously namely

$$V_n^k := \sum_{j=2}^{n-1} \mathbb{E}_{j-1} \left| \sum_{i=1}^{j-1} h_{i,j}^{(1)}(X_i, X_{j-1}, X_j) \right|^k = \sum_{j=2}^{n-1} \mathbb{E}_{j-1} \left| \sum_{i=1}^{j-1} \left(I_{i,j}^{(1)}(X_i, X_j) - \mathbb{E}_{j-1}[I_{i,j}^{(1)}(X_i, X_j)] \right) \right|^k,$$

where
$$I_{i,j}^{(1)}(x,z) = \int_w h_{i,j}(x,w) P(z,dw).$$

Using as previously Lemma 2 with $\varepsilon = 1/2$, we get

$$\begin{aligned} V_n^k &= \sum_{j=2}^{n-1} \mathbb{E}_{j-1} \left| \sum_{i=1}^{j-1} \left(I_{i,j}^{(1)}(X_i, X_j) - \mathbb{E}_{\tilde{X}\sim\pi}[I_{i,j}^{(1)}(X_i, \tilde{X})] + \mathbb{E}_{\tilde{X}\sim\pi}[I_{i,j}^{(1)}(X_i, \tilde{X})] - \mathbb{E}_{j-1}[I_{i,j}^{(1)}(X_i, X_j)] \right) \right|^k \\ &\leq (3/2)^{k-1} \sum_{j=2}^{n-1} \mathbb{E}_{j-1} \left| \sum_{i=1}^{j-1} \left(I_{i,j}^{(1)}(X_i, X_j) - \mathbb{E}_{\tilde{X}\sim\pi}[I_{i,j}^{(1)}(X_i, \tilde{X})] \right) \right|^k \\ &\quad + 3^{k-1} \sum_{j=2}^{n-1} \mathbb{E}_{j-1} \left| \sum_{i=1}^{j-1} \left(\mathbb{E}_{\tilde{X}\sim\pi}[I_{i,j}^{(1)}(X_i, \tilde{X})] - \mathbb{E}_{j-1}[I_{i,j}^{(1)}(X_i, X_j)] \right) \right|^k. \end{aligned}$$

Again, basic computations and Jensen's inequality lead to

$$\sum_{j=2}^{n-1} \mathbb{E}_{j-1} \left| \sum_{i=1}^{j-1} \left(\mathbb{E}_{\tilde{X}\sim\pi}[I_{i,j}^{(1)}(X_i, \tilde{X})] - \mathbb{E}_{j-1}[I_{i,j}^{(1)}(X_i, X_j)] \right) \right|^k = \sum_{j=2}^{n-1} \mathbb{E}_{j-1} \left| \sum_{i=1}^{j-1} p_{i,j}^{(1)}(X_i, X_j) \right|^k,$$

where
$$p_{i,j}^{(1)}(x,z) := I_{i,j}^{(1)}(x,z) - \mathbb{E}_{\tilde{X}\sim\pi}[I_{i,j}^{(1)}(x, \tilde{X})].$$

Hence, using Assumption 1 and Lemma 5 exactly like in the previous section, we get

$$V_n^k = 2 \times 3^{k-1} \sum_{j=2}^{n-1} \mathbb{E}_{j-1} \left| \sum_{i=1}^{j-1} p_{i,j}^{(1)}(X_i, X_j) \right|^k \leq 2 \times 3^{k-1} \delta_M \sum_{j=2}^{n-1} \mathbb{E}_{X_j'} \left| \sum_{i=1}^{j-1} p_{i,j}^{(1)}(X_i, X_j') \right|^k.$$

Then, one can use the same duality trick to show that the V_n^k can be controlled using the supremum of a sum of functions of the Markov chain $(X_i)_{i\geq 1}$ using [39, Theorem 3].

- Bounding $\exp(w_n^k \alpha^k / k!)$
The terms a_2 and a_3 can be bounded in a similar way. For the term a_1, we only need to show that $p_{i,j}^{(1)}$ satisfies $\mathbb{E}_{X_i\sim\pi}[p_{i,j}^{(1)}(X_i, z)] = 0,\ \forall z \in E$ in order to apply as previously a Bernstein's type inequality.

$$\begin{aligned} \mathbb{E}_{X_i\sim\pi}[p_{i,j}^{(1)}(X_i, z)] &= \int_{x_i} d\pi(x_i) \int_w h_{i,j}(x_i, w) P(z, dw) - \mathbb{E}_{X\sim\pi}\mathbb{E}_{\tilde{X}\sim\pi}[I_{i,j}^{(1)}(X, \tilde{X})] \\ &= \mathbb{E}_\pi[h_{i,j}] - \mathbb{E}_\pi[h_{i,j}] \quad \text{(Using Assumption 3)} \\ &= 0. \end{aligned}$$

- Conclusion of the proof
Let us consider the constants A_1 and B_1 defined as the counterparts of the constants A and B_0 by replacing the functions $\left(p_{i,j}\right)_{i,j}$ by $\left(p_{i,j}^{(1)}\right)_{i,j}$. One can easily see that $A_1 = A$. Let us give details about B_1.

For any $x \in E$,

$$\begin{aligned}
\mathbb{E}_{X'\sim\nu}\left[(p_{i,j}^{(1)})^2(x,X')\right] &= \int_z \left(I_{i,j}^{(1)}(x,z) - \mathbb{E}_{\tilde{X}\sim\pi}[I_{i,j}^{(1)}(x,\tilde{X})]\right)^2 d\nu(z)\\
&= \int_z \left(\int_w h_{i,j}(x,w)P(z,dw) - \int_w h_{i,j}(x,w)\underbrace{\int_a P(a,dw)d\pi(a)}_{=d\pi(w)}\right)^2 d\nu(z)\\
&= \mathbb{E}_{X'\sim\nu}\left[\mathbb{E}_{X\sim P(X',\cdot)}h_{i,j}(x,X) - \mathbb{E}_\pi[h_{i,j}]\right]^2,
\end{aligned}$$

and for any $y \in E$,

$$\begin{aligned}
\mathbb{E}_{\tilde{X}\sim\pi}\left[(p_{i,j}^{(1)})^2(\tilde{X},y)\right] &= \int_x \left(I_{i,j}^{(1)}(x,y) - \mathbb{E}_{\tilde{X}\sim\pi}[I_{i,j}^{(1)}(x,\tilde{X})]\right)^2 d\pi(x)\\
&= \int_x \left(\int_w h_{i,j}(x,w)P(y,dw) - \int_w h_{i,j}(x,w)\underbrace{\int_a P(a,dw)d\pi(a)}_{=d\pi(w)}\right)^2 d\pi(x)\\
&= \mathbb{E}_{\tilde{X}\sim\pi}\left[\mathbb{E}_{X\sim P(y,\cdot)}h_{i,j}(\tilde{X},X) - \mathbb{E}_\pi[h_{i,j}]\right]^2.
\end{aligned}$$

Hence we get that

$$B_1^2 := \max\left[\max_i \left\|\sum_{j=i+1}^n \mathbb{E}_{X\sim\nu}\left[(p_{i,j}^{(1)})^2(\cdot,X)\right]\right\|_\infty, \max_j \left\|\sum_{i=1}^{j-1}\mathbb{E}_{X\sim\pi}[(p_{i,j}^{(1)})^2(X,\cdot)]\right\|_\infty\right] \le B^2, \quad (11)$$

where we recall that

$$\begin{aligned}
B^2 = \max\Bigg[&\sup_{0\le k\le t_n}\max_i \sup_x \sum_{j=i+1}^n \mathbb{E}_{X'\sim\nu}\left[\mathbb{E}_{X\sim P^k(X',\cdot)}h_{i,j}(x,X) - \mathbb{E}_\pi[h_{i,j}]\right]^2,\\
&\sup_{0\le k\le t_n}\max_j \sup_y \sum_{i=1}^{j-1}\mathbb{E}_{\tilde{X}\sim\pi}\left[\mathbb{E}_{X\sim P^k(y,\cdot)}h_{i,j}(\tilde{X},X) - \mathbb{E}_\pi[h_{i,j}]\right]^2\Bigg].
\end{aligned}$$

This allows us to get a concentration inequality similar to the one of the previous subsection, namely for any $u > 0$, it holds with probability at least $1-(1+e)e^{-u}$,

$$\sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1} h_{i,j}^{(1)}(X_i,X_{j-1},X_j) \le \kappa\left(A\sqrt{n}\sqrt{u} + (A+B\sqrt{n})u + 2A\sqrt{n}u^{3/2} + Au^2\right)$$

Going back to (10), we get that for any $u > 0$, it holds with probability at least $1-(1+e)e^{-u}$,

$$\begin{aligned}
&\sum_{i<j}\left(\mathbb{E}_{j-1}\left[h_{i,j}(X_i,X_j)\right] - \mathbb{E}_{j-2}\left[h_{i,j}(X_i,X_j)\right]\right)\\
\le\quad &\kappa\left(A\sqrt{n}\sqrt{u} + (A+B\sqrt{n})u + 2A\sqrt{n}u^{3/2} + Au^2\right) + nA \qquad (12)
\end{aligned}$$

One can do the same analysis for the t_n first terms in the decomposition (2). Hence for any $u > 0$, it holds with probability at least $1-(1+e)e^{-u}t_n$,

$$\begin{aligned}
&\sum_{k=1}^{t_n}\sum_{i<j}\left(\mathbb{E}_{j-k+1}[h_{i,j}(X_i,X_j)] - \mathbb{E}_{j-k}[h_{i,j}(X_i,X_j)]\right)\\
\le\quad &\kappa t_n\left(A\sqrt{n}\sqrt{u} + (A+B\sqrt{n})u + 2A\sqrt{n}u^{3/2} + Au^2\right) + nt_nA. \qquad (13)
\end{aligned}$$

3.1.3 Bounding the remaining statistic with uniform ergodicity

In the previous steps of the proof, we decompose U_{stat} in t_n+1 terms (see (2)). The martingale structure of the first t_n terms of this decomposition allowed us to derive a concentration inequality for each of them. It remains to control the last term of this decomposition, namely

$$\sum_{i<j}\left(\mathbb{E}_{j-t_n}\left[h_{i,j}(X_i,X_j)\right]-\mathbb{E}\left[h_{i,j}(X_i,X_j)\right]\right),$$

where $t_n=\lfloor q\log n\rfloor$ with $q>2(\log(1/\rho))^{-1}$. In the following, we assume that $t_n\le n$, otherwise the last term of the decomposition (2) is an empty sum. Using our convention which states that for all $k<1$, $\mathbb{E}_k[\cdot]:=\mathbb{E}[\cdot]$, we need to control

$$\left|\sum_{i<j}\left(\mathbb{E}_{j-t_n}\left[h_{i,j}(X_i,X_j)\right]-\mathbb{E}\left[h_{i,j}(X_i,X_j)\right]\right)\right|\le(1)+(2),$$

with denoting $H_{i,j}=\mathbb{E}_{j-t_n}\left[h_{i,j}(X_i,X_j)\right]-\mathbb{E}\left[h_{i,j}(X_i,X_j)\right]$,

$$(1):=\left|\sum_{i=1}^{n-t_n}\sum_{j=i+t_n}^{n}H_{i,j}\right|=\left|\sum_{j=t_n+1}^{n}\sum_{i=1}^{j-t_n}H_{i,j}\right|,\quad (2):=\left|\sum_{i=1}^{n-1}\sum_{j=i+1}^{(i+t_n-1)\wedge n}H_{i,j}\right|=\left|\sum_{j=2}^{n}\sum_{i=(j-t_n+1)\vee 1}^{j-1}H_{i,j}\right|.$$

Let us first bound the term (1) splitting it in two terms,

$$(1)=\left|\sum_{j=t_n+1}^{n}\sum_{i=1}^{j-t_n}\mathbb{E}_{j-t_n}\left[h_{i,j}(X_i,X_j)\right]-\mathbb{E}\left[h_{i,j}(X_i,X_j)\right]\right|\le(1a)+(1b).$$

Using Assumption 3, it holds $\mathbb{E}_{\pi}[h_{i,j}]=\mathbb{E}_{\tilde{X}\sim\pi}[h_{i,j}(X_i,\tilde{X})]=\int_x h_{i,j}(X_i,x)d\pi(x)$. Hence we get that

$$\begin{aligned}
(1a)&:=\left|\sum_{j=t_n+1}^{n}\sum_{i=1}^{j-t_n}\mathbb{E}_{j-t_n}\left[h_{i,j}(X_i,X_j)\right]-\mathbb{E}_{\pi}[h_{i,j}]\right|\\
&\le\sum_{j=t_n+1}^{n}\left|\int_{x_j}\sum_{i=1}^{j-t_n}h_{i,j}(X_i,x_j)\left(P^{t_n}(X_{j-t_n},dx_j)-d\pi(x_j)\right)\right|\\
&\le\sum_{j=t_n+1}^{n}\sup_{x_j}\left|\sum_{i=1}^{j-t_n}h_{i,j}(X_i,x_j)\right|\sup_z\|P^{t_n}(z,\cdot)-\pi\|_{TV}\\
&\le\sum_{j=t_n+1}^{n}\sup_{x_j}\left|\sum_{i=1}^{j-t_n}h_{i,j}(X_i,x_j)\right|L\rho^{t_n}\le\sum_{j=t_n+1}^{n}\sup_{x_j}\left|\sum_{i=1}^{j-t_n}h_{i,j}(X_i,x_j)\right|L\frac{1}{n^2}\le LA,
\end{aligned}$$

where in the penultimate inequality we used that $\rho^{t_n}\le\rho^{q\log(n)}=n^{q\log(\rho)}\le n^{-2}$. Indeed $2+q\log(\rho)<0$ because we choose q such that $q>2(\log(1/\rho))^{-1}$.

