Mixing rates for potentials of non-summable variations

CHRISTOPHE GALLESCO† and DANIEL Y. TAKAHASHI‡

† *Departmento de Estatística, Instituto de Matemática,*
Estatística e Ciência de Computação, Universidade de Campinas, Campinas, Brasil
(e-mail: gallesco@unicamp.br)
‡ *Instituto do Cérebro, Universidade Federal do Rio Grande do Norte, Natal, Brasil*
(e-mail: takahashiyd@gmail.com)

Abstract. Mixing rates, relaxation rates, and decay of correlations for dynamics defined by potentials with summable variations are well understood, but little is known for non-summable variations. This paper exhibits upper bounds for these quantities for dynamics defined by potentials with square-summable variations. We obtain these bounds as corollaries of a new block coupling inequality between pairs of dynamics starting with different histories. As applications of our results, we prove a new weak invariance principle and a Hoeffding-type inequality.

1. *Introduction*

Let A be a countable set, called an *alphabet*. Consider a measurable function $\phi : A \times A^{\{-1,-2,\ldots\}} \to \mathbb{R}$ such that $\sum_{a \in A} e^{\phi(a,x)} = 1$ for all $x \in A^{\{-1,-2,\ldots\}}$. The function ϕ is called a *normalized potential*, and the probability kernel $g := e^{\phi}$ (also known as *g-function*) is a natural generalization of Markov kernels. Let $\eta = (\eta_n)_{\mathbb{Z}}$ be the canonical projections on $A^{\mathbb{Z}}$, that is, for all $x \in A^{\mathbb{Z}}$ and all $n \in \mathbb{Z}$, $\eta_n(x) = x_n$. For $y \in A^{\{0,-1,\ldots\}}$, let μ^y be the probability measure on $A^{\mathbb{Z}}$ such that $\mu^y[(\eta_0, \eta_{-1}, \ldots) \in B] = \delta_y[B]$ for all $B \subset A^{\{0,-1,\ldots\}}$ measurable. For $n \geqslant 0$, $\mu^y[\eta_{n+1} = a \mid (\eta_n, \eta_{n-1}, \ldots) = x] = e^{\phi(a,x)}$ for every $a \in A$ and μ^y-a.e. (almost every) x in $A^{\{-1,-2,\ldots\}}$. Let T denote the shift operator on $A^{\{0,-1,\ldots\}}$. We indicate by d_{TV} the total variation distance, that is, if P and Q are two

probability measures on the same σ-algebra $\mathcal{F}$,

$$d_{\mathrm{TV}}(P, Q) = \sup_{F \in \mathcal{F}} |P[F] - Q[F]|.$$

In this paper, we obtain upper bounds, respectively, for the relaxation rate

$$L(n) := \sup_{y,z} d_{\mathrm{TV}}(\mu^y[\eta_n \in \cdot\,], \mu^z[\eta_n \in \cdot\,]),$$

the mixing rate

$$M(n) := \sup_{y,z} d_{\mathrm{TV}}(\mu^y[(\eta_j)_{j \geqslant n} \in \cdot\,], \mu^z[(\eta_j)_{j \geqslant n} \in \cdot\,]),$$

and the decay rate of correlations

$$\rho_{f,\hat{f}}(n) := \left| \int f \circ T^n\, \hat{f}\, d\tilde{\mu} - \int f\, d\tilde{\mu} \int \hat{f}\, d\tilde{\mu} \right|$$

when $\tilde{\mu}$ is the unique shift-invariant measure compatible with ϕ (see the next section for the definition of *compatibility*) and f, $\hat{f}$ are suitable functions (see Theorem 3.6). Bressaud, Fernández and Galves [**BFG99**] and Pollicott [**Pol00**, Proposition 1] obtained upper bounds for $L(n)$, $M(n)$, and $\rho_{f,\hat{f}}(n)$ for potentials of summable variations and finite alphabets. Gouëzel [**Gou04**] obtained sharp lower bounds for the decay of correlations for dynamics with Hölder continuous (that is, exponentially decaying) potentials and countable alphabet. Our contribution is twofold. We obtain upper bounds for $L(n)$, $M(n)$, and $\rho_{f,\hat{f}}(n)$ when the variation rate $\mathrm{var}_k(\phi)$ decays as $\mathcal{O}(k^{-(1/2+\delta')})$ for any $\delta' > 0$. Moreover, our results also hold for a countably infinite alphabet A. Theorem 3.2 is our main result, showing a new upper bound for the coupling error between μ^y and μ^z. Corollary 3.3 answers a question posed in [**JÖ08**], in which the authors ask for a bound for $L(n)$ when the variation of ϕ is not summable. Corollary 3.4 shows a bound for $M(n)$, which cannot be achieved by simply using the union bound and Corollary 3.3. The result is new even for the case of summable variations. The interest in $M(n)$ stands from the fact that it is the natural generalization of mixing times for Markov chains. Gallesco, Gallo and Takahashi [**GGT18**] showed that $M(n)$ converges to 0 only when $\mathrm{var}_k(\phi)$ is square summable $\tilde{\mu}$-a.s. (almost surely) and hence Corollary 3.4 covers the main cases of interest. Theorem 3.6 gives an upper bound for the speed of decay of correlations, extending Theorem 3.2 in [**BFG99**]. Johansson, Öberg and Pollicott [**JOP12**] showed, when the alphabet A is finite, that there is a unique shift-invariant measure $\tilde{\mu}$ compatible with ϕ when $\mathrm{var}_k(\phi) \in \mathcal{O}(k^{-(1/2)})$. Moreover, Berger, Hoffman and Sidoravicius [**BHS18**] proved that whenever $\mathrm{var}_k(\phi) \in \mathcal{O}(k^{-(1/2-\delta)})$ for any $\delta > 0$, there exists a normalized potential ϕ that exhibits multiple compatible shift-invariant measures. Therefore, Theorem 3.6 also covers the main variation rates of interest under uniqueness of the compatible shift-invariant measure. In Corollary 3.9, we use Theorem 3.6 to obtain upper bounds on the rate of correlation decay for non-normalized potentials. We illustrate the application of our inequalities in three cases. The first application proves a novel weak invariance principle for additive functionals of dynamics with non-summable variations. The second application shows that we can obtain Hoeffding-type bounds for averages of random variables when

the variation of ϕ is not summable. The third example illustrates how we can apply our results on a Poisson autoregression model, which is popular in applied works.

The proof technique is based on a renewal equation and coupling inequalities. These ideas were developed in [**BFG99**, **CFF02**, **CQ98**]. We improve on the coupling bounds obtained in [**BFG99**] by using a coupling between blocks of coordinates, instead of one coordinate at a time. A block coupling idea was used in [**JOP12**] to obtain sharp conditions for uniqueness of the equilibrium measure for ϕ on a finite alphabet A, but mixing rate was not obtained. A difference between [**JOP12**] and our approach is that we upper bound the block coupling using different renewal processes leading to a distinct renewal equation. This new renewal equation allows us to upper bound the speed of decay of the coupling inequality even when the variation is not summable (see Theorem 3.2).

2. *Definitions*

Let the alphabet A be a countable set, $\mathcal{X} = A^{\mathbb{Z}}$, and $\mathcal{X}_- = A^{\mathbb{Z}_-}$, where $\mathbb{Z}_- = \{0, -1, -2, \ldots\}$. We endow $\mathcal{X}$ and $\mathcal{X}_-$ with the product topology and its corresponding Borel σ-algebra. The topologies and σ-algebras considered on subsets of $\mathcal{X}$ and $\mathcal{X}_-$ will always be the trace topologies and σ-algebras. We denote by x_i the ith coordinate of $x \in \mathcal{X}$ and, for $-\infty < i \leqslant j < \infty$, we write $x_{-j}^{-i} := (x_{-i}, \ldots, x_{-j})$, $x_{-\infty}^{-i} := (x_{-i}, x_{-i-1}, \ldots)$, and $x_i^\infty := (\ldots, x_{i+1}, x_i)$. If $i < j$, $x_j^i = \phi$. For $x \in \mathcal{X}$ and $y \in \mathcal{X}_-$, a *concatenation* $x_{-i}^0 y$ is a new sequence $z \in \mathcal{X}_-$ with $z_{-i}^0 = x_{-i}^{-1}$ and $z_{-\infty}^{-i-1} = y$. We take ϕ to be the neutral element of the concatenation operation, that is, $\phi x = x$ for all $x \in \mathcal{X}_-$. Note that we are using the convention, consistent with the concatenation operation, that when we scan an element $x \in \mathcal{X}$ from the left to the right we go further into the past.

