

On asymptotics for Vaserstein coupling of a Markov chain

O. A. Butkovsky*, A. Yu. Veretennikov†

October 21, 2011

Abstract

We prove that strong ergodicity of a Markov process is linked with a spectral radius of a certain “associated” semigroup operator, although, not a “natural” one. We also give sufficient conditions for weak ergodicity and provide explicit estimates of the convergence rate. To establish these results we construct a modification of the Vaserstein coupling. Some applications including mixing properties are also discussed.

1 Introduction

A general question about equilibrium distributions for homogeneous Markov processes may be posed as follows. If a Markov process converges to the stationary distribution, then how fast is convergence? In this paper we focus on quantitative estimates on the convergence rate in the total variation metric.

Recall that the total variation distance between two probability measures μ and ν on a measurable space (E, \mathcal{E}) is defined by

$$\|\mu - \nu\|_{TV} := 2 \sup_{A \in \mathcal{E}} |\mu(A) - \nu(A)|.$$

Let X_n , $n \in \mathbb{Z}_+$ be a Markov process on (E, \mathcal{E}) and assume it has a stationary measure π . Consider the measure $\mu_n^x(A) := P_x(X_n \in A)$, $x \in E$, $A \in \mathcal{E}$. The process X_n is called strongly ergodic if there exist $C > 0$, $\lambda > 0$ such that

$$\sup_{x \in E} \|\mu_n^x - \pi\|_{TV} \leq Ce^{-\lambda n}. \quad (1.1)$$

The process X_n is called weakly ergodic if for all $x \in E$ we have $\|\mu_n^x - \pi\|_{TV} \rightarrow 0$ as $n \rightarrow \infty$.

One possible approach to estimate the constant λ in (1.1) was introduced by Diaconis and Stroock [2]. They have shown that if an irreducible finite state-space Markov chain is reversible, then this Markov chain is strongly ergodic with $\lambda > \ln \text{gap}(\mathcal{P})$, where \mathcal{P} is the

*Dept. of Mathematics and Mechanics, Moscow State University, Moscow, 119992, Russia, email: oleg.butkovskiy@gmail.com; the work was partially supported by RFBR grant 10-01-00397-a.

†School of Mathematics, University of Leeds, LS2 9JT, Leeds, UK, & Institute of Information Transmission Problems, Moscow, 127994, GSP-4, Russia; email: a.veretennikov@leeds.ac.uk.

transition probability matrix and $\text{gap}(\mathcal{P})$ denotes the largest out of the eigenvalue moduli which are strictly less than one. (For positive operators this value is called spectral gap; notice that for reversible Markov chain its transition matrix is self-adjoint.)

In this paper a more general situation is considered, which also includes non-reversible processes. To estimate the convergence rate in strongly ergodic as well as in weakly ergodic cases we use the coupling method.

This method dates back to Doeblin [4] and was later developed by Doob [5], Vaserstein [19], Pitman [15], Griffeath [7], Nummelin [14] and many others. Pitman and Griffeath proposed different constructions of *maximal coupling*, in which the probability of coupling is as big as possible at any time. However, as shown in [7], any construction of maximal coupling can not be Markovian, which significantly complicates a convergence rate estimation. On the other hand, Vaserstein [19] proposed a “maximal” Markovian coupling. This coupling is “maximal” in the sense that the probability of coupling in one step is as big as possible.

We modify the Vaserstein’s construction and adapt his ideas to a more general state-space. This enables us to find a new sufficient condition which guarantees strong ergodicity of a Markov process. A natural name for it seems to be *a local Markov–Dobrushin’s condition*, because it localizes Dobrushin’s ergodic coefficient from [3] and, in turn, this coefficient considered in time–homogeneous case is based on the construction proposed by Markov himself [12, 13].

It turned out that the decay rate λ is related to a spectral radius of a certain semigroup operator. A further modification of the coupling technique allows to deal with weakly ergodic case and obtain estimates which depend on the initial state of a Markov process. We consider both exponential and polynomial convergence.

Notice also that in this paper we establish only upper bounds of the rate of convergence. For the corresponding lower bounds, see, e.g., [21].

We ought to comment about relation to some other well-known coupling methods, see, for example, [1, 11, 14, 16], etc. Most of them are based on so called generalized regeneration. Our approach is not a generalized regeneration.

Recurrence assumptions used in this paper are formulated in terms of moments of stopping times. To the authors’ point of view, this is an adequate alternative to conditions in terms of Lyapunov functions, although, they are, clearly, close enough. The latter approach is widely known in the literature, but will not be used in this paper. So, we give only a minimal number of references about it, see [9, 10, 18].

The rest of the paper is organized as follows. We describe a modification of the Vaserstein construction in Section 2.1. The main results on convergence, including both strong and weak ergodicity, are formulated in Section 2.2. Section 3 contains some examples and applications, including estimation of mixing coefficients and central limit theorem. All proofs are placed in Section 4.

2 Construction of Coupling and Main Results

2.1 Coupling

The Vaserstein’s coupling construction was proposed in [19]. It provides a coupling for two Markov chains which have a countable state space. This chapter contains two main lemmas which allow to construct a Vaserstein-type coupling for two homogeneous Markov

processes. The lemmas are crucial in establishing a new representation for the convergence rate of a Markov process to its stationary regime and will be used in the sequel.

Throughout this paper we assume that $(X_n^1)_{n \in \mathbb{Z}_+}$ and $(X_n^2)_{n \in \mathbb{Z}_+}$ are homogeneous Markov processes on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with the same transition measures. Furthermore, we suppose that the transition measure has a density $p(u, v)$ with respect to some non-negative sigma-finite measure Λ , i.e.

$$\mathbb{P}(X_{n+1}^1 \in B | X_n^1 = u) = \int_B p(u, v) \Lambda(dv), \quad (2.1)$$

for all $B \in \mathcal{B}(\mathbb{R})$. As it is natural for Markov processes (see, e.g., [6]) we assume that the function $p(\cdot, y)$ is Borel measurable for each y .