Using again Assumption 3, it holds $\mathbb{E}_{\pi}[h_{i,j}]=\int_{x_i}\chi P^i(dx_i)\int_x h_{i,j}(x_i,x)d\pi(x)$. Hence, denoting χ the initial distribution of the Markov chain $(X_i)_{i\ge1}$, we get that

$$\begin{aligned}
(1b)&:=\left|\sum_{j=t_n+1}^{n}\sum_{i=1}^{j-t_n}\mathbb{E}_{\pi}[h_{i,j}]-\mathbb{E}\left[h_{i,j}(X_i,X_j)\right]\right|\\
&\le\sum_{j=t_n+1}^{n}\sum_{i=1}^{j-t_n}\left|\int_{x_i}\int_{x_j}h_{i,j}(x_i,x_j)\chi P^i(dx_i)\left(P^{j-i}(x_i,dx_j)-d\pi(x_j)\right)\right|\\
&\le\sum_{j=t_n+1}^{n}\sum_{i=1}^{j-t_n}\int_{x_i}\int_{x_j}\|h_{i,j}\|_\infty\chi P^i(dx_i)\left|P^{j-i}(x_i,dx_j)-d\pi(x_j)\right|
\end{aligned}$$

$$\leq \sum_{j=t_n+1}^{n} \sum_{i=1}^{j-t_n} \|h_{i,j}\|_\infty \underbrace{\int_{x_i} \chi P^i(dx_i)}_{=1} \sup_z \int_{x_j} \left| P^{j-i}(z, dx_j) - d\pi(x_j) \right|$$

$$= \sum_{j=t_n+1}^{n} \sum_{i=1}^{j-t_n} \|h_{i,j}\|_\infty \sup_z \|P^{j-i}(z,\cdot) - \pi\|_{TV}$$

$$\leq \sum_{j=t_n+1}^{n} \sum_{i=1}^{j-t_n} \|h_{i,j}\|_\infty L\rho^{j-i} \leq \sum_{j=t_n+1}^{n} \sum_{i=1}^{j-t_n} \|h_{i,j}\|_\infty L\rho^{t_n} \leq LA,$$

where in the penultimate inequality we used that $\rho^{t_n} \leq \rho^{q\log(n)} = n^{q\log(\rho)} \leq n^{-2}$.

Finally we bound coarsely (2) as follows

$$(2) = \left| \sum_{j=2}^{n} \sum_{i=(j-t_n+1)\vee 1}^{j-1} \mathbb{E}_{j-t_n}\left[h_{i,j}(X_i,X_j)\right] - \mathbb{E}\left[h_{i,j}(X_i,X_j)\right] \right| \leq Ant_n.$$

We deduce that

$$\sum_{i<j} \left(\mathbb{E}_{j-t_n}\left[h_{i,j}(X_i,X_j)\right] - \mathbb{E}\left[h_{i,j}(X_i,X_j)\right] \right) \leq A(2L + nt_n), \tag{14}$$

Coupling this result with the concentration result (13) concludes the proof of Theorem 1. Hence, for any $u > 0$ it holds with probability at least $1-(1+e)e^{-u}t_n$,

$$\sum_{i<j} \left(h_{i,j}(X_i,X_j) - \mathbb{E}\left[h_{i,j}(X_i,X_j)\right]\right) \leq \kappa t_n \left(A\sqrt{n}\sqrt{u} + (A + B\sqrt{n})u + 2A\sqrt{n}u^{3/2} + A\left[u^2 + n\right]\right),$$

for some constant $\kappa > 0$.

3.2 Proof of Theorem 2

To prove Theorem 2, one only needs to follow closely the steps of the proof of Theorem 1 and to bound coarsely the constant B^2 by nA^2.

3.3 Proof of Theorem 3

The proof of Theorem 3 is obtained from the one of Theorem 1 by avoiding the use of coarse bounds at two key steps. Note that we work under Assumption 4.(ii) which trivially holds since the chain is stationary.

First step: Do not use a coarse bound for the residual terms in the induction process.
The first change needed in the poof of Theorem 1 is at equation (10). In this part the proof of Theorem 1, we show by induction that our approach works with a logarithmic depth and we upper-bounded directly $\left|\sum_{i=1}^{n-1} u_i(X_{i-1},X_i)\right|$ by $2n\max_{i,j}\|h_{i,j}\|_\infty$. Considering now that the Markov chain is stationary, we are able to give a better control on this term. To convince the reader that our approach will work for all steps of the induction process of Section 3.1.2, let us consider a step $l \geq 1$ of the induction. In this situation, we want to control

$$\sum_{i<j} \left(\mathbb{E}_{j-l}\left[h_{i,j}(X_i,X_j)\right] - \mathbb{E}_{j-l-1}\left[h_{i,j}(X_i,X_j)\right]\right)$$
$$= \sum_{i=1}^{n-1} \sum_{j=i+l+1}^{n} h_{i,j}^{(l)}(X_i, X_{j-l-1}, X_{j-l}) + \sum_{i=1}^{n-1} \sum_{j=i+1}^{(i+l)\wedge n} \left(\mathbb{E}_{j-l}\left[h_{i,j}(X_i,X_j)\right] - \mathbb{E}_{j-l-1}\left[h_{i,j}(X_i,X_j)\right]\right),$$

where $h_{i,j}^{(l)}(x,y,z) := \int_w h_{i,j}(x,w)P^l(z,dw) - \int_w h_{i,j}(x,w)P^{l+1}(y,dw)$. We have already proved how to control the term of the left hand side in Section 3.1.2 and our purpose is now to explain how the term on the right hand side can be handled without using a coarse bound as in the proof of Theorem 1.

Note that by stationary of the chain we have

$$\mathbb{E}\left\{\mathbb{E}_{j-l}\left[h_{i,j}(X_i,X_j)\right]-\mathbb{E}_{j-l-1}\left[h_{i,j}(X_i,X_j)\right]\right\}=0.$$

Hence, denoting

$$f(X_1,\dots,X_n):=\sum_{i=1}^{n-1}\sum_{j=i+1}^{(i+l)\wedge n}\left(\mathbb{E}_{j-l}\left[h_{i,j}(X_i,X_j)\right]-\mathbb{E}_{j-l-1}\left[h_{i,j}(X_i,X_j)\right]\right),$$

we have that $\mathbb{E}[f(X_1,\dots,X_n)]=0$. Moreover, one can easily check that for any $(x_1,\dots,x_n)\in E^n$ and $(x'_1,\dots,x'_n)\in E^n$, it holds (since $l\le t_n$)

$$\left|f(x_1,\dots,x_n)-f(x'_1,\dots,x'_n)\right|\le\sum_{i=1}^{n}At_n\mathbb{1}_{x_i\neq x'_i}.$$

This follows from the fact that $\mathbb{E}_{j-l}\left[h_{i,j}(X_i,X_j)\right]$ is a function of the single variable X_{j-l} for $j\in\{i,\dots,(i+l)\wedge n\}$ and $i\in\{1,\dots,n-1\}$. Using McDiarmid's inequality for Markov chain (see [35, Corollary 2.10 and Remark 2.11]), we get that for any $u>0$, it holds with probability at least $1-2e^{-u}$,

$$\left|\sum_{i=1}^{n-1}\sum_{j=i+1}^{(i+l)\wedge n}\left(\mathbb{E}_{j-l}\left[h_{i,j}(X_i,X_j)\right]-\mathbb{E}_{j-l-1}\left[h_{i,j}(X_i,X_j)\right]\right)\right|\le 3At_n\sqrt{t_{mix}nu},$$

where t_{mix} is the mixing time of the Markov chain and is given by

$$t_{mix}:=\min\left\{t\ge 0\ :\ \sup_x\|P^t(x,\cdot)-\pi\|_{TV}<\frac{1}{4}\right\}.$$

Hence, coming back to Section 3.1.2 of the proof of Theorem 1, we obtain that for any $u>0$, it holds with probability at least $1-(1+e+2)e^{-u}$,

$$\begin{aligned}&\sum_{i<j}\left(\mathbb{E}_{j-l}\left[h_{i,j}(X_i,X_j)\right]-\mathbb{E}_{j-l-1}\left[h_{i,j}(X_i,X_j)\right]\right)\\ \le\quad&\kappa\left([C+At_n\sqrt{n}]\sqrt{u}+(A+B\sqrt{n})u+2A\sqrt{n}u^{3/2}+Au^2\right),\end{aligned}$$

for some constant $\kappa>0$. We conclude that for any $u>0$, it holds with probability at least $1-(3+e)e^{-u}t_n$,

$$\begin{aligned}&\sum_{k=1}^{t_n}\sum_{i<j}\left(\mathbb{E}_{j-k+1}[h_{i,j}(X_i,X_j)]-\mathbb{E}_{j-k}[h_{i,j}(X_i,X_j)]\right)\\ \le\quad&\kappa t_n\left(\left[C+At_n\sqrt{n}\right]\sqrt{u}+(A+B\sqrt{n})u+2A\sqrt{n}u^{3/2}+Au^2\right).\end{aligned}\tag{15}$$

To conclude the proof of Theorem 3, it remains to change a second and last part of the proof of Theorem 1.