Consider a measurable function $\phi : \mathcal{X}_- \to \mathbb{R}$, which we call a potential. We say that ϕ is *normalized* if it satisfies

$$\sum_{a \in A} e^{\phi(ax)} = 1$$

for all $x \in \mathcal{X}_-$. To a normalized potential ϕ we can associate a *probability kernel* g on the alphabet A by defining $g = e^\phi$. The *variation* of order $k \geqslant 0$ of ϕ is defined by

$$\mathrm{var}_k(\phi) := \frac{1}{2} \sup_{z \in \mathcal{X}} \sup_{x, y \in \mathcal{X}_-} \sum_{b \in A} |\phi(bz_{-k}^{-1}x) - \phi(bz_{-k}^{-1}y)|.$$

When A is finite, the variation is usually defined by taking the supremum over $b \in A$ instead of the sum. Nevertheless, our definition is more convenient when the alphabet is infinite and has appeared in the literature before [**CGT20**]. The constant $1/2$ in the definition relates $\mathrm{var}_k(e^\phi)$ to total variation distance when ϕ is normalized.

We also define, for $k \geqslant 0$, the χ^2-*variation* of order k of ϕ as

$$\chi_k^2(\phi) = \sup_{z \in \mathcal{X}} \sup_{x, y \in \mathcal{X}_-} \sum_{b \in A} \frac{(e^{\phi(bz_{-k}^{-1}x)} - e^{\phi(bz_{-k}^{-1}y)})^2}{e^{\phi(bz_{-k}^{-1}y)}}.$$

The use of χ^2-variation to measure the regularity of potentials seems to be new; therefore, it is interesting to compare it to variation, which is more standard. When ϕ is normalized, by using the Cauchy–Schwarz inequality, we have that $\mathrm{var}_k^2(e^\phi) \leqslant \frac{1}{4}\chi_k^2(\phi)$

for any $k \geqslant 1$. When the alphabet A is finite and ϕ is normalized, $\mathrm{var}_k^2(\phi)$ and $\chi_k^2(\phi)$ are comparable, that is, there exist positive constants K_1 and K_2 such that $K_1 \, \mathrm{var}_k^2(\phi) \leqslant \chi_k^2(\phi) \leqslant K_2 \, \mathrm{var}_k^2(\phi)$. The χ^2-variation introduced in this work will be particularly useful to study asymptotic properties of positive probability kernels on infinite A (cf. §8.3).

Let $\eta = (\eta_n)_{\mathbb{Z}}$ be the canonical projections on $\mathcal{X}$, that is, for all $x \in \mathcal{X}$, $\eta_n(x) = x_n$ for all $n \in \mathbb{Z}$. We say that a probability measure μ on $\mathcal{X}$ is *compatible* with a normalized potential ϕ if there exists a probability measure P on $\mathcal{X}_-$ such that

$$\mu[\eta_{-\infty}^0 \in B] = P[B]$$

for all $B \subset \mathcal{X}_-$ measurable and if, for $n \geqslant 0$,

$$\mu[\eta_{n+1} = a \mid \eta_{-\infty}^n = x] = e^{\phi(ax)}$$

for every $a \in A$ and μ-a.e. x in $\mathcal{X}_-$. Johansson, Öberg and Pollicott [**JÖP07**] showed that if

$$\sum_{k=0}^{\infty} \sup_{z \in \mathcal{X}} \sup_{x,y \in \mathcal{X}_-} \sum_{b \in A} (e^{\phi(bz_{-k}^{-1}x)/2} - e^{\phi(bz_{-k}^{-1}y)/2})^2 < \infty,$$

then there is at most one shift-invariant invariant compatible measure with ϕ. From [**Rei12**, Lemma 3.3.9], for $k \geqslant 0$, we have

$$\sum_{b \in A} (e^{\phi(bz_{-k}^{-1}x)/2} - e^{\phi(bz_{-k}^{-1}y)/2})^2 \leqslant \sum_{b \in A} \frac{(e^{\phi(bz_{-k}^{-1}x)} - e^{\phi(bz_{-k}^{-1}y)})^2}{e^{\phi(bz_{-k}^{-1}y)}}$$

and hence the summability of $\chi_k^2(\phi)$ implies the existence of at most one shift-invariant invariant compatible measure.

When ϕ is not normalized, the definition of a compatible measure loses its meaning. Nevertheless, we can associate a set of shift-invariant measures called equilibrium states for not necessarily normalized ϕ [**Wal75**]. Equilibrium states are characterized via a variational principle and coincide with shift-invariant compatible measures when ϕ is normalized. An equilibrium state $\tilde{\mu}$ compatible with a normalized ϕ is also called a *g*-measure in ergodic theory [**Kea72**]. In probability literature, *g*-measures are known as chains of complete connections [**DF37**, **IG90**], chains of infinite order [**Har55**, **Kea72**], random-step Markov processes [**Kal90**], and uniform martingales [**Kal90**]. Compatible measures that are not necessarily shift-invariant are called *g*-chains [**JOP12**] or stochastic chains of unbounded memory [**GGT18**]. When there is more than one shift-invariant measure compatible with ϕ, we say that there is a *phase transition*, otherwise we say that the shift-invariant compatible measure is *unique*.

3. *Results*
In this paper, we will work under the following assumption.

Assumption $(\mathcal{A})$. ϕ is a potential on $\mathcal{X}_-$ such that for all $k \geqslant 1$,

$$\chi_k^2(\phi) \leqslant \frac{C}{k^{1+\delta}} \tag{1}$$

for some $C > 0$ and $\delta > 0$.

Remark 3.1. When the alphabet A is finite and ϕ is normalized, Assumption $(\mathcal{A})$ is equivalent to

$$\mathrm{var}_k(\phi) \leqslant \frac{C'}{k^{(1+\delta)/2}}$$

for some $C' > 0$ and the same δ as in (1). Observe that $\mathrm{var}_k(\phi)$ is not summable when $\delta \in (0, 1]$.

Now, consider $\mathcal{X} \times \mathcal{X}$ with the projection maps $\hat{\eta} = (\hat{\eta}_n)_{n \in \mathbb{Z}}$ and $\hat{\omega} = (\hat{\omega}_n)_{n \in \mathbb{Z}}$ such that for $(x, y) \in \mathcal{X} \times \mathcal{X}$, $\hat{\eta}_n(x, y) = x_n$ and $\hat{\omega}_n(x, y) = y_n$ for all $n \in \mathbb{Z}$. Let us also denote by $\widehat{\mathcal{C}}(\phi)$ the set of probability measures P on $\mathcal{X} \times \mathcal{X}$ such that the pushforward measures $\hat{\eta}_* P$ and $\hat{\omega}_* P$ are compatible with ϕ. We also introduce the process $X = (X_n)_{n \geqslant 1}$ such that for all $n \geqslant 1$,

$$X_n = \mathbb{1}\{\text{there exists } j \in [K_n, K_{n+1}), \hat{\eta}_j \neq \hat{\omega}_j\},$$

where $(K_n)_{n \geqslant 1}$ is a fixed strictly increasing sequence of natural numbers such that $K_1 = 1$. Here is our main result followed by two corollaries.

THEOREM 3.2. *Let ϕ be a normalized potential that satisfies Assumption $(\mathcal{A})$. Let $K_n = \lfloor n^\beta \rfloor$ for $\beta \geqslant 1$ and $\beta > 1/\delta$. For all measures μ and ν compatible with ϕ, there exists $\mathbb{P} \in \widehat{\mathcal{C}}(\phi)$ such that $\hat{\eta}_* \mathbb{P} = \mu$, $\hat{\omega}_* \mathbb{P} = \nu$, and, for $n \geqslant 1$,*

$$\mathbb{P}[X_n = 1] \leqslant \frac{C_1}{n^{(\beta\delta+1)/2}},$$

where C_1 is a positive constant depending on C, δ, and β.