We recall that a coupling is a bivariate processes $\tilde{X}_n = (\tilde{X}_n^1, \tilde{X}_n^2)$ such that $\tilde{X}_n^1 \stackrel{d}{=} X_n^1$, $\tilde{X}_n^2 \stackrel{d}{=} X_n^2$ for all $n = 0, 1, \dots$ and $X_n^1 = X_n^2$ for all $n > n_0(\omega)$. In other words, two copies of the Markov process start from different states and are pasted together after they reach the same state. For motivation of this definition and related discussion see, e.g., [15].

We start with a simple lemma which demonstrates our ideas and provides a coupling between two random variables.

Let X^1 and X^2 be two random variables with densities $p^1(t)$ and $p^2(t)$ with respect to Λ , correspondingly. Define

$$q := \int_{-\infty}^{+\infty} p^1(t) \wedge p^2(t) \Lambda(dt).$$

and assume that $0 < q < 1$. Further we will explain how to deal with degenerate cases $q = 0$ and $q = 1$. Let us introduce independent random variables η_1, η_2 and ξ , with the following densities:

$$\begin{aligned} p_{\eta_1}(t) &= (1 - q)^{-1} (p^1(t) - p^1(t) \wedge p^2(t)), \\ p_{\eta_2}(t) &= (1 - q)^{-1} (p^2(t) - p^1(t) \wedge p^2(t)), \\ p_{\xi}(t) &= q^{-1} (p^1(t) \wedge p^2(t)). \end{aligned}$$

Let ζ be a random variable independent of η^1, η^2 and ξ taking values in $\{0, 1\}$ such that

$$\mathbb{P}(\zeta = 0) = q, \quad \mathbb{P}(\zeta = 1) = 1 - q.$$

Lemma 1. *Assume that $q \neq 0$ and $q \neq 1$. Define the random variables \tilde{X}^1 and \tilde{X}^2 by the following formula:*

$$\begin{aligned} \tilde{X}^1 &:= \eta^1 \mathbb{I}(\zeta = 1) + \xi \mathbb{I}(\zeta = 0), \\ \tilde{X}^2 &:= \eta^2 \mathbb{I}(\zeta = 1) + \xi \mathbb{I}(\zeta = 0). \end{aligned}$$

Then $\tilde{X}^1 \stackrel{d}{=} X^1$ and $\tilde{X}^2 \stackrel{d}{=} X^2$. Moreover, $\mathbb{P}(\tilde{X}^1 = \tilde{X}^2) = q$.

Now let us generalize lemma 1 to a sequence of random variables and present our coupling construction. This construction will be used in the following to obtain an estimate of the total variation distance between the random variables X_n^1 and X_n^2 .

Let us define

$$q(u, v) := \int_{-\infty}^{+\infty} p(u, t) \wedge p(v, t) \Lambda(dt), \quad (2.2)$$

$$q_0 := \int_{-\infty}^{+\infty} p_{X_0^1}(t) \wedge p_{X_0^2}(t) \Lambda(dt). \quad (2.3)$$

It is clear that $0 \leq q(u, v) \leq 1$ for all u, v .

We assume that X_0^1 and X_0^2 have different distributions, so $q_0 < 1$. Otherwise we obviously have $X_n^1 \stackrel{d}{=} X_n^2$ for all n , and the coupling is trivial, namely $\tilde{X}_n^1 = \tilde{X}_n^2 := X_n^1$.

Introduce a Markov process $(\eta_n^1, \eta_n^2, \xi_n, \zeta_n)$. If $q_0 = 0$ then we set

$$\eta_0^1 := X_0^1, \eta_0^2 := X_0^2, \xi_0 := 0, \zeta_0 := 1.$$

Otherwise if $0 < q_0 < 1$ we apply lemma 1 to the random variables X_0^1 and X_0^2 to create $\eta_0^1, \eta_0^2, \xi_0$ and ζ_0 .

Define the transition probability density φ with respect to Λ for this process,

$$\varphi(x, y) := \varphi_1(x, y_1)\varphi_2(x, y_2)\varphi_3(x, y_3)\varphi_4(x, y_4), \quad (2.4)$$

where $x = (x^1, x^2, x^3, x^4)$, $y = (y^1, y^2, y^3, y^4)$ and if $0 < q(x^1, x^2) < 1$, then

$$\begin{aligned} \varphi_1(x, u) &:= (1 - q(x^1, x^2))^{-1} (p(x^1, u) - p(x^1, u) \wedge p(x^2, u)), \\ \varphi_2(x, u) &:= (1 - q(x^1, x^2))^{-1} (p(x^2, u) - p(x^1, u) \wedge p(x^2, u)), \\ \varphi_3(x, u) &:= \mathbb{I}(x^4 = 1)q(x^1, x^2)^{-1}p(x^1, u) \wedge p(x^2, u) + \mathbb{I}(x^4 = 0)p(x^3, u), \\ \varphi_4(x, u) &:= \mathbb{I}(x^4 = 1) (\delta_1(u)(1 - q(x^1, x^2)) + \delta_0(u)q(x^1, x^2)) + \mathbb{I}(x^4 = 0)\delta_0(u). \end{aligned} \quad (2.5)$$

If $q(x^1, x^2) = 0$, then we set¹ $\varphi_3(x, u) := \mathbb{I}(x^4 = 1) \mathbb{I}(x^1 < u < x^1 + 1) + \mathbb{I}(x^4 = 0)p(x^3, u)$ and if $q(x^1, x^2) = 1$ then we set $\varphi_1(x, u) = \varphi_2(x, u) := \mathbb{I}(x^1 < u < x^1 + 1)$. Here notation $\delta_a(u)$ stands for the delta measure concentrated at a . As it follows easily from the construction, the random variables $(\eta_{n+1}^1, \eta_{n+1}^2, \xi_{n+1}, \zeta_{n+1})$ are conditionally independent given $(\eta_n^1, \eta_n^2, \xi_n, \zeta_n)$.