Second step: Do not use a coarse bound to handle the remaining statistic with uniform ergodicity. In Section 3.1.3, we bounded coarsely the term (2). Considering now that the chain is stationary, we handle more finely this term and we split (2) in three different contributions.

$$(2)=\left|\sum_{j=2}^{n}\sum_{i=(j-t_n+1)\vee 1}^{j-1}\mathbb{E}_{j-t_n}\left[h_{i,j}(X_i,X_j)\right]-\mathbb{E}\left[h_{i,j}(X_i,X_j)\right]\right|\le(2a)+(2b)+(2c),$$

with

$$\begin{aligned}(2a)&:=\left|\sum_{j=2}^{n}\sum_{i=(j-t_n+1)\vee 1}^{j-\lfloor\frac{t_n}{2}\rfloor}\mathbb{E}_{j-t_n}\left[h_{i,j}(X_i,X_j)\right]-\mathbb{E}_{\pi}\left[h_{i,j}\right]\right|,\\ (2b)&:=\left|\sum_{j=2}^{n}\sum_{i=(j-t_n+1)\vee 1}^{j-\lfloor\frac{t_n}{2}\rfloor}\mathbb{E}_{\pi}\left[h_{i,j}\right]-\mathbb{E}\left[h_{i,j}(X_i,X_j)\right]\right|,\\ \text{and}\quad(2c)&:=\left|\sum_{j=2}^{n}\sum_{i=(j-\lfloor\frac{t_n}{2}\rfloor+1)\vee 1}^{j-1}\mathbb{E}_{j-t_n}\left[h_{i,j}(X_i,X_j)\right]-\mathbb{E}\left[h_{i,j}(X_i,X_j)\right]\right|.\end{aligned}$$

Using Assumption 3, we have that $\mathbb{E}_\pi[h_{i,j}] = \int_{x_i} P^{i-j+t_n}(X_{j-t_n}, dx_i) \int_{x_j} h_{i,j}(x_i, x_j) d\pi(x_j)$. Hence we get,

$$
\begin{aligned}
(2a) &:= \left| \sum_{j=2}^{n} \sum_{i=(j-t_n+1)\vee 1}^{j-\lfloor \frac{t_n}{2} \rfloor} \mathbb{E}_{j-t_n}\left[h_{i,j}(X_i, X_j)\right] - \mathbb{E}_\pi\left[h_{i,j}\right] \right| \\
&\leq \sum_{j=2}^{n} \sum_{i=(j-t_n+1)\vee 1}^{j-\lfloor \frac{t_n}{2} \rfloor} \left| \int_{x_i} \int_{x_j} h_{i,j}(x_i, x_j) P^{i-j+t_n}(X_{j-t_n}, dx_i) \left(P^{j-i}(x_i, dx_j) - d\pi(x_j)\right) \right| \\
&\leq \sum_{j=2}^{n} \sum_{i=(j-t_n+1)\vee 1}^{j-\lfloor \frac{t_n}{2} \rfloor} \|h_{i,j}\|_\infty \int_{x_i} P^{i-j+t_n}(X_{j-t_n}, dx_i) \int_{x_j} \left|P^{j-i}(x_i, dx_j) - d\pi(x_j)\right| \\
&\leq \sum_{j=2}^{n} \sum_{i=(j-t_n+1)\vee 1}^{j-\lfloor \frac{t_n}{2} \rfloor} \|h_{i,j}\|_\infty \underbrace{\int_{x_i} P^{i-j+t_n}(X_{j-t_n}, dx_i)}_{=1} \sup_y \int_{x_j} \left|P^{j-i}(y, dx_j) - d\pi(x_j)\right| \\
&\leq \sum_{j=2}^{n} \sum_{i=(j-t_n+1)\vee 1}^{j-\lfloor \frac{t_n}{2} \rfloor} \|h_{i,j}\|_\infty \sup_y \|P^{j-i}(y, \cdot) - \pi\|_{TV} \\
&\leq \sum_{j=2}^{n} \sum_{i=(j-t_n+1)\vee 1}^{j-\lfloor \frac{t_n}{2} \rfloor} \|h_{i,j}\|_\infty L\rho^{j-i} \leq \sum_{j=2}^{n} \sum_{i=(j-t_n+1)\vee 1}^{j-\lfloor \frac{t_n}{2} \rfloor} \|h_{i,j}\|_\infty L\rho^{t_n/2} \leq LAt_n,
\end{aligned}
$$

where we used that $\rho^{t_n/2} \leq \rho^{q\log(n)/2} = n^{q\log(\rho)/2} \leq n^{-1}$. Indeed $1 + q\log(\rho)/2 < 0$ because we choose q such that $q > 2(\log(1/\rho))^{-1}$.

$$
\begin{aligned}
(2b) &:= \left| \sum_{j=2}^{n} \sum_{i=(j-t_n+1)\vee 1}^{j-\lfloor \frac{t_n}{2} \rfloor} \mathbb{E}_\pi\left[h_{i,j}\right] - \mathbb{E}\left[h_{i,j}(X_i, X_j)\right] \right| \\
&\leq \sum_{j=2}^{n} \sum_{i=(j-t_n+1)\vee 1}^{j-\lfloor \frac{t_n}{2} \rfloor} \left| \int_{x_i} \int_{x_j} h_{i,j}(x_i, x_j) d\pi(x_i) \left(d\pi(x_j) - P^{j-i}(x_i, dx_j)\right) \right| \\
&\leq \sum_{j=2}^{n} \sum_{i=(j-t_n+1)\vee 1}^{j-\lfloor \frac{t_n}{2} \rfloor} \|h_{i,j}\|_\infty \int_{x_i} d\pi(x_i) \int_{x_j} \left|d\pi(x_j) - P^{j-i}(x_i, dx_j)\right| \\
&\leq \sum_{j=2}^{n} \sum_{i=(j-t_n+1)\vee 1}^{j-\lfloor \frac{t_n}{2} \rfloor} \|h_{i,j}\|_\infty \underbrace{\int_{x_i} d\pi(x_i)}_{=1} \sup_y \int_{x_j} \left|d\pi(x_j) - P^{j-i}(y, dx_j)\right| \\
&\leq \sum_{j=2}^{n} \sum_{i=(j-t_n+1)\vee 1}^{j-\lfloor \frac{t_n}{2} \rfloor} \|h_{i,j}\|_\infty \sup_y \|P^{j-i}(y, \cdot) - \pi\|_{TV} \\
&\leq \sum_{j=2}^{n} \sum_{i=(j-t_n+1)\vee 1}^{j-\lfloor \frac{t_n}{2} \rfloor} \|h_{i,j}\|_\infty L\rho^{j-i} \leq \sum_{j=2}^{n} \sum_{i=(j-t_n+1)\vee 1}^{j-\lfloor \frac{t_n}{2} \rfloor} \|h_{i,j}\|_\infty L\rho^{t_n/2} \leq LAt_n,
\end{aligned}
$$

where we used that $\rho^{t_n/2} \leq \rho^{q\log(n)/2} = n^{q\log(\rho)/2} \leq n^{-1}$.

$$
\begin{aligned}
(2c) &:= \left| \sum_{j=2}^{n} \sum_{i=(j-\lfloor \frac{t_n}{2} \rfloor+1)\vee 1}^{j-1} \mathbb{E}_{j-t_n}\left[h_{i,j}(X_i,X_j)\right] - \mathbb{E}\left[h_{i,j}(X_i,X_j)\right]\right| \\
&\leq \sum_{j=2}^{n} \sum_{i=(j-\lfloor \frac{t_n}{2} \rfloor+1)\vee 1}^{j-1} \left| \int_{x_j} \int_{x_i} P^{j-i}(x_i,dx_j) h_{i,j}(x_i,x_j)\left(P^{i-j+t_n}(X_{j-t_n},dx_i) - d\pi(x_i)\right)\right| \\
&\leq \sum_{j=2}^{n} \sum_{i=(j-\lfloor \frac{t_n}{2} \rfloor+1)\vee 1}^{j-1} \|h_{i,j}\|_\infty \sup_z \|P^{i-j+t_n}(z,\cdot)-\pi\|_{TV} \\
&\leq \sum_{j=\lfloor \frac{t_n}{2} \rfloor}^{n} \sum_{i=(j-\lfloor \frac{t_n}{2} \rfloor+1)}^{j-1} \|h_{i,j}\|_\infty L\rho^{i-j+t_n} + \sum_{j=2}^{\lfloor \frac{t_n}{2} \rfloor} \sum_{i=(j-\lfloor \frac{t_n}{2} \rfloor+1)\vee 1}^{j-1} \|h_{i,j}\|_\infty L\rho^{i-j+t_n} \\
&\leq \sum_{j=\lfloor \frac{t_n}{2} \rfloor}^{n} \sum_{i=(j-\lfloor \frac{t_n}{2} \rfloor+1)}^{j-1} \|h_{i,j}\|_\infty L\rho^{t_n/2} + t_n^2\|h_{i,j}\|_\infty L \leq LA\left(t_n^2 + nt_n\rho^{t_n/2}\right) \leq 2LAt_n^2,
\end{aligned}
$$

where we used that $\rho^{t_n/2} \leq \rho^{q\log(n)/2} = n^{q\log(\rho)/2} \leq n^{-1}$. We deduce that

$$
\sum_{i<j}\left(\mathbb{E}_{j-t_n}\left[h_{i,j}(X_i,X_j)\right] - \mathbb{E}\left[h_{i,j}(X_i,X_j)\right]\right) \leq AL\left(2+2q\log(n)+2q^2\log(n)^2\right),
$$

Coupling this result with the concentration result (15) concludes the proof of Theorem 3. Hence, for any $u>0$ it holds with probability at least $1-(3+e)e^{-u}t_n$,

$$
\begin{aligned}
& \sum_{i<j}\left(h_{i,j}(X_i,X_j) - \mathbb{E}\left[h_{i,j}(X_i,X_j)\right]\right) \\
\leq \quad & \kappa \log n\left(\left[C + A\log(n)\sqrt{n}\right]\sqrt{u} + (A+B\sqrt{n})u + 2A\sqrt{n}u^{3/2} + A\left[u^2+\log n\right]\right).
\end{aligned}
$$

Acknowledgements

This work was supported by a grant from Région Ile de France.

References

[1] R. Adamczak. A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electronic Journal of Probability*, 13, 10 2007.

[2] Aliprantis and Border. *Infinite dimensional analysis*. Springer, 2006.

[3] P. Ango Nze. Critères d'ergodicité géométrique ou arithmétique de modèles linéaires pertubés à représentation Markovienne. *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics*, 326(3):371 – 376, 1998.

[4] M. A. Arcones and E. Gine. Limit theorems for U-processes. *Ann. Probab.*, 21(3):1494–1542, 07 1993.

[5] E. Behrends. *Introduction to Markov chains*, volume 228. Springer, 2000.

[6] P. Bertail and S. Clémençon. A renewal approach to Markovian U-statistics. *Mathematical Methods of Statistics*, 20(2):79–105, 2011.

[7] I. Borisov and N. Volodko. A note on exponential inequalities for the distribution tails of canonical von Mises statistics of dependent observations. *Statistics & Probability Letters*, 96:287–291, Jan 2015.

[8] T. R. Boucher. A Hoeffding inequality for Markov chains using a generalized inverse. *Statistics & probability letters*, 79(8):1105–1107, 2009.

[9] S. Clémençon, G. Ciolek, and P. Bertail. Concentration inequalities for regenerative and Harris recurrent Markov chains with applications to statistical learning. In *Séminaire généraliste de l'équipe de Probabilités et Statistiques*, Nancy, France, May 2017.

[10] Y. De Castro and Q. Duchemin. Markov Random Geometric Graph (MRGG): A Growth Model for Temporal Dynamic Networks. working paper or preprint, June 2020.

[11] Y. De Castro, C. Lacour, and T. M. Pham Ngoc. Adaptive estimation of nonparametric geometric graphs. *Mathematical Statistics and Learning*, 2020.

[12] V. de la Pena and E. Giné. Decoupling, from dependence to independence, randomly stopped processes, U-statistics and processes, martingales and beyond. *Journal of the American Statistical Association*, 95, 09 2000.

[13] V. H. de la Peña and S. J. Montgomery-Smith. Decoupling inequalities for the tail probabilities of multivariate U-statistics. *The Annals of Probability*, pages 806–816, 1995.

[14] P. Del Moral and A. Guionnet. Central limit theorem for nonlinear filtering and interacting particle systems. *Ann. Appl. Probab.*, 9(2):275–297, 05 1999.

[15] P. Doukhan and M. Ghindes. Estimations dans le processus: "$x_{n+1} = f(x_n) + \varepsilon_n$". *C.R. Acad. Sci. Paris Sér. A*, 291:61–64, 01 1980.

[16] P. Eichelsbacher and U. Schmock. Large deviations for products of empirical measures of dependent sequences. *Markov Process. Related Fields*, 7(3):435–468, 2001.

[17] P. Eichelsbacher and U. Schmock. Rank-dependent moderate deviations of U-empirical measures in strong topologies. *Probability theory and related fields*, 126(1):61–90, 2003.

[18] J. Fan, B. Jiang, and Q. Sun. Hoeffding's lemma for Markov chains and its applications to statistical learning. *The Journal of Machine Learning Research*, 2018.

[19] D. Ferré, L. Hervé, and J. Ledoux. Limit theorems for stationary Markov processes with L2-spectral gap. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 48(2):396–423, May 2012.