COROLLARY 3.3. *Let ϕ be a normalized potential that satisfies Assumption $(\mathcal{A})$. If $\delta > 1$, we have for all $n \geqslant 1$,*

$$L(n) \leqslant \frac{C_2}{n^{(1+\delta)/2}},$$

where C_2 is a positive constant depending on C and δ.

If $\delta \in (0, 1]$, we have for all $n \geqslant 1$ and $\delta' < \delta$,

$$L(n) \leqslant \frac{C_3}{n^{\delta'}},$$

where C_3 is a positive constant that depends on C, δ, and δ'.

COROLLARY 3.4. *Let ϕ be a normalized potential that satisfies Assumption $(\mathcal{A})$. For all $\delta' < \delta$, we have for all $n \geqslant 1$,*

$$M(n) \leqslant \frac{C_4}{n^{\delta'/2}},$$

where C_4 is a positive constant that depends on C, δ, and δ'.

Remark 3.5. When $\delta > 1$ and A is finite, we can use [**BFG99**, Theorem 1] and the union bound to obtain

$$M(n) \leqslant \frac{C_5}{n^{(\delta-1)/2}},$$

where $C_5 > 0$ is a constant that depends on C and δ. Hence, the result in Corollary 3.4 gives a sharper upper bound, even when the potential is summable and the alphabet A is finite.

We now look at the correlations decay for the shift-invariant measure compatible with a potential ϕ. For this, we need the following definitions. Consider the shift operator $T : \mathcal{X}_- \to \mathcal{X}_-$ such that for all $x \in \mathcal{X}_-$, $Tx = Tx_{-\infty}^0 = x_{-\infty}^{-1}$. For non-constant ϕ, let us consider the seminorm

$$\|f\|_\phi = \sup_{k \geqslant 1} \frac{\mathrm{var}_k(f)}{\mathrm{var}_k(e^\phi)}$$

and the subspace of $\mathcal{C}(\mathcal{X}_-, \mathbb{R})$ defined by

$$V_\phi = \{f \in \mathcal{C}(\mathcal{X}_-, \mathbb{R}) : \|f\|_\phi < \infty\}.$$

THEOREM 3.6. *Let ϕ be a normalized potential that satisfies Assumption $(\mathcal{A})$. Assume that a shift-invariant probability measure $\tilde{\mu}$ compatible with ϕ exists. Let $f \in L^1(\tilde{\mu})$ and $\hat{f} \in V_\phi$.*
  If $\delta > 1$, we have for all $n \geqslant 1$,

$$\rho_{f,\hat{f}}(n) \leqslant \frac{C_6}{n^{(1+\delta)/2}} \|f\|_1 \|\hat{f}\|_\phi,$$

where C_6 is a positive constant that depends on C and δ.
  If $\delta \in (0, 1]$, we have for all $n \geqslant 1$ and $\delta' < \delta$,

$$\rho_{f,\hat{f}}(n) \leqslant \frac{C_7}{n^{\delta'}} \|f\|_1 \|\hat{f}\|_\phi,$$

where C_7 is a positive constant that depends on C, δ, and δ'.

Remark 3.7. When $\delta > 1$ and A is finite, Theorem 3.6 recovers the rate obtained in [**BFG99**, Theorem 1].

Remark 3.8. When A is finite, continuity of ϕ guarantees the existence of a compatible shift-invariant measure; therefore, the assumption on the existence of a compatible measure in Theorem 3.6 is redundant. When A is infinite, the existence of a shift-invariant compatible measure is not immediate. Sufficient conditions for existence of shift-invariant compatible measures when A is infinite are given in [**FM05**, **JÖP07**]. See §8.3 for a concrete example. Whenever a shift-invariant compatible measure exists, Assumption $(\mathcal{A})$ implies uniqueness of $\tilde{\mu}$ in Theorem 3.6 [**JÖP07**], although uniqueness is not a priori necessary for Theorem 3.6.

A natural question is whether we can obtain an upper bound for the rate of correlations decay for a potential ϕ that is not normalized. When A is finite, we can use the same strategy as in [**BFG99**, **Pol00**, **Wal75**]. The idea is to study normalized potentials ψ that are cohomologous to ϕ, that is, $\psi = \phi + h - h \circ T + c$ for some $h \in \mathcal{C}(\mathcal{X}_-, \mathbb{R})$ and $c \in \mathbb{R}$. If ϕ and ψ are cohomologous, then both functions have the same associated equilibrium states [**Wal75**]. Hence, properties of equilibrium states for ϕ can be obtained by studying shift-invariant measures compatible with ψ. Walters [**Wal75**] proved that

when the rate of variation of ϕ is summable, there exist a unique h and a unique c such that ψ is a normalized potential. Moreover, from the construction of h in [**Wal75**], we have that $\mathrm{var}_k(h) \leqslant \sum_{j \geqslant k} \mathrm{var}_j(\phi)$. This implies that $\mathrm{var}_k(\psi) \leqslant 3 \sum_{j \geqslant k} \mathrm{var}_j(\phi)$. Using these results, we obtain the following corollary, which improves the results in [**BFG99**, **Pol00**].

COROLLARY 3.9. *Let the alphabet A be finite and ϕ be a potential not necessarily normalized. Assume that there exist a constant $C > 0$ and $\delta > 0$ such that*

$$\mathrm{var}_k(\phi) \leqslant \frac{C}{k^{(3+\delta)/2}}.$$

Let $\tilde{\mu}$ be an equilibrium state for ϕ, $f \in L^1(\tilde{\mu})$, and $\hat{f} \in V_\phi$.
   If $\delta > 1$, we have for all $n \geqslant 1$,

$$\rho_{f,\hat{f}}(n) \leqslant \frac{C_8}{n^{(1+\delta)/2}} \|f\|_1 \|\hat{f}\|_\phi,$$

where C_8 is a positive constant that depends on C and δ.
   If $\delta \in (0, 1]$, we have for all $n \geqslant 1$ and $\delta' < \delta$,

$$\rho_{f,\hat{f}}(n) \leqslant \frac{C_9}{n^{\delta'}} \|f\|_1 \|\hat{f}\|_\phi,$$

where C_9 is a positive constant that depends on C, δ, and δ'.

Remark 3.10. When $\delta > 1$, Corollary 3.9 recovers the rate obtained in [**Pol00**, Theorem 1(1)]. To generalize Corollary 3.9 to an infinite alphabet, we need a result equivalent to [**Wal75**, Theorem 3.3] for an infinite alphabet, which is currently unavailable.

4. *Technical lemmas*
Here we collect some results that we will use to prove Theorem 3.2. We first recall the definitions of the Kullback–Leibler and Pearson χ^2 divergences. Let P and Q be two probabilities on some discrete space $\mathcal{Y}$. We define

$$D_{\mathrm{KL}}(P\|Q) = \sum_{y \in \mathcal{Y}} P(y) \ln \left(\frac{P(y)}{Q(y)} \right)$$

and

$$D_{\chi^2}(P\|Q) = \sum_{y \in \mathcal{Y}} \frac{(P(y) - Q(y))^2}{Q(y)}.$$

It is well known that $D_{\mathrm{KL}}(P\|Q) \leqslant D_{\chi^2}(P\|Q)$ (cf. [**SV16**, eq. 5]).