Lemma 2. Define random variables \tilde{X}_n^1 and \tilde{X}_n^2 , $n \in \mathbb{Z}_+$, by the following formulae:

$$\begin{aligned} \tilde{X}_n^1 &:= \eta_n^1 \mathbb{I}(\zeta_n = 1) + \xi_n \mathbb{I}(\zeta_n = 0), \\ \tilde{X}_n^2 &:= \eta_n^2 \mathbb{I}(\zeta_n = 1) + \xi_n \mathbb{I}(\zeta_n = 0). \end{aligned}$$

Then $\tilde{X}_n^1 \stackrel{d}{=} X_{n_2}^1$, $\tilde{X}_n^2 \stackrel{d}{=} X_n^2$ for all $n \in \mathbb{Z}_+$.

Moreover, $\tilde{X}_n^1 = \tilde{X}_n^2$ for all $n \geq n_0(\omega) := \inf\{k \geq 0 : \zeta_k = 0\}$ and

$$\mathbb{P}(\tilde{X}_n^1 \neq \tilde{X}_n^2) \leq (1 - q_0) \mathbb{E} \prod_{i=0}^{n-1} (1 - q(\eta_i^1, \eta_i^2)). \quad (2.6)$$

¹As it follows from the proof below, in degenerated cases $q(x^1, x^2) = 0$ (respectively $q(x^1, x^2) = 1$) functions $\varphi_3(x, u) \mathbb{I}(x^4 = 1)$ (respectively $\varphi_1(x, u)$ and $\varphi_2(x, u)$) can be defined arbitrarily.

Informally speaking, the processes η_n^1 and η_n^2 represent X_n^1 and X_n^2 , correspondingly, under condition that the coupling was not successful until time n . On the other hand, the process ξ_n represents both X_n^1 and X_n^2 , if the coupling occurs before time n . Finally, the process ζ_n represents the moment of coupling. Namely, if $\zeta_n = 0$ then the coupling occurs before time n and vice versa. As it follows from (2.5),

$$\begin{aligned} P(\zeta_{n+1} = 0 | \zeta_n = 0) &= 1, \\ P(\zeta_{n+1} = 0 | \zeta_n = 1, \eta_n^1 = x^1, \eta_n^2 = x^2) &= q(x^1, x^2). \end{aligned}$$

Thus, if two processes were coupled at time n , then they remain coupled at time $n + 1$, and if they were not coupled, then the coupling occurs with the probability $q(\eta_n^1, \eta_n^2)$. We stress once again that at each time probability of coupling at the next step is as large as possible, given the current states.

Remark 1. $(\tilde{X}_n^1)_{n \in \mathbb{Z}_+}$ and $(\tilde{X}_n^2)_{n \in \mathbb{Z}_+}$ are homogeneous Markov processes with respect to their natural filtration. Moreover,

$$(\tilde{X}_n^1)_{n \in \mathbb{Z}_+} \stackrel{d}{=} (X_n^1)_{n \in \mathbb{Z}_+} \quad \text{and} \quad (\tilde{X}_n^2)_{n \in \mathbb{Z}_+} \stackrel{d}{=} (X_n^2)_{n \in \mathbb{Z}_+}. \quad (2.7)$$

Thus the constructed coupling is Markovian.

2.2 Main results

In this chapter we establish some results about convergence of homogeneous Markov processes in the total variation metric.

We recall that the total variation distance between two random variables X and Y is defined as the total variation distance between their laws, i.e.

$$d_{TV}(X, Y) := 2 \sup_{A \in \mathcal{B}(\mathbb{R})} |P(X \in A) - P(Y \in A)|.$$

Remind that $(X_n^1)_{n \in \mathbb{Z}_+}$ and $(X_n^2)_{n \in \mathbb{Z}_+}$ are homogeneous Markov processes with the same transition probability densities $p(u, v)$ with respect to measure Λ . Without loss of generality, we may assume that the processes $(X_n^1)_{n \in \mathbb{Z}_+}$ and $(X_n^2)_{n \in \mathbb{Z}_+}$ are independent. Indeed, if this is not the case, then we can consider their independent copies.

Let us introduce the operator A

$$Af(x) := (1 - q(x))E \{ f(\eta_1) | \eta_0 = x \}, \quad (2.8)$$

where $x = (x^1, x^2)$ is a 2-dimensional vector, $q(x)$ is defined in (2.2), and $\eta_i = (\eta_i^1, \eta_i^2)$ is the first two coordinates of the Markov process $(\eta_i^1, \eta_i^2, \xi_i, \zeta_i)$ defined in (2.4).

Our goal is to find an explicit upper bound of the total variation distance between X_n^1 and X_n^2 . In particular, if X_n^2 has a stationary distribution π , then we get a convergence rate of X_n^1 to its stationary regime. We discuss two different approaches.

The first approach is based on spectral properties of the operator A . It is easy to see that definition (2.8) is equivalent to the following:

$$Af(x^1, x^2) = (1 - q(x^1, x^2)) \int_{\mathbb{R}} \int_{\mathbb{R}} f(t^1, t^2) \psi(x^1, x^2, t^1) \psi(x^2, x^1, t^2) \Lambda(dt^1) \Lambda(dt^2), \quad (2.9)$$

where the function $\psi(u, v, w) := (1 - q(u, v))^{-1} (p(u, w) - p(u, w) \wedge p(v, w))$ if $q(u, v) < 1$ and $\psi(u, v, w) := \mathbf{I}(u < w < u + 1)$ if $q(u, v) = 1$. Therefore A is an integral operator and there exist a number of methods for estimating its spectral radius in various function spaces.

It turns out that strong ergodicity of a Markov process is related to a spectral radius of A ,

$$r(A) = \overline{\lim}_{n \rightarrow \infty} \sqrt[n]{\|A^n\|}. \quad (2.10)$$

Here the norm $\|A^n\|$ should be specified, however any functional space norm may be used if $\|1\| = 1$ in this space. We use the spaces L_∞ and L_p . Although estimates in the space C_b of bounded continuous functions might provide a better upper bound of the constant C in (1.1), they lead to the same bound of λ and require additional conditions on the transition density $p(u, v)$. Therefore we do not consider these estimates.