[20] G. Fort, E. Moulines, P. Priouret, and P. Vandekerkhove. A simple variance inequality for U-statistics of a Markov chain with applications. *Statistics and Probability Letters*, 82(6):1193–1201, 2012.

[21] E. Giné and R. Nickl. *Mathematical foundations of infinite-dimensional statistical models*, volume 40. Cambridge University Press, 2016.

[22] E. Giné and R. Nickl. *Mathematical foundations of infinite-dimensional statistical models*. Cambridge university press, 2021.

[23] E. Giné, R. Latała, and J. Zinn. Exponential and moment inequalities for U-statistics. *High Dimensional Probability II*, page 13–38, 2000.

[24] P. W. Glynn and D. Ormoneit. Hoeffding's inequality for uniformly ergodic Markov chains. *Statistics & probability letters*, 56(2):143–146, 2002.

[25] F. Han. An exponential inequality for U-statistics under mixing conditions. *Journal of Theoretical Probability*, 31(1):556–578, Nov 2016.

[26] C. Houdré and P. Reynaud-Bouret. Exponential inequalities for U-statistics of order two with constants. *Stochastic Inequalities and Applications. Progress in Probability*, 56, 01 2002.

[27] B. Jiang, Q. Sun, and J. Fan. Bernstein's inequality for general Markov chains. *arXiv preprint arXiv:1805.10721*, 2018.

[28] E. Joly and G. Lugosi. Robust estimation of U-statistics. *Stochastic Processes and their Applications*, In Memoriam: Evarist Giné:3760–3773, 2016.

[29] G. L. Jones. On the Markov chain central limit theorem. *Probab. Surveys*, 1:299–320, 2004.

[30] S. Kwapien and W. Woyczynski. *Random Series and Stochastic Integrals: Single and Multiple: Single and Multiple*. Probability and Its Applications. Birkhäuser Boston, 2002.

[31] M. Lerasle, N. M. Magalhães, and P. Reynaud-Bouret. Optimal kernel selection for density estimation. *Progress in Probability*, page 425–460, 2016.

[32] F. Lindsten, R. Douc, and E. Moulines. Uniform ergodicity of the particle Gibbs sampler. *Scandinavian Journal of Statistics*, 42(3):775–797, Feb 2015.

[33] P. Massart. *Concentration inequalities and model selection*, volume 6. Springer, 2007.

[34] S. Meyn and R. Tweedie. *Markov Chains and Stochastic Stability*, volume 92. 01 1993.

[35] D. Paulin. Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electronic Journal of Probability*, 20(0), 2015.

[36] S. Rao et al. A Hoeffding inequality for Markov chains. *Electronic Communications in Probability*, 24, 2019.

[37] G. O. Roberts and J. S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probability Surveys*, 1(0):20–71, 2004.

[38] R. Rockafellar. *Convex Analysis*. Princeton Landmarks in Mathematics and Physics. Princeton University Press, 1970.

[39] P.-M. Samson et al. Concentration of measure inequalities for Markov chains and Phi-mixing processes. *The Annals of Probability*, 28(1):416–461, 2000.

[40] Y. Shen, F. Han, and D. Witten. Exponential inequalities for dependent V-statistics via random Fourier features. *Electronic Journal of Probability*, 25(0), 2020.

[41] M. A. Suchard, R. E. Weiss, and J. S. Sinsheimer. Bayesian selection of continuous-time Markov chain evolutionary models. *Molecular biology and evolution*, 18(6):1001–1013, 2001.

[42] R. S. Sutton and A. G. Barto. Reinforcement learning: An introduction. *Cambridge, MA: MIT Press*, 2011.

[43] A. Van Der Vaart and J. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer New York, 2013.

[44] Y. Wang, R. Khardon, D. Pechyony, and R. Jones. Online learning with pairwise loss functions. *CoRR*, 2013.

SUPPLEMENTARY MATERIAL

For the sake of completeness, we provide in this supplementary material complete proofs of some Lemmas and we recall important definitions and properties on Markov chains useful for our paper.

- Section A: Definitions and properties on Markov chains
  In the section, we recall useful definitions and results on Markov chains regarding ergodicity, spectral gaps and the splitting method.

- Section B: Additional proofs
  This section contains the proofs of some Lemmas useful to establish Theorems 1, 2 and 3.

A Definitions and properties for Markov chains

We recall some definitions and well-known results on Markov chains that are useful for our paper.

A.1 Spectral gap

This section is largely inspired from [18]. Let us consider that the Markov chain $(X_i)_{i\geq 1}$ admits a unique invariant distribution π on E.
For any real-valued, Σ-measurable function $h : E \to \mathbb{R}$, we define $\pi(h) := \int h(x)d\pi(x)$. The set

$$\mathscr{L}_2(E,\Sigma,\pi) := \{h : \pi(h^2) < \infty\}$$

is a Hilbert space endowed with the inner product

$$\langle h_1, h_2\rangle_\pi = \int h_1(x)h_2(x)d\pi(x),\ \forall h_1,h_2 \in \mathscr{L}^2(E,\Sigma,\pi).$$

The map

$$\|\cdot\|_\pi : h \in \mathscr{L}_2(E,\Sigma,\pi) \mapsto \|h\|_\pi = \sqrt{\langle h,h\rangle_\pi},$$

is a norm on $\mathscr{L}_2(E,\Sigma,\pi)$. $\|\cdot\|_\pi$ naturally allows to define the norm of a linear operator T on $\mathscr{L}_2(E,\Sigma,\pi)$ as

$$N_\pi(T) = \sup\{\|Th\|_\pi : \|h\|_\pi = 1\}.$$

To each transition probability kernel $P(x,B)$ with $x \in E$ and $B \in \Sigma$ invariant with respect to π, we can associate a bounded linear operator $h \mapsto \int h(y)P(\cdot,dy)$ on $\mathscr{L}_2(E,\Sigma,\pi)$. Denoting this operator P, we get

$$Ph(x) = \int h(y)P(x,dy),\ \forall x \in E,\ \forall h \in \mathscr{L}_2(E,\Sigma,\pi).$$

Let $\mathscr{L}_2^0(\pi) := \{h \in \mathscr{L}_2(E,\Sigma,\pi)\ :\ \pi(h) = 0\}$. We define the absolute spectral gap of a Markov operator.

Definition 2 *(Spectral gap) A Markov operator P reversible admits a spectral gap $1-\lambda$ if*

$$\lambda := \sup\left\{\frac{\|Ph\|_\pi}{\|h\|_\pi}\ :\ h \in \mathscr{L}_2^0(\pi),\ h \neq 0\right\} < 1.$$

Remark Uniform ergocity ensures that the chain admits a spectral gap (see [19, Section 2.3]).

A.2 The splitting method

We describe the construction of the split chain that we use in our proof. Let us consider some Markov chain $(X_n)_n$ on the probability space $(E, \Sigma, \mathbb{P})$ with transition kernel P. We assume that there exists a small set $C \in \Sigma$, there exist a positive integer m, a constant $\delta_m > 0$, and a probability measure μ on E with the following minorisation condition holds

$$\forall x \in C, \quad \forall A \in \Sigma, \quad P^m(x, A) \geq \delta_m \mu(A).$$

We give only the construction of the split chain in the case where $m = 1$ and we refer to [34] for further details on the construction of the split chain. Each point x in E is splitted in $x_0 = (x, 0) \in E_0 = E \times \{0\}$ and $x_1 = (x, 1) \in E_1 = E \times \{1\}$. Each set B in Σ is splitted in $B_0 = B \times \{0\}$ and $B_1 = B \times \{1\}$. Thus, we have defined a new probability space (E^*, Σ^*) where $E^* = E_0 \cup E_1$ and $\Sigma^* = \sigma(B_0, B_1 : B \in \Sigma)$. A measure η on Σ splits into two measures on E_0 and E_1 by defining η^* on (E^*, Σ^*) through

$$\begin{aligned}
\eta^*(B_0) &= \eta(B \cap C)(1 - \delta_m) + \eta(B \cap C^c) \\
\eta^*(B_1) &= \eta(B \cap C)\delta_m.
\end{aligned}$$

Now we define a new transition probability $P^*(\cdot, \cdot)$ on (E^*, Σ^*) with

$$P^*((x, e), \cdot) = \begin{cases} P(x, \cdot)^* & \text{if } x \notin C \text{ and } e = 0 \\ (1 - \delta_m)^{-1} (P(x, \cdot)^* - \delta_m \mu^*) & \text{if } x \in C \text{ and } e = 0 \\ \mu^* & \text{if } e = 1 \end{cases}.$$

The split chain associated with $(X_n)_n$ is then defined as the Markov chain $((\overline{X}_n, R_n))_n$ on E^* with transition kernel P^*. After the previous rigorous construction of the split chain, we can give the following interpretation of the method. Each time the chain reaches C, there is a possibility for the chain to regenerate. Each time the chain is at $x \in C$, a coin is tossed with probability of success δ_m. If the toss is successful, then the chain is moved according to the probability distribution μ, otherwise, according to $(1 - \delta_m)^{-1} (P(x, \cdot) - \delta_m \mu(\cdot))$. Overall, the dynamic of the chain is not affected by this coin toss, but at each time the toss is successful, the chains regenerates with regeneration distribution μ independent from x. In particular if Assumption 1 holds, the whole space is small and if in addition $m = 1$, the chain regenerates with probability δ_m at each time step.

B Additional proofs

B.1 Proof of Proposition 1

Since the split chain has the same distribution as the original Markov chain, we get that $(\overline{X}_i)_i$ is ψ-irreducible for some measure ψ and uniformly ergodic. From [34, Theorem 16.0.2], Assumption 1 ensures that for every measurable set $A \subset E \times \{0, 1\}$ such that $\psi(A) > 0$, there exists some $\kappa_A > 1$ such that

$$\sup_x \mathbb{E}[\kappa_A^{\tau_A} | \overline{X}_1 = x] < \infty,$$

where $\tau_A := \inf\{n \geq 1 \ : \ \overline{X}_n \in A\}$ is the first hitting time of the set A. Let us recall that T_1 and T_2 are defined as hitting times of the atom of the split chain $E \times \{1\}$ which is accessible (i.e. the atom has a positive ψ-measure). Hence, there exist $C > 0$ and $\kappa > 1$ such that,

$$\sup_x \mathbb{E}[\kappa^{\tau_{E \times \{1\}}} | \overline{X}_1 = x] = \sup_x \mathbb{E}[\exp(\tau_{E \times \{1\}} \log(\kappa)) | \overline{X}_1 = x] \leq C.$$

Considering $k \geq 1$ such that $C^{1/k} \leq 2$, a straight forward application of Jensen inequality gives that $\max(\|T_1\|_{\psi_1}, \|T_2\|_{\psi_1}) \leq k / \log(\kappa)$. ■

B.2 Talagrand inequality for Markov chains

In the section, we show that in the proof of Theorem 1, we can use the concentration inequality for the supremum of an empirical process of [39, Theorem 3].
Let us consider the sequence of random variables $W=(W_1,\dots,W_n)$ on the probability space $(\Omega,\mathscr{A},\mathbb{P})$ taking values in the measurable space $S=E\times F^{n-1}$ where F is the subset of the set $\mathscr{F}(E,\mathbb{R})$ of all measurable functions from (E,Σ) to $(\mathbb{R},\mathscr{B}(\mathbb{R}))$ that are bounded by A. Note that

$$\{0_{\mathscr{F}(E,\mathbb{R})}\}\cup\{p_{i,j}(x,\cdot)\ :\ x\in E,\ i,j\in[n]\}\subset F.$$

We define

$$\mathscr{P}:=\left\{D\in\mathscr{P}(F)\ :\ \forall i\in[n-1],\forall j\in\{i+1,\dots,n\},\ f_{i,j}^{-1}(D)\in\Sigma\right\},$$

where $\mathscr{P}(F)$ is the powerset of F and where $\forall i\in[n-1]$, $\forall j\in\{i+1,\dots,n\}$,

$$\begin{aligned} f_{i,j}:(E,\Sigma)&\to(F,\mathscr{P}(F))\\ x&\mapsto p_{i,j}(x,\cdot).\end{aligned}$$

Then we have the following straightforward result.