LEMMA 4.1. *Let $x, y \in \mathcal{X}_-$ and $\mu, \nu \in \mathcal{P}(\phi)$ such that $\mu[\eta^0_{-\infty} \in \cdot] = \delta_x(\cdot)$ and $\nu[\eta^0_{-\infty} \in \cdot] = \delta_y(\cdot)$. For all $n \geqslant 1$, $0 \leqslant k \leqslant n-1$, and all $a, b, c \in \mathcal{X}$, we have*

$$D_{KL}(\mu[\eta^{K_{n+1}-1}_{K_n} \in \cdot \mid \eta^{K_n-1}_1 = a^{K_n-1}_{K_{n-k}} b^{K_{n-k}-1}_1]$$
$$\times \| \nu[\eta^{K_{n+1}-1}_{K_n} \in \cdot \mid \eta^{K_n-1}_1 = a^{K_n-1}_{K_{n-k}} c^{K_{n-k}-1}_1])$$
$$\leqslant \sum_{j=K_n}^{K_{n+1}-1} \chi^2_{j-K_{n-k}}(\phi). \tag{2}$$

Proof. Let us simply denote by D the left-hand term of inequality (2). We have by the chain rule property of the Kullback–Leibler divergence [**CT06**, Theorem 2.5.3] that

$$D = \sum_{i=K_n}^{K_{n+1}-1} D_{KL}(\mu[\eta_i \in \cdot \mid \eta^{i-1}_1 = z^{i-1}_{K_n} a^{K_n-1}_{K_{n-k}} b^{K_{n-k}-1}_1]$$

$$\times \| \nu[\eta_i \in \cdot \mid \eta^{i-1}_1 = z^{i-1}_{K_n} a^{K_n-1}_{K K_{n-k}} c^{K_{n-k}-1}_1])$$

$$=: \sum_{i=K_n}^{K_{n+1}-1} D_i.$$

Then we use the well-known bound

$$D_i \leqslant D_{\chi^2}(\mu[\eta_i \in \cdot \mid \eta^{i-1}_1 = z^{i-1}_{K_n} a^{K_n-1}_{K_{n-k}} b^{K_{n-k}-1}_1]$$
$$\times \| \nu[\eta_i \in \cdot \mid \eta^{i-1}_1 = z^{i-1}_{K_n} a^{K_n-1}_{K_{n-k}} c^{K_{n-k}-1}_1])$$
$$\leqslant \chi^2_{i-K_n}(\phi)$$

to conclude the proof. $\square$

LEMMA 4.2. *For $\alpha > 1$ and $0 < a < b$, we have*

$$\frac{(b+1)^\alpha - a^\alpha}{b^\alpha - a^\alpha} \geqslant \frac{(b+1)^{\alpha-1} - a^{\alpha-1}}{b^{\alpha-1} - a^{\alpha-1}}. \tag{3}$$

Proof. By algebraic computations, we obtain that (3) is equivalent to

$$\left(\frac{b}{a}\right)^{\alpha-1} \geqslant 1 + (b-a)\left(1 - \left(\frac{b}{b+1}\right)^{\alpha-1}\right).$$

This last inequality is obtained from the Bernoulli inequality $(1+x)^r \geqslant 1 + rx$, for $r > 0$ and $x > -1$, observing that

$$\left(\frac{b}{a}\right)^{\alpha-1} = \left(1 + \frac{b-a}{a}\right)^{\alpha-1} \geqslant 1 + (\alpha-1)\frac{b-a}{a}$$

and

$$\left(\frac{b}{b+1}\right)^{\alpha-1} = \left(1 - \frac{1}{b+1}\right)^{\alpha-1} \geqslant 1 - (\alpha-1)\frac{1}{b+1}. \qquad \square$$

Define, for all $\delta > 0$, $\beta \geqslant 1$, $k \geqslant 3$, and $n \geqslant k + 1$,

$$\Delta_k^n := (n^\beta - (n - k)^\beta - 2)^{-\delta} - ((n + 1)^\beta - (n - k)^\beta)^{-\delta}.$$

LEMMA 4.3. *For all $\delta > 0$, $\beta \geqslant 1$, and $k \geqslant 3$, Δ_k^n is a non-increasing function of $n \geqslant k + 1$.*

Proof. The statement of the lemma is trivial for $\beta = 1$. For $\beta > 1$, consider the function $f : [4, \infty) \to \mathbb{R}^+$ defined by

$$f(x) = (x^\beta - (x - k)^\beta - 2)^{-\delta} - ((x + 1)^\beta - (x - k)^\beta)^{-\delta}.$$

In order to prove the result, it is enough to show that the derivative of f is negative. Since

$$\begin{aligned} f'(x) = -\delta\beta[&(x^\beta - (x - k)^\beta - 2)^{-\delta-1}(x^{\beta-1} - (x - k)^{\beta-1}) \\ &- ((x + 1)^\beta - (x - k)^\beta)^{-\delta-1}((x + 1)^{\beta-1} - (x - k)^{\beta-1})], \end{aligned}$$

it is enough to show that

$$\frac{(x + 1)^\beta - (x - k)^\beta}{x^\beta - (x - k)^\beta - 2} \geqslant \frac{(x + 1)^{\beta-1} - (x - k)^{\beta-1}}{x^{\beta-1} - (x - k)^{\beta-1}}.$$

But this last inequality follows from Lemma 4.2. $\qquad\square$

LEMMA 4.4. *For all $\delta > 0$, $\beta \geqslant 1$, and $k \geqslant 3$, we have*

$$\Delta_k^{k+1} \leqslant 4\frac{2^\beta 4^\delta \beta}{k^{\delta\beta+1}}.$$

Proof. Observe that for $k \geqslant 3$,

$$\begin{aligned} \Delta_k^{k+1} &= \int_{(k+1)^\beta - 3}^{(k+2)^\beta - 1} \frac{1}{x^{1+\delta}} dx \\ &\leqslant \frac{(k + 2)^\beta - (k + 1)^\beta + 2}{((k + 1)^\beta - 3)^{1+\delta}} \leqslant \frac{\beta(k + 2)^{\beta-1} + 2}{((k + 1)^\beta - 3)^{1+\delta}} \leqslant 4\frac{2^\beta 4^\delta \beta}{k^{\delta\beta+1}}, \end{aligned} \qquad (4)$$

where, to obtain the second inequality in (4), we used the inequality

$$(a + b)^\alpha \leqslant a^\alpha + \alpha b(a + b)^{\alpha-1}$$

for $\alpha \geqslant 1$ and $a, b \geqslant 0$. This inequality can be obtained using the fundamental theorem of calculus applied to the function $f(x) = x^\alpha$.

To obtain the last inequality in (4), we used that for $k \geqslant 3$,

$$\beta(k + 2)^{\beta-1} + 2 \leqslant 2\beta(k + 2)^{\beta-1} \leqslant 2^\beta \beta k^{\beta-1}$$

and

$$(k + 1)^\beta - 3 = (k + 1)^\beta \left(1 - \frac{3}{(k + 1)^\beta}\right) \geqslant \frac{(k + 1)^\beta}{4} \geqslant \frac{k^\beta}{4}. \qquad\square$$

Finally, we recall the following lemma in [BFG99] (see also Lemma A.4 in [Gia07]) that gives an estimate for the renewal sequence that will appear in the proof of Theorem 3.2. We state the lemma using a notation that is adapted to our purpose.

LEMMA 4.5. (Proposition 2, item (iv) in [BFG99]) *Let $(f_k)_{k \geqslant 1}$ be a sequence of positive real numbers such that $\sum_{k=1}^{\infty} f_k < 1$. Suppose that $(u_k)_{k \geqslant 1}$ is a sequence with $u_0 = 1$ and satisfies the renewal equation*

$$u_n = \sum_{k=1}^{n} f_k u_{n-k}.$$

If $f_n \leqslant c_1 / n^{1+\alpha}$ for some $\alpha > 0$ and a positive constant c_1, then $u_n \leqslant c_2 / n^{1+\alpha}$, where c_2 is a constant that depends on $(f_k)_{k \geqslant 1}$.

5. *Proof of Theorem 3.2*

Let $x, y \in \mathcal{X}_-$ and μ, ν compatible measures such that $\mu[\eta^0_{-\infty} \in \cdot] = \delta_x(\cdot)$ and $\nu[\eta^0_{-\infty} \in \cdot] = \delta_y(\cdot)$. We now construct the coupling of μ and ν, that we call $\mathbb{P}^{x,y}$, as follows. We start by defining

$$\mathbb{P}^{x,y}[\hat{\eta}^0_{-\infty} \in \cdot, \hat{\omega}^0_{-\infty} \in \cdot] = \delta_x \otimes \delta_y.$$

Then, for all $n \geqslant 1$, given the pasts $\hat{\eta}^{K_n-1}_{-\infty}$ and $\hat{\omega}^{K_n-1}_{-\infty}$, we maximally couple $\hat{\eta}^{K_{n+1}-1}_{K_n}$ and $\hat{\omega}^{K_{n+1}-1}_{K_n}$ to complete the construction of $\mathbb{P}^{x,y}$.