Theorem 1. *The operator A can be considered as a bounded linear operator on $L_\infty = L_\infty(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2), \Lambda \times \Lambda)$ and has a spectral radius $r(A) \leq 1$.*

Furthermore, if $r(A) \neq 1$ then for any $\varepsilon > 0$ there exists N such that for $n > N$

$$d_{TV}(X_{n+1}^1, X_{n+1}^2) \leq 2(1 - q_0)e^{-n(|\ln r(A)| - \varepsilon)}. \quad (2.11)$$

In some cases it may be easier to work with L_p spaces, in particular, with L_2 . Hence, we state one immediate corollary of this type.

Corollary 1. *Let the following conditions hold:*

- 1) *The operator A is well-defined as $A: L_p \rightarrow L_p$, $1 \leq p < +\infty$ where we denote $L_p = L_p(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2), \mu)$, the measure μ is finite and is of form $\mu(dx) = m(x) \Lambda(dx^1) \Lambda(dx^2)$, $m(x) \neq 0$.*
- 2) *The random variables X_0^1 and X_0^2 have densities $g(u)$ and $h(u)$ correspondingly with respect to the measure Λ .*
- 3) *The spectral radius $r(A)$ of the operator A is less than 1.*

Then for any $\varepsilon > 0$ there exists N such that for $n > N$

$$d_{TV}(X_n^1, X_n^2) \leq C e^{-n(|\ln r(A)| - \varepsilon)}, \quad (2.12)$$

where

$$C := \frac{2}{(1 - q_0)} \left\| \frac{(g(u) - g(u) \wedge h(u))(h(v) - g(v) \wedge h(v))}{m(u, v)} \right\|_{L_q},$$

and $p^{-1} + q^{-1} = 1$.

The second approach uses spectral properties along with recurrence and allows to obtain a non-uniform estimate of convergence rate under more general conditions. It is based on the following simple observation. If the operator A is considered on the space L_∞ (as in Theorem 1) then we obviously have $r(A) < \|A\|_\infty = 1 - \inf_{x \in \mathbb{R}} q(x)$. However this infimum might equal 0 in some cases, making this estimate useless. So we introduce a “good” set

$$K(\varepsilon) := \{(x^1, x^2) : q(x^1, x^2) \geq \varepsilon\},$$

where $0 \leq \varepsilon \leq 1$. Roughly speaking, if the process (X_n^1, X_n^2) visits $K(\varepsilon)$ frequently enough, then the rate of convergence can be exponential or polynomial depending on the frequency of visits. This is the main idea of the alternative approach. It should be noticed that, of course, the use of recurrence in general is one of the main ideas in the whole area. We only link it here to the newly introduced operator A and its spectral radius.

Let $B \in \mathcal{B}(\mathbb{R}^2)$ and define $\tau^B := \inf\{t > 0 : \eta_t \in B\}$. In other words, τ^B is a first time when the process η_n “hits” the set B . Similarly we denote $T^B := \inf\{t > 0 : (X_t^1, X_t^2) \in B\}$.

Proposition 1. *Assume that there exist $\varepsilon > 0$, $\lambda > 0$, $M > 0$, $B \subset K(\varepsilon)$ such that the following conditions hold:*

- 1) $Q := \mathbb{E}e^{\lambda\tau^B} < \infty$.
- 2) For all $x = (x^1, x^2) \in B$ we have $\mathbb{E}_x e^{\lambda\tau^B} \leq M$.

Then there exists a constant $C > 0$, which does not depend on initial distribution (X_0^1, X_0^2) such that

$$d_{TV}(X_n^1, X_n^2) \leq CQe^{-n\theta}, \quad (2.13)$$

where

$$\theta = \frac{|\ln(1 - \varepsilon)|\lambda}{\ln M + |\ln(1 - \varepsilon)|}.$$

Remark 2. *Actually, we can change the first condition of proposition 1 to*

$$Q_1 := \mathbb{E}I(q(\eta_0) < 1)e^{\lambda\tau^B} < \infty.$$

Moreover it is sufficient to check the second condition of the above theorem only for $x \in B \setminus K(1)$.

We also can change inequality (2.13) to the following

$$d_{TV}(X_n^1, X_n^2) \leq CQ_1e^{-n\theta}.$$

Conditions 1 and 2 in proposition 1 are formulated in terms of the derived process η_n . Let us reformulate these conditions in terms of a first hit time of the original process (X_n^1, X_n^2) .

Theorem 2. *Assume that there exist $\varepsilon > 0$, $\lambda > 0$, $M > 0$, $B \subset K(\varepsilon)$ such that the following conditions hold:*

- 1) $Q_2 := \mathbb{E}e^{\lambda T^B} < \infty$.
- 2) For all $x = (x^1, x^2) \in B \setminus K(1)$ we have $\mathbb{E}_x e^{\lambda T^B} < M$.

Then there exists a constant $C > 0$ which does not depend on the initial distribution (X_0^1, X_0^2) such that

$$d_{TV}(X_n^1, X_n^2) \leq CQ_2e^{-n\theta_1}, \quad (2.14)$$

where

$$\theta_1 = \frac{|\ln(1 - \varepsilon)|\lambda}{\ln M + 3|\ln(1 - \varepsilon)|}. \quad (2.15)$$

Notice that setting $B = K(\varepsilon)$ in Theorem 2 may not be optimal. Namely, denote all sets (λ, M) which suit the conditions 1) and 2) of Theorem 2 for a given $B \subset K(\varepsilon)$ by $\mathcal{A}(B)$. Then it may happen that $B_1 \subset B_2 \subset K(\varepsilon)$, but

$$\sup_{(\lambda, M) \in \mathcal{A}(B_1)} \theta_1(\lambda, M) > \sup_{(\lambda, M) \in \mathcal{A}(B_2)} \theta_1(\lambda, M).$$

In other words, for given ε , it may make sense to consider an optimization problem and possibly not to take the maximal $K(\varepsilon)$ itself. In particular, even if the process satisfies the global Markov–Dobrushin assumption, the best constant in the strong ergodicity condition may be better than $|\ln(1 - \varepsilon)|$ (cf. below Example 1) if a localized version is used with some B , see Example 4 below.