Lemma 7 *$\mathscr{P}$ is a σ-algebra on F.*

In the following, we endow the space F with the σ-algebra $\mathscr{P}$ and we consider on S the product σ-algebra given by

$$\mathscr{S}:=\sigma\left(\left\{C\times D_2\times\cdots\times D_n\ :\ C\in\Sigma,\ D_j\in\mathscr{P}\ \forall j\in\{2,\dots,n\}\right\}\right).$$

For all $i\in[n]$, we define W_i by

$$W_i:=\left(X_i,\underbrace{0,\dots,0}_{(i-1)\text{ times}},p_{i,i+1}(X_i,\cdot),p_{i,i+2}(X_i,\cdot),\dots,p_{i,n}(X_i,\cdot)\right).$$

Hence for all $i\in[n]$, W_i is $\sigma(X_i)$-measurable. Let us consider for any $i\in[n-1]$,

$$\begin{aligned}\Phi_i:(E,\Sigma)&\to(S,\mathscr{S})\\ x&\mapsto\left(x,\underbrace{0_F,\dots,0_F}_{(i-1)\text{ times}},p_{i,i+1}(x,\cdot),\dots,p_{i,n}(x,\cdot)\right).\end{aligned}$$

Then, one can directly see that for all $i\in[n-1]$, $W_i=\Phi_i(X_i)$ and by construction of $\mathscr{P}$ and $\mathscr{S}$, Φ_i is measurable. Indeed, each coordinate of Φ_i is measurable by construction of $\mathscr{P}$ and this ensures that Φ_i is measurable thanks to the following Lemma.

Lemma 8 *(See [2, Lemma 4.49]) Let $(X,\Sigma),(X_1,\Sigma_1)$ and (X_2,Σ_2) be measurable spaces, and let $f_1:X\to X_1$ and $f_2:X\to X_2$. Define $f:X\to X_1\times X_2$ by $f(x)=(f_1(x),f_2(x))$. Then $f:(X,\Sigma)\to(X_1\times X_2,\Sigma_1\otimes\Sigma_2)$ is measurable if and only if the two functions $f_1:(X,\Sigma)\to(X_1,\Sigma_1)$ and $f_2:(X,\Sigma)\to(X_2,\Sigma_2)$ are both measurable.*

Then it holds for any $i\in\{2,\dots,n-1\}$ and any $G\in\mathscr{S}$,

$$\begin{aligned}&\mathbb{P}\left(W_i\in G\mid W_{i-1}\right)\\ &=\mathbb{P}\left(\Phi_i(X_i)\in G\mid W_{i-1}\right)\\ &=\mathbb{P}\left(\Phi_i(X_i)\in G\mid X_{i-1}\right)\\ &=\mathbb{P}\left(X_i\in\Phi_i^{-1}(G)\mid X_{i-1}\right)\\ &=P\left(X_{i-1},\Phi_i^{-1}(G)\right)\\ &=\left[(\Phi_i)_\#P(X_{i-1},\cdot)\right](G),\end{aligned}\tag{16}$$

where $(\Phi_i)_\# P(X_{i-1},\cdot)$ denotes the pushforward measure of the measure $P(X_{i-1},\cdot)$ by the measurable map Φ_i. We deduce that W_i is non-homogeneous Markov chain. Moreover, (16) proves that the transition kernel of the Markov chain $(W_k)_k$ from state $i-1$ to state i is given by $K^{(i-1,i)}$ where for all $(x,p_2,\dots,p_n)\in S$ and for all $G\in\mathscr{S}$,

$$K^{(i-1,i)}((x,p_2,\dots,p_n),G)=[(\Phi_i)_\# P(x,\cdot)](G).$$

One can easily generalize this notation. Let us consider some $i,j\in[n]$ with $i<j$ and let us denote $K^{(i,j)}$ the transition kernel of the Markov chain $(W_k)_k$ from state i to state j. Then for all $x\in E$, for all $p_2,\dots,p_n\in F$ and for all $G\in\mathscr{S}$,

$$K^{(i,j)}((x,p_2,\dots,p_n)\,,G)=\left[(\Phi_j)_\# P^{j-i}(x,\cdot)\right](G),$$

We introduce the mixing matrix $\Gamma=(\gamma_{i,j})_{1\le i,j\le n-1}$ where coefficients are defined by

$$\gamma_{i,j}:=\sup_{w_i\in S}\sup_{z_i\in S}\|\mathscr{L}(W_j|W_i=w_i)-\mathscr{L}(W_j|W_i=z_i)\|_{TV}.$$

For any $w\in S=E\times F^{n-1}$, we denote $w^{(1)}$ the first coordinate of the vector w. Hence, $w^{(1)}$ is an element of E. Then

$$\begin{aligned}
\gamma_{i,j}&=\sup_{w_i\in S}\sup_{z_i\in S}\sup_{G\in\mathscr{S}}\left|\left[(\Phi_j)_\# P^{j-i}(w_i^{(1)},\cdot)\right](G)-\left[(\Phi_j)_\# P^{j-i}(z_i^{(1)},\cdot)\right](G)\right|\\
&=\sup_{w_i\in S}\sup_{z_i\in S}\sup_{G\in\mathscr{S}}\left|P^{j-i}\left(w_i^{(1)},\Phi_j^{-1}(G)\right)-P^{j-i}\left(z_i^{(1)},\Phi_j^{-1}(G)\right)\right|\\
&\le\sup_{w_i\in S}\sup_{z_i\in S}\sup_{C\in\Sigma}\left|P^{j-i}\left(w_i^{(1)},C\right)-P^{j-i}\left(z_i^{(1)},C\right)\right|\\
&=\sup_{x_i\in E}\sup_{x_i'\in E}\sup_{C\in\Sigma}\left|P^{j-i}(x_i,C)-P^{j-i}\left(x_i',C\right)\right|\\
&=\sup_{x_i\in E}\sup_{x_i'\in E}\|P^{j-i}(x_i,\cdot)-\pi(\cdot)+\pi(\cdot)-P^{j-i}(x_i',\cdot)\|_{TV}\\
&\le\sup_{x_i\in E}\|P^{j-i}(x_i,\cdot)-\pi(\cdot)\|_{TV}+\sup_{x_i'\in E}\|P^{j-i}(x_i',\cdot)-\pi(\cdot)\|_{TV}\\
&\le 2L\rho^{j-i},
\end{aligned}$$

where in the first inequality we used that $\Phi_j:(E,\Sigma)\to(S,\mathscr{S})$ is measurable and in the last inequality we used the uniform ergodicity of the Markov chain $(X_i)_{i\ge1}$. We deduce that

$$\|\Gamma\|\le 2L\left\|Id+\sum_{l=1}^{n-1}\rho^l N_l\right\|,$$

where $N_l=\left(n_{i,j}^{(l)}\right)_{1\le i,j\le n-1}$ represents the nilpotent matrix of order l defined by

$$n_{i,j}^{(l)}=\begin{cases}1 & \text{if } j-i=l\\ 0 & \text{otherwise.}\end{cases}$$

Since for each $1\le l\le n-1$, $\|N_l\|\le 1$, it follows from the triangular inequality that

$$\|\Gamma\|\le 2L\sum_{l=0}^{n-1}\rho^l\le\frac{2L}{1-\rho}.$$

To conclude the proof and get the concentration result stated in Lemma 4, one only needs to apply [39, Theorem 3] with the class of functions $\mathscr{F}$ and with the Markov chain $(W_k)_k$. Let us recall that $\mathscr{F}$ is defined by $\mathscr{F}=\{f_\xi:\sum_{j=2}^n\mathbb{E}|\xi_j(X_j')|^{k/(k-1)}=1\}$ where for any $\xi=(\xi_2,\dots,\xi_n)\in\prod_{i=2}^n L^{k/(k-1)}(\nu)$,

$$\forall w=(x,p_2,\dots,p_n)\in E\times F^{n-1},\quad f_\xi(w)=\sum_{j=2}^n\int p_j(y)\xi_j(y)d\nu(y).$$

B.3 Hoeffding inequality for uniformly ergodic Markov chains

A large number of different Hoeffding inequalities for Markov chains can be found in the literature such as in [18],[8],[36] or [24]. We need such concentration result in our proofs and our goal is to find one that match our assumptions and that does not ask condition on the initial distribution of the chain. As stated in [8, Lemma 1], for a ψ-irreducible and aperiodic Markov chain, a Hoeffding inequality with exponents independent of the initial state of the chain exists if and only if the chain is uniformly ergodic. Some Hoeffding inequalities for uniformly ergodic Markov chains without condition on the initial distribution already exist (see [24] or [8]), but they require n to be large enough to hold or can involve quantities related to the chain that we did not use so far (such that the Drazin inverse of $I-P$). This is the reason why we propose here to prove briefly a different Hoeffding inequality for uniformly ergodic Markov chains that holds for any sample size n, any initial distribution of the chain and that only uses the notations from Section 2.

Proposition 3 *Let $(X_i)_{i\geq 1}$ be a Markov chain on E uniformly ergodic (namely satisfying Assumption 1) with invariant distribution π and let us consider some function $f: E \to \mathbb{R}$ such that $\mathbb{E}_{X\sim\pi}[f(X)] = 0$ and $\|f\|_\infty \leq A$. Then it holds for any $t \geq 0$*

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} f(X_i)\right| \geq t\right) \leq 16\exp\left(-\frac{1}{K}\min\left(\frac{t^2(\mathbb{E}T_2)}{nA^2m^2\tau^2}, \frac{t}{Am\tau}\right)\right),$$

for some universal constant K.
In particular, since by definition of τ and of the Orlicz norm we have $\mathbb{E}[T_2] \leq \left(\mathbb{E}[e^{T_2/\tau}]-1\right)\tau \leq \tau$, it holds for any $t \geq 0$

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} f(X_i)\right| \geq t\right) \leq 16\exp\left(-\frac{1}{K(m,\tau)}\frac{t^2}{nA^2}\right),$$

where $K(m,\tau) = 2Km^2\tau^2$ for some universal constant $K > 0$.

Proof of Proposition 3.
Let us consider $N = \sup\{i \in \mathbb{N} : mS_{i+1} + m - 1 \geq n\}$. Then,

$$\begin{aligned}\left|\sum_{i=1}^{n} f(X_i)\right| &= \left|\sum_{l=0}^{N} Z_l + \sum_{i=m(S_N+1)}^{n} f(X_i)\right| \\ &\leq \left|\sum_{l=0}^{\lfloor N/2\rfloor} Z_{2l}\right| + \left|\sum_{l=0}^{\lfloor (N-1)/2\rfloor} Z_{2l+1}\right| + \left|\sum_{i=m(S_N+1)}^{n} f(X_i)\right|. \end{aligned} \tag{17}$$

We have $|\sum_{i=m(S_N+1)}^{n} f(X_i)| \leq AmT_{N+1}$. So using the definition of the Orlicz norm and the fact that the random variables $(T_i)_{i\geq 2}$ are i.i.d., it holds for any $t \geq 0$,

$$\begin{aligned}\mathbb{P}\left(\left|\sum_{i=m(S_N+1)}^{n} f(X_i)\right| \geq t\right) &\leq \mathbb{P}(T_{N+1} \geq \frac{t}{Am}) \\ &\leq \mathbb{P}(\max(T_1,T_2) \geq \frac{t}{Am}) \\ &\leq \mathbb{P}(T_1 \geq \frac{t}{Am}) + \mathbb{P}(T_2 \geq \frac{t}{Am}) \\ &\leq 4\exp\left(-\frac{t}{Am\tau}\right).\end{aligned}$$

In order to control the first two terms in (17), we need to describe the tail behaviour of the random variable N with Lemma 9.

Lemma 9 *(See [1, Lemma 5]) If $\|T_1\|_{\psi_1}, \|T_2\|_{\psi_1} \leq \tau$, then*

$$\mathbb{P}(N > R) \leq 2\exp\left(-\frac{n\mathbb{E}T_2}{8\tau^2}\right),$$

where $R = \lfloor 3n/(\mathbb{E}T_2)\rfloor$.