Next, we show that $\mathbb{P}^{x,y}$ satisfies the inequality in Theorem 3.2. For all $n \geqslant 1$ and $0 \leqslant k \leqslant n-1$, define

$$q_k^n = \sup_{x,y,a,b \in \mathcal{X}} \mathbb{P}^{x,y}[X_n = 1 \mid X_{n-k}^{n-1} = 0, \hat{\eta}^{K_{n-k}-1}_1 = a_1^{K_{n-k}-1}, \hat{\omega}^{K_{n-k}-1}_1 = b_1^{K_{n-k}-1}]$$

with the convention that if $k > l$, then elements of the form a_k^l are dropped from the conditional part. The shorthand notation $X_{n-k}^{n-1} = 0$ means that $X_{n-k} = 0, \ldots, X_{n-1} = 0$. Observe that for all $n \geqslant 1$ and $0 \leqslant k \leqslant n-2$, we have $q_k^n \geqslant q_{k+1}^n$.

We start by proving the following lemma.

LEMMA 5.1. *Suppose that $(\chi_n^2(\phi))_{j \geqslant 0} \in \ell^1$. Then there exists $\varepsilon > 0$ such that for all $k \geqslant 0$ and all $n \geqslant k + 1$,*

$$q_k^n \leqslant \sqrt{1 - \exp\left(-\sum_{j=0}^{\infty} \chi_j^2(\phi)\right)} \leqslant 1 - \varepsilon. \tag{5}$$

For $k \geqslant 1$ and $n \geqslant k + 1$, we also have

$$q_k^n \leqslant \sqrt{\frac{1}{2} \sum_{j=K_n}^{K_{n+1}-1} \chi^2_{j-K_{n-k}}(\phi)}. \tag{6}$$

Proof. Inequality (5) is a direct consequence of the Bretagnolle–Huber inequality (cf. [SV16, eq. 4]) and Lemma 4.1. Inequality (6) is a direct consequence of the Pinsker inequality (cf. [SV16, eq. 1]) and again Lemma 4.1. □

Now, on some probability space $(\Omega, \mathcal{F}, P)$, consider the random process $Y = (Y_n)_{n \geq 0}$ with values in $\{0, 1\}$ such that $Y_0 = 1$ and, for $n \geq 1$ and $0 \leq k \leq n - 1$,

$$P[Y_n = 1 \mid Y_{n-k}^{n-1} = 0, Y_{n-k-1} = 1, Y_1^{n-k-2}] = q_k^n.$$

For all $m \geq 1$ and $a, b \in \{0, 1\}^m$, we say that $a \geq b$ if $a_i \geq b_i$ for $i \in \{1, \ldots, m\}$. By construction, for all $n \geq 2$, $a, b \in \{0, 1\}^{n-1}$, and $a \geq b$, we have $P[Y_n = 1 | Y_1^{n-1} = a] \geq \mathbb{P}^{x,y}[X_n = 1 | X_1^{n-1} = b]$. Therefore, by applying Strassen's theorem on stochastic domination [**Lin99**] inductively on n, we can construct a coupling measure Q such that, for $a, b \in \{0, 1\}^{n-1}$ and $a \geq b$, we have $Q[Y_n \geq X_n | Y_1^{n-1} = a, X_1^{n-1} = b] = 1$. Therefore, for all $n \geq 1$, we have $Q[Y_n \geq X_n] = 1$, which implies that

$$P[Y_n = 1] \geq \mathbb{P}^{x,y}[X_n = 1] \tag{7}$$

for all $n \geq 1$.

Now, consider the process $Z = (Z_n)_{n \geq 0}$ with values in $\{0, 1\}$ such that $Z_0 = 1$ and, for $n \geq 1$ and $0 \leq k \leq n - 1$,

$$P[Z_n = 1 \mid Z_{n-k}^{n-1} = 0, Z_{n-k-1} = 1, Z_1^{n-k-2}] = b_k := \sup_{n \geq k+1} q_k^n. \tag{8}$$

Observe that, for all $k \geq 0$, we have $b_k \geq b_{k+1}$. Using the same argument used to show (7), we have that $P[Z_n = 1] \geq P[Y_n = 1]$ for all $n \geq 1$. Also, by (6) and Lemmas 4.3 and 4.4, we have that for $k \geq 3$ and $n \geq k + 1$,

$$2(q_k^n)^2 \leq \sum_{j=K_n}^{K_{n+1}-1} \chi_{j-K_{n-k}}^2(\phi)$$

$$\leq C \sum_{j=\lfloor n^\beta \rfloor}^{\lfloor (n+1)^\beta \rfloor - 1} \frac{1}{(j - \lfloor (n-k)^\beta \rfloor)^{1+\delta}} = C \sum_{j=\lfloor n^\beta \rfloor - \lfloor (n-k)^\beta \rfloor}^{\lfloor (n+1)^\beta \rfloor - \lfloor (n-k)^\beta \rfloor - 1} \frac{1}{j^{1+\delta}}$$

$$\leq C \int_{n^\beta - (n-k)^\beta - 2}^{(n+1)^\beta - (n-k)^\beta} \frac{1}{x^{1+\delta}} dx = C \frac{\Delta_k^n}{\delta} \leq C \delta^{-1} \Delta_k^{k+1} \leq 4C \frac{2^\beta 4^\delta \beta \delta^{-1}}{k^{\delta\beta+1}}.$$

Using (5), we obtain that $b_k \leq (2C(2^\beta 4^\delta \beta \delta^{-1})/(k^{\delta\beta+1}))^{1/2} \wedge (1 - \varepsilon)$ for all $k \geq 3$ and $b_k \leq 1 - \varepsilon$ for $0 \leq k \leq 2$.

Next, let $f_i := b_{i-1} \prod_{k=0}^{i-2}(1 - b_k)$ for $i \geq 1$ (with the convention that $\prod_{j=0}^{-1} = 1$) and $u_i := P[Z_i = 1]$ for $i \geq 0$. We have that

$$P[Z_n = 1] = \sum_{k=1}^{n} P[Z_n = 1, Z_{n-k+1}^{n-1} = 0 | Z_{n-k} = 1] P[Z_{n-k} = 1]$$

and hence the following renewal equation holds:

$$u_n = \sum_{k=1}^{n} f_k u_{n-k}.$$

By definition, we have that $\sum_{k=1}^{\infty} f_k = 1 - \prod_{k=1}^{\infty}(1 - b_k)$. If $\beta > \delta^{-1}$, we have $\sum_{k=0}^{\infty} b_k < \infty$. Hence, $\sum_{k=1}^{\infty} f_k < 1$. Moreover, when $\beta > \delta^{-1}$, we have that $b_k \leq$

$c_1 k^{-(\delta\beta+1)/2}$ for some positive constant c_1 that depends on $C, \delta,$ and β. From Lemma 4.5, we have that, for all $n \geqslant 1$,

$$u_n \leqslant \frac{C_1}{n^{(\delta\beta+1)/2}},$$

where C_1 is a positive constant that depends on $C, \delta,$ and β. Because $P[Z_n = 1] \geqslant \mathbb{P}^{x,y}[X_n = 1]$, we obtain that for all $n \geqslant 1$,

$$\mathbb{P}^{x,y}[X_n = 1] \leqslant \frac{C_1}{n^{(\delta\beta+1)/2}}$$

for $\beta > \delta^{-1}$. Because the bound is uniform on $x, y \in \mathcal{X}_-$, we obtain the desired result.

6. *Proofs of Corollaries 3.3 and 3.4*

6.1. *Proof of Corollary 3.3.*   For $k \in [K_n, K_{n+1})$ and all $y, z \in \mathcal{X}_-$, using the coupling inequality for total variation distance (cf. [**Tho00**]), we have

$$d_{\mathrm{TV}}(\mu^y[\eta_k \in \cdot], \mu^z[\eta_k \in \cdot]) \leqslant \mathbb{P}[\hat{\eta}_k \neq \hat{\omega}_k] \leqslant \mathbb{P}[X_n = 1].$$

Then, by Theorem 3.2, we obtain

$$\mathbb{P}[\hat{\eta}_k \neq \hat{\omega}_k] \leqslant \frac{C_1}{n^{(\beta\delta+1)/2}}$$

for all $\beta \geqslant 1$ and $\beta > \delta^{-1}$. If $\delta > 1$, just take $\beta = 1$. In this case $k = n$ and thus we obtain Corollary 3.3 with a constant C_2 that depends on C and δ. If $\delta \in (0, 1)$, since $k \leqslant (n+1)^\beta$, we have $n \geqslant k^{1/\beta} - 1$. This leads to

$$\mathbb{P}[\hat{\eta}_k \neq \hat{\omega}_k] \leqslant \frac{C_3}{k^{(\beta\delta+1)/2\beta}}$$

for all $k \geqslant 1$, where C_3 is a positive constant that depends on $C, \delta,$ and β. Now, observe that for any $0 < \delta' < \delta$, we can choose β such that $\beta \geqslant 1, \beta > \delta^{-1},$ and $(\beta\delta + 1)/2\beta \geqslant \delta'$.