Now let us consider the case when T^B has only polynomial moments, rather than exponential. It means that the process (X_n^1, X_n^2) visits B less frequently. So it is natural, that in this case the rate of convergence is slower than exponential. Namely the following theorem is true.

Theorem 3. *Assume that there exist $\varepsilon > 0$, $\lambda \geq 1$, $M > 0$, $B \subset K(\varepsilon)$ such that the following conditions hold:*

- 1) $Q_3 := \mathbb{E} (T^B)^\lambda < \infty$.
- 2) For all $x = (x^1, x^2) \in B \setminus K(1)$ we have $\mathbb{E}_x (T^B)^\lambda < M$.

Then for each $0 < \lambda_1 < \lambda$ there exists a constant $C > 0$ which does not depend on the initial distribution (X_0^1, X_0^2) such that

$$d_{TV}(X_n^1, X_n^2) \leq C Q_3 n^{-\lambda_1}. \quad (2.16)$$

3 Examples and applications

As an illustration to theorems 1 and 2 let us give some specific examples when $d_{TV}(X_n^1, X_n^2)$ converges to zero with exponential rate.

Example 1 (see e.g. [17]). *Suppose the transition probability density $p(u, v)$ satisfies the global Markov–Dobrushin’s condition, namely,*

$$\inf_{u, v} \int_{-\infty}^{+\infty} p(u, t) \wedge p(v, t) \Lambda(dt) = \varepsilon > 0. \quad (3.1)$$

Then

$$d_{TV}(X_n^1, X_n^2) \leq 2(1 - \varepsilon)^n.$$

Condition (3.1) is of a global nature. We give weaker local conditions below.

Example 2. *Suppose there exist $\varepsilon > 0$, $\delta > 0$, $K > 0$ such that*

$$\inf_{\substack{|u| < K \\ |v| < K}} \int_{-\infty}^{+\infty} p(u, t) \wedge p(v, t) \Lambda(dt) = \varepsilon \quad (3.2)$$

and

$$\mathbb{P}(|X_1^1| < K | X_0^1 = u) \geq \delta \quad (3.3)$$

for all u .

Then there exist $C > 0$, $\theta > 0$ such that $d_{TV}(X_n^1, X_n^2) \leq Ce^{-n\theta}$.

Obviously, conditions (3.2) and (3.3) are weaker than the global Markov–Dobrushin’s condition above (example 1). Let us give a more general example with global condition (3.3) replaced to its local version.

We denote $S^D := \inf\{t > 0 : X_t^1 \in D\}$, where $D \in \mathcal{B}(\mathbb{R}^1)$.

Example 3. Assume that there exist $\varepsilon > 0$, $\delta > 0$, $K > 0$, $\lambda > 0$, $M > 0$ such that

- 1) Condition (3.2) is met;
- 2) $\mathbb{P}(|X_1^1| < K | X_0^1 = u) > \delta$ for all $|u| < K$;
- 3) $\mathbb{E}_u e^{\lambda S^{(-K;K)}} < \infty$ for all u ;
- 4) $\mathbb{E}_u e^{\lambda S^{(-K;K)}} < M$ for all $|u| < K$.

Then there exist $C > 0$, $\theta > 0$ such that

$$d_{TV}(X_n^1, X_n^2) \leq Ce^{-n\theta}.$$

Example 3 shows that if a Markov process often visits a certain “good” bounded set, then it converges to the stationary distribution with exponential rate.

The next example illustrates the point mentioned in Section 2.1: even if the global Markov–Dobrushin’s condition is satisfied, the convergence rate may be better, in fact, arbitrarily better, than the rate provided by this condition.

Example 4. Consider a Markov chain on the state space $\mathcal{S} = \{1, 2, 3, 4\}$ with a transition matrix with a small parameter $0 < \delta \ll 1$:

$$\mathcal{P} = \begin{pmatrix} 2/3 - \delta & 1/3 - \delta & \delta & \delta \\ 1/3 - \delta & 2/3 - \delta & \delta & \delta \\ 1 - 3\delta & \delta & \delta & \delta \\ \delta & 1 - 3\delta & \delta & \delta \end{pmatrix}.$$

Here the global Markov–Dobrushin’s condition holds and, according to the Example 1, guarantees the rate

$$2(1 - 4\delta)^n.$$

This rate, apparently, may be arbitrarily slow if δ is close enough to zero. On the other hand, if we denote $B = \{1, 2\} \times \{1, 2\}$, then Theorem 2 provides the following bound (probably, not optimal yet)

$$C \exp(-\theta n),$$

with

$$\theta \geq \frac{\ln 3(|\ln \delta| - \ln 6)}{|\ln \delta| + 4 \ln 3}.$$

Whence, the value of θ tends to optimal rate $\ln 3$ if we let $\delta \rightarrow 0$.

Indeed, this result follows from (2.15) with $\varepsilon = 2/3$, $\lambda = -\ln 6 - \ln \delta - \ln(1 - \delta)$, $M = (1 - 2\delta)^2 / (2\delta(1 - \delta))$.

Let us give some applications of theorems 1 and 3. We shall show how these theorems allow to estimate mixing coefficients of Markov process and establish the central limit theorem.

Let X_n , $n \in \mathbb{Z}_+$ be a homogeneous Markov process with a transitional density $p(u, v)$ with respect to the measure Λ . We assume that the process X_n has a stationary distribution π and let X_n^π be a stationary version of the process X_n .

We recall that β -mixing and φ -mixing coefficients of the process X_n (see, e.g., [8]) are defined by

$$\beta(n) = \sup_{t \geq 0} \mathbb{E} \sup_{K \in \mathcal{F}_{\geq t+n}^X} |P(K | \mathcal{F}_t^X) - P(K)|, \quad (3.4)$$

$$\varphi(n) = \sup_{t \geq 0} \sup_{\substack{A \in \mathcal{F}_t^X \\ B \in \mathcal{F}_{\geq t+n}^X}} |P(B|A) - P(B)|,$$

where $\mathcal{F}_t^X := \sigma\{X_i, i \leq t\}$ and by $\mathcal{F}_{\geq u}^X$ we denote a σ -field generated by random variables $\{X_s, s \geq u\}$.