The random variable Z_{2l} is $\sigma(X_{m(S_{2l}+1)}, \dots, X_{m(S_{2l+1}+1)-1})$-measurable. Hence the random variables $(Z_{2l})_l$ are independent (see Section 2.3). Moreover, one has that for any l, $\mathbb{E}[Z_{2l}] = 0$. This is due to [34, Eq.(17.23) Theorem 17.3.1] together with the assumption that $\mathbb{E}_{X\sim\pi}[f(X)] = 0$. Let us finally notice for any $l \geq 0$, $|Z_{2l}| \leq AmT_{2l+1}$, so $\|Z_{2l}\|_{\psi_1} \leq Am\max(\|T_1\|_{\psi_1}, \|T_2\|_{\psi_1}) \leq Am\tau$. One can similarly get that $(Z_{2l+1})_l$ are independent with $\mathbb{E}[Z_{2l+1}] = 0$ and $\|Z_{2l+1}\|_{\psi_1} \leq Am\tau$ for all $l \in \mathbb{N}$. Using these facts we have for any $t \geq 0$,

$$\begin{aligned}
&\mathbb{P}\left(\left|\sum_{l=0}^{\lfloor N/2\rfloor} Z_{2l}\right| + \left|\sum_{l=0}^{\lfloor (N-1)/2\rfloor} Z_{2l+1}\right| \geq t\right)\\
&\leq \mathbb{P}\left(\left|\sum_{l=0}^{\lfloor N/2\rfloor} Z_{2l}\right| + \left|\sum_{l=0}^{\lfloor (N-1)/2\rfloor} Z_{2l+1}\right| \geq t\ , N \leq R\right) + 2\exp\left(-\frac{n\mathbb{E}T_2}{8\tau^2}\right)\\
&\leq \mathbb{P}\left(\max_{0\leq s\leq \lfloor R/2\rfloor}\left|\sum_{l=0}^{s} Z_{2l}\right| \geq t/2\right) + \mathbb{P}\left(\max_{0\leq s\leq \lfloor (R-1)/2\rfloor}\left|\sum_{l=0}^{s} Z_{2l+1}\right| \geq t/2\right) + 2\exp\left(-\frac{n\mathbb{E}T_2}{8\tau^2}\right)\\
&\leq 3\mathbb{P}\left(\left|\sum_{l=0}^{\lfloor R/2\rfloor} Z_{2l}\right| \geq t/6\right) + 3\mathbb{P}\left(\left|\sum_{l=0}^{\lfloor (R-1)/2\rfloor} Z_{2l+1}\right| \geq t/6\right) + 2\exp\left(-\frac{n\mathbb{E}T_2}{8\tau^2}\right) \quad \text{(Using Lemma 10)}\\
&\leq 12\exp\left(-\frac{1}{8}\min\left(\frac{t^2}{36RA^2m^2\tau^2}, \frac{t}{6Am\tau}\right)\right) + 2\exp\left(-\frac{n\mathbb{E}T_2}{8\tau^2}\right),
\end{aligned}$$

where we used Lemma 6 in the last inequality.

Lemma 10 *(see [30, Proposition 1.1.1]) If $X_1, X_2, \dots$ are independent Banach space valued random variables (not necessarily identically distributed), and if $S_k = \sum_{i=1}^{k} X_i$, then*

$$\mathbb{P}\left(\max_{1\leq j\leq k}\|S_j\| > t\right) \leq 3\max_{1\leq j\leq k}\mathbb{P}\left(\|S_j\| > t/3\right).$$

Gathering the previous results, we obtain that for any $t \geq 0$

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} f(X_i)\right| \geq t\right) \leq 12\exp\left(-\frac{1}{8}\min\left(\frac{t^2(\mathbb{E}T_2)}{36\times 12\times nA^2m^2\tau^2}, \frac{t}{12Am\tau}\right)\right) + 2\exp\left(-\frac{n\mathbb{E}T_2}{8\tau^2}\right) + 4\exp\left(-\frac{t}{2Am\tau}\right).$$

Since the left hand side of the previous inequality is zero for $t \geq nA$, and since $m \geq 1$, we obtain Proposition 3. ■

B.4 Bernstein's inequality for non-stationary Markov chains

This subsection is dedicated to the proof of Proposition 4 which is used in the proof of Theorem 1. A Bernstein type concentration inequality was proved in [27] for stationary Markov chains. Following the work of [18], we extend the previous Bernstein inequality to the framework of non-stationary Markov chains with initial distribution that satisfies Assumption 4.(ii). Contrary to Proposition 3, this concentration handles the setting where the sum involves different functions $(f_i)_i$. This is of interest for Theorem 1 when we allow the kernel functions $h_{i,j}$ to depend on both i and j.

Proposition 4 *Suppose that the sequence $(X_i)_{i\geq 1}$ is a Markov chain satisfying Assumptions 1 and 4.(ii) with invariant distribution π and with an absolute spectral gap $1-\lambda > 0$. Let us consider some $n \in \mathbb{N}^*$ and* *bounded real valued functions $(f_i)_{1\leq i\leq n}$ such that for any $i\in\{1,\dots,n\}$, $\int f_i(x)d\pi(x)=0$ and $\|f_i\|_\infty\leq c$ for some $c>0$. Let $\sigma^2=\sum_{i=1}^n\int f_i^2(x)d\pi(x)/n$. Then for any $\varepsilon\geq 0$ it holds*

$$\mathbb{P}\left(\sum_{i=1}^n f_i(X_i)\geq\varepsilon\right)\leq\left\|\frac{d\chi}{d\pi}\right\|_{\pi,p}\exp\left(-\frac{\varepsilon^2/(2q)}{A_2 n\sigma^2+A_1 c\varepsilon}\right),$$

where $A_2:=\frac{1+\lambda}{1-\lambda}$ and $A_1:=\frac{1}{3}\mathbb{1}_{\lambda=0}+\frac{5}{1-\lambda}\mathbb{1}_{\lambda>0}$. q is the constant introduced in Assumption 4.(ii). Stated otherwise, for any $u>0$ it holds

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n f_i(X_i)>\frac{2quA_1c}{n}+\sqrt{\frac{2quA_2\sigma^2}{n}}\right)\leq\left\|\frac{d\chi}{d\pi}\right\|_{\pi,p}e^{-u}.$$

B.4.1 Proof of Proposition 4.

Let us recall that we denote indifferently $\mathbb{P}_\chi$ or $\mathbb{P}$ the probability distribution of the Markov chain $(X_i)_{i\geq 1}$ when the distribution of the first state X_1 is χ, whereas $\mathbb{P}_\pi$ refers to the distribution of the Markov chain when the distribution of the first state X_1 is the invariant measure π.
In [27], they proved that for any $0\leq t<(1-\lambda)/(5c)$, it holds

$$\mathbb{E}_\pi\left[e^{t\sum_{i=1}^n f_i(X_i)}\right]\leq\exp\left(\frac{n\sigma^2}{c^2}(e^{tc}-tc-1)+\frac{n\sigma^2\lambda t^2}{1-\lambda-5ct}\right).$$

We deduce that for any $0\leq t<(1-\lambda)/(5cq)$,

$$\begin{aligned}
\mathbb{E}_\chi\left[e^{t\sum_{i=1}^n f_i(X_i)}\right]&\leq\mathbb{E}_\pi\left[\frac{d\chi}{d\pi}(X_1)e^{t\sum_{i=1}^n f_i(X_i)}\right]\\
&\leq\left\{\mathbb{E}_\pi\left[\left|\frac{d\chi}{d\pi}(X_1)\right|^p\right]\right\}^{1/p}\left\{\mathbb{E}_\pi\left[e^{qt\sum_{i=1}^n f_i(X_i)}\right]\right\}^{1/q}\\
&=\left\|\frac{d\chi}{d\pi}\right\|_{\pi,p}\left\{\mathbb{E}_\pi\left[e^{qt\sum_{i=1}^n f_i(X_i)}\right]\right\}^{1/q}\\
&\leq\left\|\frac{d\chi}{d\pi}\right\|_{\pi,p}\left\{\exp\left(\frac{n\sigma^2}{c^2}(e^{tqc}-tqc-1)+\frac{n\sigma^2\lambda q^2t^2}{1-\lambda-5cqt}\right)\right\}^{1/q}\\
&=\left\|\frac{d\chi}{d\pi}\right\|_{\pi,p}\exp\left(\frac{n\sigma^2}{qc^2}(e^{tqc}-tqc-1)+\frac{n\sigma^2\lambda qt^2}{1-\lambda-5cqt}\right).
\end{aligned}\tag{18}$$

Let us define

$$g_1(t)=\begin{cases}0 & \text{if } t<0\\ \frac{n\sigma^2}{qc^2}(e^{tqc}-tqc-1) & \text{if } t\geq 0\end{cases}$$

and

$$g_2(t)=\begin{cases}0 & \text{if } t<0\\ \frac{n\sigma^2\lambda qt^2}{1-\lambda-5cqt} & \text{if } 0\leq t<\frac{1-\lambda}{5cq}\\ +\infty & \text{if } t\geq\frac{1-\lambda}{5cq}\end{cases}.$$

In order to lower-bound the convex conjugate of the function g_1+g_2, we will need the convex conjugate of g_1 and g_2 which are provided by Lemma 11. The proof of Lemma 11 can be found in Section B.4.2.

Lemma 11 *g_1 and g_2 are closed proper convex functions with convex conjugates*

$$\forall\varepsilon_1\in\mathbb{R},\quad g_1^*(\varepsilon_1)=\begin{cases}\frac{n\sigma^2}{qc^2}h_1\left(\frac{\varepsilon_1 c}{n\sigma^2}\right) & \text{if } \varepsilon_1\geq 0\\ +\infty & \text{if } \varepsilon_1<0\end{cases}\tag{19}$$

with $h_1(u) = (1+u)\log(1+u) - u \geq \frac{u^2}{2(1+u/3)}$ *for any* $u \geq 0$, *and*

$$\forall \varepsilon_2 \in \mathbb{R}, \quad g_2^*(\varepsilon_2) = \begin{cases} \frac{(1-\lambda)\varepsilon^2}{qn\sigma^2\lambda} h_2(\frac{5c\varepsilon_2}{n\sigma^2\lambda}) & \text{if } \varepsilon_2 \geq 0 \\ +\infty & \text{if } \varepsilon_2 < 0 \end{cases} \tag{20}$$

with $h_2(u) = \left(\frac{(\sqrt{u+1}-1}{u}\right)^2 \geq \frac{1}{2(u+2)}$.