6.2. *Proof of Corollary 3.4.*   Consider $k \in [K_n, K_{n+1})$. Let

$$\theta = \inf\{n \geqslant 1 : \hat{\eta}_k = \hat{\omega}_k \text{ for all } k \geqslant n\}$$

with the convention that $\inf \emptyset = \infty$. We start by observing that

$$\mathbb{P}[\theta > k] \leqslant \mathbb{P}\left[\bigcup_{j \geqslant n}\{X_j = 1\}\right] \leqslant \sum_{j \geqslant n}\mathbb{P}[X_j = 1].$$

By Theorem 3.2, we obtain

$$\mathbb{P}[\theta > k] \leqslant C_1 \sum_{j \geqslant n}\frac{1}{n^{(\beta\delta+1)/2}} \leqslant \frac{C_1'}{n^{(\beta\delta-1)/2}}$$

for $\beta \geqslant 1, \beta > \delta^{-1},$ and C_1' a positive constant that depends on $C, \delta,$ and β. Since $n \geqslant k^{1/\beta} - 1$, we obtain

$$\mathbb{P}[\theta > k] \leqslant \frac{C_4}{k^{(\beta\delta-1)/2\beta}}$$

for all $k \geqslant 1$ and C_4 a positive constant that depends on C, δ, and β. Finally, notice that for all $\delta' < \delta$, we can choose β large enough such that $(\beta\delta - 1)/2\beta \geqslant \delta'$. Using the coupling inequality (cf. [**Tho00**]), we conclude that

$$M(n) := \sup_{y,z} d_{\mathrm{TV}}(\mu^y[(\eta_j)_{j \geqslant n} \in \cdot\,], \mu^z[(\eta_j)_{j \geqslant n} \in \cdot\,]) \leqslant \mathbb{P}[\theta > k].$$

7. *Proof of Theorem 3.6*

Consider $K_m = \lfloor m^\beta \rfloor$ for $m \geq 1$. For each $x, y \in \mathcal{X}_-$, we consider a probability space $(\Omega, \mathcal{F}, \mathbb{P}^{x,y})$ that supports the random elements $\tilde{\eta}$, $\tilde{\omega}$, and $\tilde{Z}$ defined as follows. Let $\tilde{\eta}_* \mathbb{P}^{x,y}$, $\tilde{\omega}_* \mathbb{P}^{x,y}$ be compatible with ϕ and $\tilde{\eta}^0_{-\infty} = x$, $\tilde{\omega}^0_{-\infty} = y$. Also, for all $m \geqslant 1$ given the pasts $\tilde{\eta}^{K_m-1}_{-\infty}$ and $\tilde{\omega}^{K_m-1}_{-\infty}$, the blocks $\tilde{\eta}^{K_{m+1}-1}_{K_m}$ and $\tilde{\omega}^{K_{m+1}-1}_{K_m}$ are maximally coupled. Under $\mathbb{P}^{x,y}$, the process $\tilde{Z}$ has the same law as the process Z defined in §5 and verifies $\tilde{Z}_m \geqslant \tilde{X}_m :=$ $\mathbb{1}\{$there exists $j \in [K_m, K_{m+1}), \tilde{\eta}_j \neq \tilde{\omega}_j\}$ for all $m \geqslant 1$ (this is indeed possible since Z stochastically dominates X; see §5). We denote by $\mathbb{E}^{x,y}$ the expectation with respect to $\mathbb{P}^{x,y}$.

Fix some $n \in \mathbb{N}$ and let k be such that $n \in [K_{k-1}, K_k)$. We will show that

$$\left| \int f \circ T^n \, \hat{f} \, d\tilde{\mu} - \int f \, d\tilde{\mu} \int \hat{f} \, d\tilde{\mu} \right| \leqslant c_1 P[Z_k = 1] \tag{9}$$

for some positive constant c_1 that depends only on C, δ, and β. From this point, Theorem 3.6 is easily obtained following the proof of Corollary 3.3. To obtain (9), we follow the argument developed in [**BFG99**, §5]. Using (3.7) in [**BFG99**], we first observe that

$$\left| \int f \circ T^n \, \hat{f} \, d\tilde{\mu} - \int f \, d\tilde{\mu} \int \hat{f} \, d\tilde{\mu} \right| \leqslant \|f\|_1 \sup_{x,y} \mathbb{E}^{x,y}[|\hat{f}(\tilde{\eta}^n_{-\infty}) - \hat{f}(\tilde{\omega}^n_{-\infty})|].$$

For $k \geqslant 1$, let

$$\theta_k = \inf\{0 \leqslant m \leqslant k : \tilde{Z}_{k-m} = 1\}.$$

We have

$$\mathbb{E}^{x,y}[|\hat{f}(\tilde{\eta}^n_{-\infty}) - \hat{f}(\tilde{\omega}^n_{-\infty})|] = \mathbb{E}^{x,y}\left[\sum_{j=0}^{k} \mathbb{1}\{\theta_k = j\} |\hat{f}(\tilde{\eta}^n_{-\infty}) - \hat{f}(\tilde{\omega}^n_{-\infty})| \right]$$

$$\leqslant \|\hat{f}\|_\phi \sum_{j=0}^{k} \mathrm{var}_{n-K_{k-j}+1}(e^\phi) \mathbb{P}^{x,y}[\theta_k = j]. \tag{10}$$

Observe that for all $0 \leqslant j \leqslant k$,

$$\mathbb{P}^{x,y}[\theta_k = j] = P[Z_k = 0, \ldots, Z_{k-j+1} = 0, Z_{k-j} = 1]$$

$$= \prod_{l=1}^{j} (1 - b_{k-j+l}) P[Z_{k-j} = 1],$$

where the b_{k-j+l} are from (8). Now, observe that for all $i \geqslant 1$, we have

$$
\begin{aligned}
P[Z_i = 1] &= \sum_{k=1}^{i} b_{i-k} P[Z_{i-1} = 0 \mid Z_k^{i-1} = 0, Z_{k-1} = 1] \\
&= \sum_{k=1}^{i} b_{i-k} \prod_{l=1}^{i-k} (1 - b_{i-k-l}) P[Z_{k-1} = 1].
\end{aligned}
$$

Thus, we have that for all $i \geqslant 1$,

$$
P[Z_i = 1] = \sum_{k=1}^{i} f_k P[Z_{i-k} = 1]
$$

with $f_k := b_{k-1} \prod_{l=0}^{k-2} (1 - b_l)$, $k \geqslant 1$. From this, we obtain

$$
\begin{aligned}
\mathbb{E}^{x,y}[|\hat{f}(\tilde{\eta}_{-\infty}^n) - \hat{f}(\tilde{\omega}_{-\infty}^n)|] \leqslant \|\hat{f}\|_\phi \Bigg(&\mathrm{var}_{n-K_k+1}(e^\phi) \sum_{l=1}^{k} f_l P[Z_{k-l} = 1] \\
&+ \sum_{l=1}^{k} \mathrm{var}_{n-K_{k-l}+1}(e^\phi) \prod_{m=1}^{l} (1 - b_{k-l+m}) P[Z_{k-l} = 1] \Bigg).
\end{aligned}
$$

We deduce that

$$
\sup_{x,y} \mathbb{E}^{x,y}[|\hat{f}(\tilde{\eta}_{-\infty}^n) - \hat{f}(\tilde{\omega}_{-\infty}^n)|] \leqslant \kappa \sum_{l=1}^{k} f_l P[Z_{k-l} = 1] = \kappa P[Z_k = 1]
$$

with

$$
\kappa := \mathrm{var}_{n-K_k+1}(e^\phi) + \sup_{1 \leqslant l \leqslant k} \frac{\mathrm{var}_{n-K_{k-l}+1}(e^\phi)}{f_l}.
$$

Finally, since $(\mathrm{var}_{n-K_{k-l}+1}(e^\phi))/(b_{l-1}) \leqslant 2$ and $\prod_{j=0}^{\infty}(1 - b_j) > 0$ (using that $b_0 < 1$ and $\sum_{j=0}^{\infty} b_j < \infty$), we observe that

$$
\kappa \leqslant 1 + \frac{\mathrm{var}_{n-K_{k-l}+1}(e^\phi)}{b_{l-1} \prod_{j=0}^{\infty}(1 - b_j)} \leqslant c_2
$$

for some positive constant c_2 depending on C, δ, and β.