It turned out that if the process X_n satisfies the conditions of Theorem 1, then it is a φ -mixing process. Alternatively, if the process X_n satisfies the conditions of Theorem 3, then it is a β -mixing process.

Theorem 4. *Assume that the operator $A : L_\infty \mapsto L_\infty$ defined by (2.8) has a spectral radius $r(A) \neq 1$. Then for any $\varepsilon > 0$ there exist $N > 0$ such that for $n > N$*

$$\varphi(n+1) \leq 4e^{-n(|\ln r(A)| - \varepsilon)}.$$

Moreover, if

- 1) $\mathbb{E}|X_1^\pi|^2 < \infty$,
- 2) $\sigma^2 := \text{Var } X_1^\pi + 2 \sum_{k=1}^{\infty} \text{cov}(X_1^\pi, X_{k+1}^\pi) \neq 0$,

then the process X_n satisfies the central limit theorem, i.e.

$$\frac{\sum_{i=1}^n X_i - n\mathbb{E}X_1^\pi}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \sigma^2) \quad \text{as } n \rightarrow \infty.$$

Theorem 5. *Assume that there exist $\varepsilon > 0$, $\lambda > 2$, $M > 0$, $B \subset K(\varepsilon)$ such that the following conditions hold:*

- 1) $\mathbb{E}_{(X_0, X_0^\pi)} (T^B)^\lambda < \infty$.
- 2) $\mathbb{E}_{(u, X_0^\pi)} (T^B)^\lambda < \infty$ for all $u \in \mathbb{R}$.
- 3) $\mathbb{E}_x (T^B)^\lambda < M$ for all $x = (x^1, x^2) \in B$.

Then for all $\lambda_1 < \lambda - 1$ there exists $C > 0$ such that for all $n \in \mathbb{N}$

$$\beta(n) \leq Cn^{-\lambda_1}.$$

Theorem 6. *Suppose the conditions of theorem 5 are satisfied. Furthermore, assume that*

- 1) $E|X_1^\pi|^{2+\delta} < \infty$, for $\delta > \frac{2}{\lambda-2}$,
- 2) $\sigma^2 := \text{Var } X_1^\pi + 2 \sum_{k=1}^{\infty} \text{cov}(X_1^\pi, X_{k+1}^\pi) \neq 0$.

Then the process X_n satisfies the central limit theorem, i.e.

$$\frac{\sum_{i=1}^n X_i - nEX_1^\pi}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \sigma^2) \quad \text{as } n \rightarrow \infty.$$

4 Proofs

4.1 Proofs of the coupling lemmas

Proof of lemma 1. First let us verify that $\tilde{X}^1 \stackrel{d}{=} X^1$. It is sufficient to prove that for any bounded measurable function $f(t)$ we have $Ef(\tilde{X}^1) = Ef(X^1)$. We claim that this is the case. Indeed,

$$\begin{aligned} Ef(\tilde{X}^1) &= Ef(\eta^1) I(\zeta = 1) + Ef(\xi) I(\zeta = 0) = \\ &= (1-q) \int_{-\infty}^{+\infty} f(t)(1-q)^{-1} (p^1(t) - p^1(t) \wedge p^2(t)) \Lambda(dt) + \\ &+ q \int_{-\infty}^{+\infty} f(t)q^{-1} (p^1(t) \wedge p^2(t)) \Lambda(dt) = \\ &= \int_{-\infty}^{+\infty} f(t)p^1(t) \Lambda(dt) = Ef(X^1). \end{aligned}$$

Similarly, $\tilde{X}^2 \stackrel{d}{=} X^2$.

To conclude the proof, it remains to note that

$$P(\tilde{X}^1 = \tilde{X}^2) \geq P(\zeta = 0) = q. \quad \square$$

Proof of lemma 2. Let us prove the first statement of the lemma by induction over n .

Basis. $n = 0$. Then by construction $\tilde{X}_0^1 \stackrel{d}{=} X_0^1$ and $\tilde{X}_0^2 \stackrel{d}{=} X_0^2$ thanks to lemma 1 .

Inductive step. Assume $\tilde{X}_n^1 \stackrel{d}{=} X_n^1$ and $\tilde{X}_n^2 \stackrel{d}{=} X_n^2$. Let us show that $\tilde{X}_{n+1}^1 \stackrel{d}{=} X_{n+1}^1$ and $\tilde{X}_{n+1}^2 \stackrel{d}{=} X_{n+1}^2$.

To prove the first equality it is sufficient to check that $Ef(\tilde{X}_{n+1}^1) = Ef(X_{n+1}^1)$ for any bounded measurable function $f(x)$. We obviously have

$$Ef(\tilde{X}_{n+1}^1) = EE \left(f(\tilde{X}_{n+1}^1) \middle| \eta_n^1, \eta_n^2, \xi_n, \zeta_n \right). \quad (4.1)$$

Let us find the conditional expectation in formula (4.1).

$$\begin{aligned}
& \mathbb{E} \left(f(\tilde{X}_{n+1}^1) \middle| \eta_n^1 = x^1, \eta_n^2 = x^2, \xi_n = x^3, \zeta_n = 1 \right) = \\
& = \mathbb{E} \left(f(\eta_{n+1}^1) \mathbb{I}(\zeta_{n+1} = 1) + f(\xi_{n+1}) \mathbb{I}(\zeta_{n+1} = 0) \middle| \eta_n^1 = x^1, \eta_n^2 = x^2, \xi_n = x^3, \zeta_n = 1 \right) = \\
& = (1 - q(x^1, x^2)) \int_{-\infty}^{+\infty} f(u) \varphi_1(x, u) \Lambda(du) + q(x^1, x^2) \int_{-\infty}^{+\infty} f(u) \varphi_3(x, u) \Lambda(du) = \\
& = \int_{-\infty}^{+\infty} f(u) p(x^1, u) \Lambda(du). \tag{4.2}
\end{aligned}$$