Since $g_1(t) = O(t^2)$ and $g_2(t) = O(t^2)$ as $t \to 0^+$, $t\varepsilon - g_1(t) - g_2(t) > 0$ for small enough $t > 0$, and $t\varepsilon - g_1(t) - g_2(t) \leq 0$ for $t \leq 0$. Hence

$$(g_1 + g_2)^*(\varepsilon) = \sup_{0 \leq t < (1-\lambda)/(5cq)} \varepsilon t - g_1(t) - g_2(t) = \sup_{t \in \mathbb{R}} \varepsilon t - g_1(t) - g_2(t).$$

• If $\lambda > 0$, then by the Moreau-Rockafellar formula [38, Theorem 16.4], the convex conjugate of $g_1 + g_2$ is the infimal convolution of their conjugates g_1^* and g_2^*, namely

$$(g_1 + g_2)^*(\varepsilon) = \inf\left\{g_1^*(\varepsilon_1) + g_2^*(\varepsilon_2) : \varepsilon = \varepsilon_1 + \varepsilon_2,\ \varepsilon_1, \varepsilon_2 \in \mathbb{R}\right\}.$$

Using (19) and (20), this reads as

$$(g_1 + g_2)^*(\varepsilon) = \inf\left\{\frac{n\sigma^2}{qc^2} h_1\left(\frac{\varepsilon_1 c}{n\sigma^2}\right) + \frac{(1-\lambda)\varepsilon_2^2}{qn\sigma^2\lambda} h_2\left(\frac{5c\varepsilon_2}{n\sigma^2\lambda}\right) : \varepsilon = \varepsilon_1 + \varepsilon_2,\ \varepsilon_1, \varepsilon_2 \geq 0\right\}.$$

Bounding $h_1(u) \geq \frac{u^2}{2(1+u/3)}$ and $h_2(u) \geq \frac{1}{2(u+2)}$, we have

$$\begin{aligned}(g_1 + g_2)^*(\varepsilon) &\geq \inf\left\{\frac{1}{qc^2} \frac{c^2\varepsilon_1^2}{2(n\sigma^2 + c\varepsilon_1/3)} + \frac{(1-\lambda)\varepsilon_2^2}{qn\sigma^2\lambda} \frac{1}{\frac{10c\varepsilon_2}{n\sigma^2\lambda} + 4} : \varepsilon = \varepsilon_1 + \varepsilon_2,\ \varepsilon_1, \varepsilon_2 \geq 0\right\} \\ &\geq \inf\left\{\frac{\varepsilon_1^2}{2q(n\sigma^2 + c\varepsilon_1/3)} + \frac{(1-\lambda)\varepsilon_2^2}{2q} \frac{1}{5c\varepsilon_2 + 2n\sigma^2\lambda} : \varepsilon = \varepsilon_1 + \varepsilon_2,\ \varepsilon_1, \varepsilon_2 \geq 0\right\}.\end{aligned}$$

Using the fact that $\varepsilon_1^2/a + \varepsilon_2^2/b \geq (\varepsilon_1 + \varepsilon_2)^2/(a+b)$ for any non-negative $\varepsilon_1, \varepsilon_2$ and positive a, b yield

$$\begin{aligned}(g_1 + g_2)^*(\varepsilon) &\geq \inf\left\{\frac{(\varepsilon_1 + \varepsilon_2)^2}{2q(n\sigma^2 + c\varepsilon_1/3) + \frac{2q}{(1-\lambda)}(5c\varepsilon_2 + 2n\sigma^2\lambda)} : \varepsilon = \varepsilon_1 + \varepsilon_2,\ \varepsilon_1, \varepsilon_2 \geq 0\right\} \\ &= \inf\left\{\frac{\varepsilon^2/(2q)}{\frac{1+\lambda}{1-\lambda} n\sigma^2 + c\varepsilon_1/3 + \frac{5c\varepsilon}{1-\lambda} - \frac{5c\varepsilon_1}{1-\lambda}} : \varepsilon = \varepsilon_1 + \varepsilon_2,\ \varepsilon_1, \varepsilon_2 \geq 0\right\} \\ &\geq \frac{\varepsilon^2/(2q)}{\frac{1+\lambda}{1-\lambda} n\sigma^2 + \frac{5c\varepsilon}{1-\lambda}},\end{aligned}$$

where we used for the last inequality that for any $\varepsilon_1 \geq 0$,

$$c\varepsilon_1/3 - \frac{5c\varepsilon_1}{1-\lambda} = \frac{c\varepsilon_1}{3(1-\lambda)}(1 - \lambda - 15) < 0.$$

• If $\lambda = 0$,

$$(g_1 + g_2)^*(\varepsilon) = g_1^*(\varepsilon) = \frac{n\sigma^2}{qc^2} h_1\left(\frac{\varepsilon_1 c}{n\sigma^2}\right) \geq \frac{\varepsilon^2/(2q)}{n\sigma^2 + c\varepsilon/3}.$$

We deduce from the previous computations that for any $t, \varepsilon \geq 0$ it holds

$$
\begin{aligned}
&\mathbb{P}_\chi\left(\sum_{i=1}^n f_i(X_i) \geq \varepsilon\right) \\
&\leq e^{-\varepsilon t}\mathbb{E}_\chi\left[e^{t\sum_{i=1}^n f_i(X_i)}\right] \text{ using Markov's inequality} \\
&\leq e^{-\varepsilon t}\left\|\frac{d\chi}{d\pi}\right\|_{\pi,p} \exp\left(\frac{n\sigma^2}{qc^2}(e^{tqc}-tqc-1)+\frac{n\sigma^2\lambda q t^2}{1-\lambda-5cqt}\right) \text{ using (18)} \\
&\leq \left\|\frac{d\chi}{d\pi}\right\|_{\pi,p} \exp\left(-(g_1+g_2)^*(\varepsilon)\right) \\
&\leq \left\|\frac{d\chi}{d\pi}\right\|_{\pi,p} \times \begin{cases} \exp\left(-\frac{\varepsilon^2/(2q)}{\frac{1+\lambda}{1-\lambda}n\sigma^2+\frac{5c\varepsilon}{1-\lambda}}\right) & \text{if } \lambda>0 \\ \exp\left(-\frac{\varepsilon^2/(2q)}{n\sigma^2+c\varepsilon/3}\right) & \text{if } \lambda = 0 \end{cases}.
\end{aligned}
$$

B.4.2 Proof of Lemma 11

The convex conjugate of g_1 is usual and follows from easy computations. We focus on the convex conjugate of g_2 which requires non-trivial computations.
Let $f_\varepsilon(t) = \varepsilon t - \frac{n\sigma^2\lambda q t^2}{1-\lambda-5cqt}$ for any $0 \leq t < (1-\lambda)/(5cq)$. We have for any $0 \leq t < (1-\lambda)/(5cq)$,

$$
f_\varepsilon'(t) = \varepsilon - \frac{2n\sigma^2\lambda qt(1-\lambda-5cqt)+5cq^2n\sigma^2\lambda t^2}{(1-\lambda-5cqt)^2}.
$$

Hence, for $0 \leq t < (1-\lambda)/(5cq)$ such that $f_\varepsilon'(t) = 0$ we have

$$
\begin{aligned}
& & \varepsilon(1-\lambda-5cqt)^2 - 2n\sigma^2\lambda qt(1-\lambda-5cqt) - 5cq^2n\sigma^2\lambda t^2 &= 0 \\
\Leftrightarrow & & \varepsilon(1-\lambda)^2 - 10\varepsilon(1-\lambda)cqt + 25\varepsilon c^2q^2t^2 - 2n\sigma^2\lambda q(1-\lambda)t + 10n\sigma^2cq^2\lambda t^2 - 5n\sigma^2cq^2\lambda t^2 &= 0 \\
\Leftrightarrow & & \varepsilon(1-\lambda)^2 - 10\varepsilon(1-\lambda)cqt + 25\varepsilon c^2q^2t^2 - 2n\sigma^2\lambda q(1-\lambda)t + 5n\sigma^2cq^2\lambda t^2 &= 0.
\end{aligned}
$$

We are looking for the roots of a polynomial of degree 2 in t. The discriminant is

$$
\begin{aligned}
\Delta &= (10\varepsilon(1-\lambda)cq + 2n\sigma^2\lambda q(1-\lambda))^2 - 4\varepsilon(1-\lambda)^2(25\varepsilon c^2q^2 + 5n\sigma^2cq^2\lambda) \\
&= 4(1-\lambda)^2q^2\left[(5\varepsilon c + n\sigma^2\lambda)^2 - \varepsilon(25c^2\varepsilon + 5n\sigma^2c\lambda)\right] \\
&= 4(1-\lambda)^2q^2\left[25\varepsilon^2c^2 + 10n\sigma^2\lambda\varepsilon c + n^2\sigma^4\lambda^2 - 25c^2\varepsilon^2 - 5n\sigma^2c\lambda\varepsilon\right] \\
&= 4(1-\lambda)^2q^2\left[5n\sigma^2\lambda\varepsilon c + n^2\sigma^4\lambda^2\right] \\
&= 4(1-\lambda)^2q^2n^2\sigma^4\lambda^2\left[u+1\right],
\end{aligned}
$$

where $u = \frac{5c\varepsilon}{n\sigma^2\lambda}$.
Hence, the roots of the polynomial of interest are of the form

$$
\begin{aligned}
&\frac{10\varepsilon(1-\lambda)cq + 2n\sigma^2\lambda q(1-\lambda) \pm \sqrt{\Delta}}{2\left[25\varepsilon c^2q^2 + 5n\sigma^2cq^2\lambda\right]} \\
&= \frac{2(1-\lambda)qn\sigma^2\lambda\left[\frac{5\varepsilon c}{n\sigma^2\lambda}+1\right] \pm \sqrt{\Delta}}{10q^2cn\sigma^2\lambda\left[\frac{5\varepsilon c}{n\sigma^2\lambda}+1\right]} \\
&= \frac{1-\lambda}{5cq} \times \frac{u+1 \pm \sqrt{u+1}}{u+1}.
\end{aligned}
$$

We deduce that the polynomial has a root in the interval $[0, \frac{1-\lambda}{5cq})$ which is given by

$$
t^* := \frac{1-\lambda}{5cq} \times \frac{u+1-\sqrt{u+1}}{u+1},
$$

and one can check that this critical point corresponds to a maximum of the function f_ε. We deduce that for any $\varepsilon > 0$,

$$
g_2^*(\varepsilon) = \varepsilon t^* - \frac{n\sigma^2\lambda q(t^*)^2}{1-\lambda-5cqt^*}
$$

$$= t^*\left\{\varepsilon - \frac{n\sigma^2\lambda q\frac{1-\lambda}{5cq}\times\frac{u+1-\sqrt{u+1}}{u+1}}{1-\lambda-5cq\frac{1-\lambda}{5cq}\times\frac{u+1-\sqrt{u+1}}{u+1}}\right\} = t^*\left\{\varepsilon - \frac{n\sigma^2\lambda q\left(u+1-\sqrt{u+1}\right)}{5cq(u+1)-5cq\left(u+1-\sqrt{u+1}\right)}\right\}$$
$$= t^*\left\{\varepsilon - \frac{n\sigma^2\lambda q\left(u+1-\sqrt{u+1}\right)}{5cq\sqrt{u+1}}\right\} = t^*\left\{\varepsilon - \frac{n\sigma^2\lambda}{5c}\times\frac{\left(u+1-\sqrt{u+1}\right)}{\sqrt{u+1}}\right\}$$
$$= t^*\left\{\varepsilon - \frac{\varepsilon\left(u+1-\sqrt{u+1}\right)}{u\sqrt{u+1}}\right\} = t^*\varepsilon\left\{\frac{u\sqrt{u+1}-u-1+\sqrt{u+1}}{u\sqrt{u+1}}\right\}$$
$$= \frac{1-\lambda}{5cq}\varepsilon\times\frac{u+1-\sqrt{u+1}}{u+1}\left\{\frac{u-\sqrt{u+1}+1}{u}\right\}$$
$$= \frac{(1-\lambda)\varepsilon}{q}\frac{1}{5c}\times\left(\sqrt{u+1}-1\right)\left\{\frac{\sqrt{u+1}-1}{u}\right\} = \frac{(1-\lambda)\varepsilon^2}{qn\sigma^2\lambda}\frac{(\sqrt{u+1}-1)^2}{u^2} = \frac{(1-\lambda)\varepsilon^2}{qn\sigma^2\lambda}h_2(u),$$

where $h_2(u) = \frac{(\sqrt{u+1}-1)^2}{u^2}$. However, the function $u \mapsto \sqrt{u+1}$ is analytic on $]0,+\infty[$ and for any $v \in]0,+\infty[$,

$$\sqrt{1+v} = \sum_{k=0}^{\infty}\frac{v^k}{k!}a_k,$$

where $a_0 = 1$ and for all $k \in \mathbb{N}^*$, $a_k = \frac{1}{2}(\frac{1}{2}-1)\dots(\frac{1}{2}-k+1)$. Hence, we have

$$\frac{\sqrt{v+1}-1}{v} = \sum_{k=1}^{\infty}\frac{v^{k-1}}{k!}\frac{1}{2}(\frac{1}{2}-1)\dots(\frac{1}{2}-k+1) = \sum_{k=0}^{\infty}\frac{v^k}{(k+1)!}\frac{1}{2}(\frac{1}{2}-1)\dots(\frac{1}{2}-k)$$
$$= \frac{1}{2}\sum_{k=0}^{\infty}\frac{v^k}{(k+1)!}b_k = \frac{1}{2}\sum_{k=0}^{\infty}\frac{(v/2)^k}{k!}b_k\underbrace{\frac{2^k}{k+1}}_{\geq 1}$$
$$\geq \frac{1}{2}\sum_{k=0}^{\infty}\frac{(v/2)^k}{k!}b_k = \frac{1}{2}(v/2+1)^{-1/2} = \frac{1}{\sqrt{2}}\frac{1}{\sqrt{v+2}},$$

where we have denoted $b_0 = 1$ and for all $k \in \mathbb{N}^*$, $b_k = \left(-\frac{1}{2}\right)\left(-\frac{1}{2}-1\right)\dots\left(-\frac{1}{2}-k+1\right)$. Hence we proved that for any $\varepsilon > 0$,

$$g_2^*(\varepsilon) = \frac{(1-\lambda)\varepsilon^2}{qn\sigma^2\lambda}h_2(u) \geq \frac{(1-\lambda)\varepsilon^2}{2qn\sigma^2\lambda}\times\frac{1}{u+2}\text{ with } u = \frac{5c\varepsilon}{n\sigma^2\lambda}.$$