7.1. *Proof of Corollary 3.9.* The potential ϕ is summable; therefore, there exists a normalized potential ψ with the same unique equilibrium state as ϕ [**Wal75**, Theorem 3.2]. If $\mathrm{var}_k(\phi) \leqslant C/k^{(3+\delta)/2}$, then $\mathrm{var}_k(\psi) \leqslant C'/k^{(1+\delta)/2}$ for some constant $C' > 0$ that depends only on C (see, for example, [**Pol00**, Proposition 1]). Closely following the proof of Theorem 3.6 using the measures compatible with ψ, we obtain the desired results. The only difference is that in (10) we use the bound

$$
\mathbb{E}^{x,y}[|\hat{f}(\tilde{\eta}_{-\infty}^n) - \hat{f}(\tilde{\omega}_{-\infty}^n)|] \leqslant \|\hat{f}\|_\phi \sum_{j=0}^{k} \mathrm{var}_{n-K_{k-j}+1}(e^\phi) \mathbb{P}^{x,y}[\theta_k = j]
$$

instead of

$$\mathbb{E}^{x,y}[|\hat{f}(\tilde{\eta}^n_{-\infty}) - \hat{f}(\tilde{\omega}^n_{-\infty})|] \leqslant \|\hat{f}\|_{\psi} \sum_{j=0}^{k} \mathrm{var}_{n-K_{k-j}+1}(e^{\psi})\mathbb{P}^{x,y}[\theta_k = j].$$

8. *Applications*

8.1. *Functional central limit theorem (FCLT) for potentials with non-summable variations.*
Let $\sigma > 0$. A function $h : A \to \mathbb{R}$ satisfies the functional central limit theorem (FCLT), also called the weak invariance principle, when the process $\{\zeta_n(t), t \in [0, 1], n \geqslant 1\}$ defined by

$$\zeta_n(t) = \frac{1}{\sigma\sqrt{n}} \sum_{i=0}^{\lfloor nt \rfloor} h \circ \eta_i$$

converges weakly to a standard Brownian motion on $D[0, 1]$. Tyran-Kamińska [**TK05**, Section 4.3] showed that the FCLT holds when the potential ϕ has summable variations. A straightforward application of Theorem 3.2 in [**TK05**] and our Corollary 3.3 is the following FCLT for potentials with non-summable variations.

PROPOSITION 8.1. *Assume that the alphabet A is finite and ϕ satisfies Assumption $(\mathcal{A})$ with $\delta \in (1/2, 1]$. Let μ be shift-invariant and compatible with ϕ. Also, let $h : A \to \mathbb{R}$ be a function such that $\int h \circ \eta_0 \, d\mu = 0$. If $\sigma^2 := \int (h \circ \eta_0)^2 \, d\mu > 0$, then h satisfies the FCLT.*

Proof. Because the alphabet is finite, we have that $\sigma^2 \leqslant \|h\|_{\infty}^2 < \infty$, as required by Theorem 3.2 in [**TK05**]. It remains to verify the condition on the mixing rate. Using the processes $\{\tilde{\eta}_i, i \in \mathbb{Z}\}$ and $\{\tilde{\omega}_i, i \in \mathbb{Z}\}$ introduced in the proof of Theorem 3.6, it is sufficient to check that there exists $\gamma > 1/2$ such that

$$\limsup_{n \to \infty} n^{\gamma} \sup_{x,y} \mathbb{E}^{x,y}[|h(\tilde{\eta}_n) - h(\tilde{\omega}_n)|] < \infty. \tag{11}$$

We have from Corollary 3.3 that, for all $x, y \in \mathcal{X}_-$,

$$\mathbb{E}^{x,y}[|h(\tilde{\eta}_n) - h(\tilde{\omega}_n)|] \leqslant 2\|h\|_{\infty}\mathbb{P}^{x,y}[\tilde{\eta}_n \neq \tilde{\omega}_n]$$
$$\leqslant \frac{c_1}{n^{\delta'}},$$

where $\delta' < \delta$ and c_1 is a positive constant that depends on C, δ, δ', and h. Taking δ' and γ such that $1/2 < \gamma \leqslant \delta' < \delta$, we obtain (11). $\square$

8.2. *Hoeffding-type inequality for potentials with non-summable variations.*
A Hoeffding-type inequality gives finite sample bounds for deviations of additive functionals from their mean. When the variation rate of the potential is summable, we have an exponential inequality [**GGT14**, **CGT20**, **Mar98**]. Nevertheless, the rate of concentration for potentials when the variation rate is not summable is an open question. Using our result, we can obtain the following stretched exponential inequality for sums of random variables.

PROPOSITION 8.2. *Assume that the alphabet A is finite and ϕ satisfies Assumption ($\mathcal{A}$) with $\delta \in (1/2, 1]$. Let μ be compatible with ϕ. For all $\delta' < 2\delta - 1$, $n \geqslant 1$, $t \geqslant 0$, and all functions $h : A \to \mathbb{R}$, we have*

$$\mu\left[\left|\frac{1}{n}\sum_{i=1}^{n}(h(\eta_i) - \mathbb{E}[h(\eta_i)])\right| \geqslant t\right] \leqslant 2\exp\left\{-\frac{C_{10}n^{\delta'}t^2}{R(h)^2}\right\},$$

where $R(h) := \max_{a \in A} h(a) - \min_{a \in A} h(a)$ and C_{10} is a constant that depends on C, δ, and δ'.

Proof. This is a consequence of Theorem 3.2 of [**CCKR07**] and Corollary 3.3. In order to apply Theorem 3.2 of [**CCKR07**], we need to estimate the terms $\|\overline{D}\|^2_{\ell^2(\mathbb{N})}$ and $\|\delta f\|^2_{\ell^2(\mathbb{N})}$ (for $f(x_1, \ldots, x_n) = (1/n)\sum_{i=1}^{n} h(x_i)$) there. We have

$$\|\overline{D}\|^2_{\ell^2(\mathbb{N})} \leqslant \left(1 + \sum_{i=1}^{n} \sup_{x,y \in \mathcal{X}_-} \mathbb{P}^{x,y}[\hat{\eta}_i \neq \hat{\omega}_i]\right)^2.$$

Now, for $\delta' < 2\delta - 1$, using Corollary 3.3, we obtain

$$\|\overline{D}\|^2_{\ell^2(\mathbb{N})} \leqslant c_1 n^{1-\delta'} \tag{12}$$

for some positive constant c_1 depending on C, δ, and δ'.

For a given function $f : A^n \to \mathbb{R}$, we define the oscillation of f at site $i \in \{1, \ldots, n\}$ by

$$\delta_i f := \sup_{x_j = x'_j, j \neq i} |f(x_1, \ldots, x_n) - f(x'_1, \ldots, x'_n)|.$$

Now, taking $f(x_1, \ldots, x_n) = (1/n)\sum_{i=1}^{n} h(x_i)$, we have $\delta_i f = R(h)/n$ for $i \in \{1, \ldots, n\}$. Thus, we obtain

$$\|\delta f\|^2_{\ell^2(\mathbb{N})} = \sum_{i=1}^{n}\left(\frac{R(h)}{n}\right)^2 = \frac{R(h)^2}{n}. \tag{13}$$

Finally, using (12) and (13) in Theorem 3.2 of [**CCKR07**], we obtain Proposition 8.2. □

8.3. *Poisson autoregression model.*   As a second application of our results, we consider a model with countable infinite alphabet called Poisson autoregression, which is popular in applications [**KF05**]. Only the Markovian case of these models were studied in the literature. We will show how we can choose the parameters of non-Markovian Poisson autoregression models to satisfy the Assumption ($\mathcal{A}$) and thus apply the results of §3.