Similarly,

$$\begin{aligned}
& \mathbb{E} \left(f(\tilde{X}_{n+1}^1) \middle| \eta_n^1 = x^1, \eta_n^2 = x^2, \xi_n = x^3, \zeta_n = 0 \right) = \\
& = \mathbb{E} \left(f(\eta_{n+1}^1) \mathbb{I}(\zeta_{n+1} = 1) + f(\xi_{n+1}) \mathbb{I}(\zeta_{n+1} = 0) \middle| \eta_n^1 = x^1, \eta_n^2 = x^2, \xi_n = x^3, \zeta_n = 0 \right) = \\
& = \int_{-\infty}^{+\infty} f(u) \varphi_3(x, u) \Lambda(du) = \int_{-\infty}^{+\infty} f(u) p(x^3, u) \Lambda(du). \tag{4.3}
\end{aligned}$$

Then it follows from (4.2) and (4.3) that

$$\begin{aligned}
\mathbb{E} \left(f(\tilde{X}_{n+1}^1) \middle| \eta_n^1, \eta_n^2, \xi_n, \zeta_n \right) &= \int_{-\infty}^{+\infty} f(u) p(\eta_n^1, u) \Lambda(du) \mathbb{I}(\zeta_n = 1) + \\
&+ \int_{-\infty}^{+\infty} f(u) p(\xi_n, u) \Lambda(du) \mathbb{I}(\zeta_n = 0) = \int_{-\infty}^{+\infty} f(u) p(\tilde{X}_n^1, u) \Lambda(du).
\end{aligned}$$

Therefore formula (4.1) and the inductive assumption imply

$$\mathbb{E} f(\tilde{X}_{n+1}^1) = \mathbb{E} \int_{-\infty}^{+\infty} f(u) p(\tilde{X}_n^1, u) \Lambda(du) = \mathbb{E} \int_{-\infty}^{+\infty} f(u) p(X_n^1, u) \Lambda(du) = \mathbb{E} f(X_{n+1}^1). \tag{4.4}$$

Consequently $\tilde{X}_{n+1}^1 \stackrel{d}{=} X_{n+1}^1$.

Likewise, one can verify that $\tilde{X}_{n+1}^2 \stackrel{d}{=} X_{n+1}^2$. Moreover, it follows from the definition, that $\tilde{X}_n^1 = \tilde{X}_n^2$ for all $n \geq \inf\{k \geq 0 : \zeta_k = 0\} = n_0$.

To establish inequality (2.6) let us note that

$$\mathbb{P}(\tilde{X}_n^1 \neq \tilde{X}_n^2) \leq \mathbb{E} \mathbb{I}(\zeta_0 = 1) \cdot \mathbb{I}(\zeta_1 = 1) \cdot \dots \cdot \mathbb{I}(\zeta_n = 1). \tag{4.5}$$

We denote $\mathcal{G}_n := \sigma(\eta_i^1, \eta_i^2, \xi_i, \zeta_i, i = 1 \dots n)$ and we prove by induction over k that the following identity

$$\mathbb{E} \left(\prod_{i=k}^n \mathbb{I}(\zeta_i = 1) \middle| \mathcal{G}_{k-1} \right) = \mathbb{E} \left(\prod_{i=k-1}^{n-1} (1 - q(\eta_i^1, \eta_i^2)) \middle| \mathcal{G}_{k-1} \right) \mathbb{I}(\zeta_{k-1} = 1) \tag{4.6}$$

holds for all $k \leq n$.

Basis. $k = n$. Then

$$\begin{aligned} \mathbb{E}(\mathbb{I}(\zeta_n = 1) | \mathcal{G}_{n-1}) &= \mathbb{E}(\mathbb{I}(\zeta_n = 1) | \eta_{n-1}^1, \eta_{n-1}^2, \zeta_{n-1}) = \\ &= \mathbb{I}(\zeta_{n-1} = 1)(1 - q(\eta_{n-1}^1, \eta_{n-1}^2)). \end{aligned}$$

Inductive step. Assume that identity (4.6) is proved for all $k \geq K + 1$. Let us prove (4.6) for $k = K$. For simplicity we denote

$$\alpha(x, y) := \mathbb{E} \left(\prod_{i=K}^{n-1} (1 - q(\eta_i^1, \eta_i^2)) \middle| \eta_K^1 = x, \eta_K^2 = y \right).$$

We have,

$$\begin{aligned} \mathbb{E} \left(\prod_{i=K}^n \mathbb{I}(\zeta_i = 1) \middle| \mathcal{G}_{K-1} \right) &= \mathbb{E} \left\{ \mathbb{I}(\zeta_K = 1) \mathbb{E} \left(\prod_{i=K+1}^n \mathbb{I}(\zeta_i = 1) \middle| \mathcal{G}_K \right) \middle| \mathcal{G}_{K-1} \right\} = \\ &= \mathbb{E} \left\{ \mathbb{I}(\zeta_K = 1) \mathbb{E} \left(\prod_{i=K}^{n-1} (1 - q(\eta_i^1, \eta_i^2)) \middle| \mathcal{G}_K \right) \middle| \mathcal{G}_{K-1} \right\} = \\ &= \mathbb{E} \left\{ \mathbb{I}(\zeta_K = 1) \mathbb{E} \left(\prod_{i=K}^{n-1} (1 - q(\eta_i^1, \eta_i^2)) \middle| \eta_K^1, \eta_K^2 \right) \middle| \mathcal{G}_{K-1} \right\} = \\ &= \mathbb{E} \left\{ \mathbb{I}(\zeta_K = 1) \alpha(\eta_K^1, \eta_K^2) \middle| \mathcal{G}_{K-1} \right\} = \\ &= \mathbb{E} \left\{ \mathbb{I}(\zeta_K = 1) \middle| \mathcal{G}_{K-1} \right\} \mathbb{E} \left\{ \alpha(\eta_K^1, \eta_K^2) \middle| \mathcal{G}_{K-1} \right\} = \\ &= \mathbb{I}(\zeta_{K-1} = 1)(1 - q(\eta_{K-1}^1, \eta_{K-1}^2)) \mathbb{E} \left(\prod_{i=K}^{n-1} (1 - q(\eta_i^1, \eta_i^2)) \middle| \mathcal{G}_{K-1} \right) = \\ &= \mathbb{E} \left(\prod_{i=K-1}^{n-1} (1 - q(\eta_i^1, \eta_i^2)) \middle| \mathcal{G}_{K-1} \right) \mathbb{I}(\zeta_{K-1} = 1). \end{aligned} \tag{4.7}$$

Equality (4.7) is correct because the random variables η_K^1 , ξ_K and ζ_K are conditionally independent given \mathcal{G}_{K-1} . Identity (4.6) is proved.