B.5 Complement for the proof of Theorem 1

In this section, we only provide the part of the proof of Theorem 1 that needs to be modified to get the result when the kernels $h_{i,j}$ depend on both i and j and when Assumption 4.(ii) holds. Keeping the notations of Theorem 1, we only want to bound $(\mathbb{E}[Z])^k$ (and thus a_1) using a different concentration result that can allow to deal with kernel functions $h_{i,j}$ that might depend on both i and j.
Let us recall that

$$(\mathbb{E}[Z])^k \leq \mathbb{E}[Z^k] \quad \text{(Using Jensen's inequality)}$$
$$= \mathbb{E}\left[\left(\sup_{f_\xi\in\mathscr{F}}\sum_{i=1}^{n-1}f_\xi(X_i)\right)^k\right]$$
$$= \mathbb{E}\left[\left(\sup_{f_\xi\in\mathscr{F}}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\mathbb{E}_{j-1}[p_{i,j}(X_i,X_j')\xi_j(X_j')]\right)^k\right]$$
$$= \mathbb{E}\left[\sum_{j=2}^{n}\mathbb{E}_{j-1}\left|\sum_{i=1}^{j-1}p_{i,j}(X_i,X_j')\right|^k\right] \quad \text{(Using Lemma 3)}$$

$$= \mathbb{E}\left[\sum_{j=2}^{n} \mathbb{E}_{|X'}\left|\sum_{i=1}^{j-1} p_{i,j}(X_i, X'_j)\right|^k\right].$$

Thus we have

$$a_1 = \frac{2\delta_M}{1+\varepsilon}\mathbb{E}\sum_{j=2}^{n}\left(\mathbb{E}_{|X'}\left[e^{\alpha(1+\varepsilon)K|C_j|}\right] - \alpha(1+\varepsilon)K\mathbb{E}_{|X'}\left[|C_j|\right] - 1\right),$$

where $C_j = \sum_{i=1}^{j-1} p_{i,j}(X_i, X'_j)$ and where the notation $\mathbb{E}_{|X'}$ refers to the expectation conditionally to the σ-algebra $\sigma(X'_2, \dots, X'_n)$.
Now we use a symmetrization trick: since $e^x - x - 1 \geq 0$ for all x and since $e^{a|x|} + e^{-a|x|} = e^{ax} + e^{-ax}$, adding $\mathbb{E}_{|X'}[\exp\left(-\alpha(1+\varepsilon)K|C_j|\right)] + \alpha(1+\varepsilon)K\mathbb{E}_{|X'}[|C_j|] - 1$ to a_1 gives

$$a_1 \leq \frac{2\delta_M}{1+\varepsilon}\mathbb{E}\sum_{j=2}^{n}\left(\mathbb{E}_{|X'}[e^{\alpha(1+\varepsilon)KC_j}] - 1 + \mathbb{E}_{|X'}[e^{-\alpha(1+\varepsilon)KC_j}] - 1\right). \tag{21}$$

Let us consider some $j \in \{2, \dots, n\}$. Conditionally on $\sigma(X'_2, \dots, X'_n)$, C_j is a sum of bounded functions (by A) depending on the Markov chain. We denote

$$v_j(X'_j) = \sum_{i=1}^{j-1} \mathbb{E}_{X_i\sim\pi}[p_{i,j}^2(X_i, X'_j)|X'_j] \leq B_0^2$$

and $V = \sum_{j=2}^{n} \mathbb{E}v_j^k(X'_j) \leq C^2 B_0^{2(k-1)}$ (with $C^2 = \sum_{j=2}^{n}\sum_{i=1}^{j-1}\mathbb{E}[p_{i,j}^2(X_i, X'_j)]$).
Remark that

$$\begin{aligned}
& \mathbb{E}_{X_i\sim\pi}[p_{i,j}(X_i, X'_j)|X'_j] \\
= \; & \mathbb{E}_{X_i\sim\pi}\left[h_{i,j}(X_i, X'_j) - \mathbb{E}_{\tilde{X}\sim\pi}[h_{i,j}(X_i, \tilde{X})] \mid X'_j\right] \\
= \; & \int_{x'}\left(\int_{x_i}(h_{i,j}(x_i, X'_j) - h_{i,j}(x_i, \tilde{x})])d\pi(x_i)\right)d\pi(\tilde{x}) \\
= \; & 0,
\end{aligned}$$

where the last equality comes from Assumption 3. We use a Bernstein inequality for Markov chain (see Proposition 4 in Section B.4). Notice from Taylor expansion that $(1-p/3)(e^p - p - 1) \leq p^2/2$ for all $p \geq 0$. Applying (18) with $t = \alpha(1+\varepsilon)K$ and $c = A$ (using the notations of the proof of Proposition 4), we get that for $\alpha < [(1+\varepsilon)K\sqrt{q}(A\sqrt{q}/3 + B_0\sqrt{3/2})]^{-1} \wedge [(1-\lambda)^{-1/2}(1+\varepsilon)K\sqrt{q}\left(5A\sqrt{q}(1-\lambda)^{-1/2} + \sqrt{3\lambda}B_0\right)]^{-1}$,

$$\begin{aligned}
& \mathbb{E}_{|X'}[e^{\alpha(1+\varepsilon)K|C_j|}] \\
\leq \; & 2\left\|\frac{d\chi}{d\pi}\right\|_{\pi,p} \times \mathbb{E}_{|X'}\left[\exp\left(\frac{\alpha^2(1+\varepsilon)^2K^2qv_j(X'_j)}{2 - 2Aq\alpha(1+\varepsilon)K/3} + \frac{v_j(X'_j)\lambda\alpha^2(1+\varepsilon)^2K^2q}{1-\lambda-5\alpha(1+\varepsilon)KAq}\right)\right].
\end{aligned}$$

Considering $\alpha < [(1+\varepsilon)K\sqrt{q}(A\sqrt{q}/3+B_0\sqrt{3/2})]^{-1} \wedge [(1-\lambda)^{-1/2}(1+\varepsilon)K\sqrt{q}\left(5A\sqrt{q}(1-\lambda)^{-1/2} + \sqrt{3\lambda}B_0\right)]^{-1}$, $\varepsilon < 1$ and using (21), this leads to

$$\begin{aligned}
\frac{a_1}{2\left\|\frac{d\chi}{d\pi}\right\|_{\pi,p}} &\leq \frac{2\delta_M}{1+\varepsilon}\sum_{j=2}^{n}\mathbb{E}\left[\exp\left(\frac{\alpha^2(1+\varepsilon)^2K^2qv_j(X'_j)}{2-2Aq\alpha(1+\varepsilon)K/3} + \frac{v_j(X'_j)\lambda\alpha^2(1+\varepsilon)^2K^2q}{1-\lambda-5\alpha(1+\varepsilon)KAq}\right) - 1\right] \\
&= \frac{2\delta_M}{1+\varepsilon}\sum_{j=2}^{n}\sum_{k=1}^{\infty}\frac{1}{k!}\left(\frac{\alpha^2(1+\varepsilon)^2K^2qv_j(X'_j)}{2-2Aq\alpha(1+\varepsilon)K/3} + \frac{v_j(X'_j)\lambda\alpha^2(1+\varepsilon)^2K^2q}{1-\lambda-5\alpha(1+\varepsilon)KAq}\right)^k \\
&= \frac{2\delta_M}{1+\varepsilon}\sum_{j=2}^{n}\sum_{k=1}^{\infty}\frac{1}{k!}\left(\frac{3}{2}\right)^{k-1}\left(\frac{\alpha^2(1+\varepsilon)^2K^2qv_j(X'_j)}{2-2Aq\alpha(1+\varepsilon)K/3}\right)^k
\end{aligned}$$

$$
\begin{aligned}
&\quad + \frac{2\delta_M}{1+\varepsilon}\sum_{j=2}^{n}\sum_{k=1}^{\infty}\frac{1}{k!}3^{k-1}\left(\frac{\nu_j(X_j')\lambda\alpha^2(1+\varepsilon)^2K^2q}{1-\lambda-5\alpha(1+\varepsilon)KAq}\right)^k \qquad \text{(Using Lemma 2)}\\
&\leq \frac{\delta_M}{3(1+\varepsilon)}\sum_{k=1}^{\infty}\frac{3^k\alpha^{2k}(1+\varepsilon)^{2k}K^{2k}q^kV}{(4-4Aq\alpha(1+\varepsilon)K/3)^k} + \frac{2\delta_M}{3(1+\varepsilon)}\sum_{k=1}^{\infty}\frac{3^kV\lambda^k\alpha^{2k}(1+\varepsilon)^{2k}K^{2k}q^k}{(1-\lambda-5\alpha(1+\varepsilon)KAq)^k}\\
&\leq \frac{\delta_M}{3(1+\varepsilon)}\sum_{k=1}^{\infty}\frac{3^k\alpha^{2k}(1+\varepsilon)^{2k}K^{2k}q^kC^2B_0^{2(k-1)}}{(2-2Aq\alpha(1+\varepsilon)K/3)^k} + \frac{2\delta_M}{3(1+\varepsilon)}\sum_{k=1}^{\infty}\frac{3^kC^2B_0^{2(k-1)}\lambda^k\alpha^{2k}(1+\varepsilon)^{2k}K^{2k}q^k}{(1-\lambda-5\alpha(1+\varepsilon)KAq)^k}\\
&= \frac{(1+\varepsilon)C^2\alpha^2K^2\delta_Mq}{2-2Aq\alpha(1+\varepsilon)K/3-3\alpha^2(1+\varepsilon)^2K^2B_0^2q} + \frac{2\delta_MC^2\lambda\alpha^2(1+\varepsilon)K^2q}{1-\lambda-5\alpha(1+\varepsilon)KAq-3B_0^2\lambda\alpha^2(1+\varepsilon)^2K^2q}\\
&= \frac{(1+\varepsilon)C^2\alpha^2K^2\delta_Mq/2}{1-Aq\alpha(1+\varepsilon)K/3-3\alpha^2(1+\varepsilon)^2K^2B_0^2q/2}\\
&\quad + \frac{2\delta_MC^2\lambda\alpha^2(1+\varepsilon)K^2q(1-\lambda)^{-1}}{1-5(1-\lambda)^{-1}\alpha(1+\varepsilon)KAq-3B_0^2\lambda(1-\lambda)^{-1}\alpha^2(1+\varepsilon)^2K^2q}\\
&\leq \frac{(1+\varepsilon)C^2\alpha^2K^2\delta_Mq/2}{1-\alpha(1+\varepsilon)K\sqrt{q}(A\sqrt{q}/3+B_0\sqrt{3/2})}\\
&\quad + \frac{2\delta_MC^2\lambda\alpha^2(1+\varepsilon)K^2q(1-\lambda)^{-1}}{1-\alpha(1-\lambda)^{-1/2}(1+\varepsilon)K\sqrt{q}\left(5A\sqrt{q}(1-\lambda)^{-1/2}+\sqrt{3\lambda}B_0\right)}.
\end{aligned}
$$

From this bound on a_1, one can follow exactly the steps of the proof of Theorem 1 to conclude the proof.