Consider an absolutely converging sequence $(\beta_i)_{i \geqslant 1}$ and a sequence of non-negative integers $(\gamma_i)_{i \geqslant 1}$ such that $S := \sum_{i=1}^{\infty} |\beta_i|\gamma_i < \infty$. Consider $A = \mathbb{Z}_+$ and the potential ϕ defined for all $x \in \mathcal{X}_-$ by

$$\phi(x) = -\lambda(x_{-\infty}^{-1}) + x_0 \log \lambda(x_{-\infty}^{-1}) - \sum_{k=0}^{x_0} \log(k),$$

where

$$\lambda(x_{-\infty}^{-1}) = \exp\left\{ \sum_{i=1}^{\infty} \beta_i(x_{-i} \wedge \gamma_i) \right\}.$$

For this model, we obtain

$$\chi_k^2(\phi) = \sup_{a \in \mathcal{X}} \sup_{x,y \in \mathcal{X}_-} (e^{\lambda(a_{-k}^{-1}y)(\lambda(a_{-k}^{-1}x)/(\lambda(a_{-k}^{-1}y))-1)^2} - 1). \tag{14}$$

Now, since $e^{-S} \leqslant \lambda(x_{-\infty}^{-1}) \leqslant e^S$ and the exponential function is locally bi-lipschitz, using (14), we have that

$$c_1^{-1}\left(\sum_{i=k+1}^{\infty} |\beta_i|\gamma_i \right)^2 \leqslant \chi_k^2(\phi) \leqslant c_1\left(\sum_{i=k+1}^{\infty} |\beta_i|\gamma_i \right)^2,$$

where c_1 is a positive constant that depends only on S.

Finally, choosing the sequences $(\beta_i)_{i \geqslant 1}$ and $(\gamma_i)_{i \geqslant 1}$ such that

$$\frac{c_2^{-1}}{i^{(3+\varepsilon)/2}} \leqslant |\beta_i|\gamma_i \leqslant \frac{c_2}{i^{(3+\varepsilon)/2}}$$

for some $\varepsilon > 0$ and $c_2 \geqslant 1$, we obtain

$$\frac{c_3^{-1}}{k^{1+\varepsilon}} \leqslant \chi_k^2(\phi) \leqslant \frac{c_3}{k^{1+\varepsilon}},$$

where c_3 is a positive constant.

Finally, for this model, we mention that the existence of a shift-invariant probability measure compatible with ϕ is obtained by applying Theorem 5.1 of [JÖP07] with $K = e^{2\sinh S}$ and π equal to the Poisson law with parameter e^S. Assumption ($\mathcal{A}$) implies the square summability of the variation, which guarantees the uniqueness of the shift-invariant probability measure [JÖP07, Corollary 4.2].

REFERENCES

[BFG99]   X. Bressaud, R. Fernández and A. Galves. Decay of correlations for non-Hölderian dynamics. A coupling approach. *Electron. J. Probab.* **4**(3) (1999), 19 pages (electronic).
[BHS18]   N. Berger, C. Hoffman and V. Sidoravicius. Non-uniqueness for specifications in $\ell^{2+\epsilon}$. *Ergod. Th. & Dynam. Sys.* **38**(4) (2018), 1342–1352.
[CCKR07] J.-R. Chazottes, P. Collet, C. Külske and F. Redig. Concentration inequalities for random fields via coupling. *Probab. Theory Related Fields* **137**(1–2) (2007), 201–225.
[CFF02]   F. Comets, R. Fernández and P. A. Ferrari. Processes with long memory: regenerative construction and perfect simulation. *Ann. Appl. Probab.* **12**(3) (2002), 921–943.

[CGT20]   J.-R. Chazottes, S. Gallo and D. Y. Takahashi. Optimal Gaussian concentration bounds for stochastic chains of unbounded memory. *Preprint*, 2020, arXiv:2001.06633.

[CQ98]   Z. Coelho and A. Quas. Criteria for $\bar{d}$-continuity. *Trans. Amer. Math. Soc.* **350**(8) (1998), 3257–3268.

[CT06]   T. M. Cover and J. A. Thomas. *Elements of Information Theory*, 2nd edn. Wiley, Hoboken, NJ, 2006.

[DF37]   W. Doeblin and R. Fortet. Sur des chaînes à liaisons complètes. *Bull. Soc. Math. France* **65** (1937), 132–148.

[FM05]   R. Fernández and G. Maillard. Chains with complete connections: general theory, uniqueness, loss of memory and mixing properties. *J. Stat. Phys.* **118**(3–4) (2005), 555–588.

[GGT14]   C. Gallesco, S. Gallo and D. Y. Takahashi. Explicit estimates in the Bramson–Kalikow model. *Nonlinearity* **27**(9) (2014), 2281–2296.

[GGT18]   C. Gallesco, S. Gallo and D. Y. Takahashi. Dynamic uniqueness for stochastic chains with unbounded memory. *Stochastic Process. Appl.* **128**(2) (2018), 689–706.

[Gia07]   G. Giacomin. *Random Polymer Models*. Imperial College Press, Singapore, 2007.

[Gou04]   S. Gouëzel. Sharp polynomial estimates for the decay of correlations. *Israel J. Math.* **139**(1) (2004), 29–65.

[Har55]   T. E. Harris. On chains of infinite order. *Pacific J. Math.* **5** (1955), 707–724.

[IG90]   M. Iosifescu and Ş. Grigorescu. *Dependence with Complete Connections and Its Applications (Cambridge Tracts in Mathematics, 96)*. Cambridge University Press, Cambridge, 1990.

[JÖ08]   A. Johansson and A. Öberg. Square summability of variations and convergence of the transfer operator. *Ergod. Th. & Dynam. Sys.* **28**(4) (2008), 1145–1151.

[JÖP07]   A. Johansson, A. Öberg and M. Pollicott. Countable state shifts and uniqueness of *g*-measures. *Amer. J. Math.* **129**(6) (2007), 1501–1511.

[JOP12]   A. Johansson, A. Öberg and M. Pollicott. Unique Bernoulli *g*-measures. *J. Eur. Math. Soc.* **14**(5) (2012), 1599–1615.

[Kal90]   S. Kalikow. Random Markov processes and uniform martingales. *Israel J. Math.* **71**(1) (1990), 33–54.

[Kea72]   M. Keane. Strongly mixing *g*-measures. *Invent. Math.* **16**(4) (1972), 309–324.

[KF05]   B. Kedem and K. Fokianos. *Regression Models for Time Series Analysis (Wiley Series in Probability and Statistics, 488)*. John Wiley & Sons, New Jersey, 2005.

[Lin99]   T. Lindvall. On Strassen's theorem on stochastic domination. *Electron. Commun. Probab.* **4** (1999), 51–59.

[Mar98]   K. Marton. Measure concentration for a class of random processes. *Probab. Theory Related Fields* **110**(3) (1998), 427–439.

[Pol00]   M. Pollicott. Rates of mixing for potentials of summable variation. *Trans. Amer. Math. Soc.* **352**(2) (2000), 843–853.

[Rei12]   R.-D. Reiss. *Approximate Distributions of Order Statistics: With Applications to Nonparametric Statistics*. Springer Science & Business Media, Springer-Verlag, New York, 2012.

[SV16]   I. Sason and S. Verdú. *f*-divergence inequalities. *IEEE Trans. Inform. Theory* **62**(11) (2016), 5973–6006.

[Tho00]   H. Thorisson. *Coupling, Stationarity, and Regeneration (Probability and Its Applications)*. Springer, New York, 2000.

[TK05]   M. Tyran-Kamińska. An invariance principle for maps with polynomial decay of correlations. *Comm. Math. Phys.* **260**(1) (2005), 1–15.

[Wal75]   P. Walters. Ruelle's operator theorem and *g*-measures. *Trans. Amer. Math. Soc.* **214** (1975), 375–387.