Setting $k = 1$ in (4.6) and combining it with (4.5) we obtain

$$\begin{aligned} \mathbb{P}(\tilde{X}_n^1 \neq \tilde{X}_n^2) &\leq \mathbb{E} \left\{ \mathbb{I}(\zeta_0 = 1) \mathbb{E} \left(\prod_{i=1}^n \mathbb{I}(\zeta_i = 1) \middle| \mathcal{G}_0 \right) \right\} = \\ &= \mathbb{E} \left\{ \mathbb{I}(\zeta_0 = 1) \mathbb{E} \left(\prod_{i=0}^{n-1} (1 - q(\eta_i^1, \eta_i^2)) \middle| \mathcal{G}_0 \right) \right\} = \\ &= (1 - q_0) \mathbb{E} \prod_{i=0}^{n-1} (1 - q(\eta_i^1, \eta_i^2)). \end{aligned}$$

Inequality (2.6) is proved. □

Proof of remark 1. Denote $\tilde{\mathcal{F}}_n^1 := \sigma(\tilde{X}_i^1, 0 \leq i \leq n)$. Then for any bounded measurable function $f(x)$ we have

$$\begin{aligned} \mathbb{E} \left(f(\tilde{X}_{n+1}^1) \middle| \tilde{\mathcal{F}}_n^1 \right) &= \mathbb{E} \left(\mathbb{E} \left(f(\tilde{X}_{n+1}^1) \middle| \mathcal{G}_n \right) \middle| \tilde{\mathcal{F}}_n^1 \right) = \mathbb{E} \left(\mathbb{E} \left(f(\tilde{X}_{n+1}^1) \middle| \eta_n^1, \eta_n^2, \xi_n, \zeta_n \right) \middle| \tilde{\mathcal{F}}_n^1 \right) = \\ &= \mathbb{E} \left(f(\tilde{X}_{n+1}^1) \middle| \tilde{X}_n^1 \right). \end{aligned}$$

Hence, $(\tilde{X}_n^1)_{n \in \mathbb{Z}_+}$ is a homogeneous Markov process and $(\tilde{X}_n^1)_{n \in \mathbb{Z}_+} \stackrel{d}{=} (X_n^1)_{n \in \mathbb{Z}_+}$. \square

4.2 Proofs of the main results

Proof of theorem 1. The proof is based on the coupling construction presented in lemma 2 and inequality (2.6).

First note that

$$r(A) \leq \|A\|_\infty = \|A1(x)\|_\infty \leq \|1 - q(x)\|_\infty \leq 1.$$

We apply lemma 2 to the random processes X_n^1 and X_n^2 and construct the processes \tilde{X}_n^1 and \tilde{X}_n^2 . Then it follows from (2.6) that

$$\frac{1}{2}d_{TV}(X_n^1, X_n^2) = \frac{1}{2}d_{TV}(\tilde{X}_n^1, \tilde{X}_n^2) \leq \mathbb{P}(\tilde{X}_n^1 \neq \tilde{X}_n^2) = (1 - q_0) \mathbb{E} \prod_{i=0}^{n-1} (1 - q(\eta_i^1, \eta_i^2)). \quad (4.8)$$

Clearly, for any function $f \in C_b(\mathbb{R}^2)$

$$\begin{aligned} \mathbb{E} \prod_{i=0}^{n-1} (1 - q(\eta_i)) f(\eta_n) &= \mathbb{E} \mathbb{E} \left(\prod_{i=0}^{n-1} (1 - q(\eta_i)) f(\eta_n) \middle| \eta_0, \dots, \eta_{n-1} \right) = \\ &= \mathbb{E} \left(\prod_{i=0}^{n-2} (1 - q(\eta_i)) (1 - q(\eta_{n-1})) \mathbb{E}(f(\eta_n) | \eta_{n-1}) \right) = \\ &= \mathbb{E} \left(\prod_{i=0}^{n-2} (1 - q(\eta_i)) A f(\eta_{n-1}) \right) = \dots = \mathbb{E} A^n f(\eta_0). \end{aligned} \quad (4.9)$$

It is easy to show that for a Borel set B with $\Lambda(B) = 0$ we have $P(\eta_1^1 \in B) = 0$. Hence combining (4.8) and (4.9) we see that

$$\frac{1}{2}d_{TV}(X_n^1, X_n^2) \leq (1 - q_0) \mathbb{E} A^{n-1} 1(\eta_1) \leq (1 - q_0) \mathbb{E} \|A^{n-1} 1\|_\infty = (1 - q_0) \|A^{n-1} 1\|_\infty. \quad (4.10)$$

Since $\|A^n 1\|_\infty = \|A^n\|_\infty$, it remains to estimate the norm of operator A . It follows from (2.10) that if $r(A) \neq 1$ then for any $\varepsilon > 0$ there exists n_0 such that for $n > n_0$

$$\|A^n\|_\infty \leq e^{-n(|\ln r(A)| - \varepsilon)}.$$

Finally, using the last inequality and (4.10) we get (2.11). This completes the proof of theorem 1. \square

Proof of corollary 1. Arguing as in the proof above, we see that formulas (4.8) and (4.9) are correct. Further,

$$\begin{aligned} \mathbb{E} A^n 1(\eta_0) &= \int_{\mathbb{R}^2} A^n 1(u, v) p_{\eta_0^1}(u) p_{\eta_0^2}(v) \Lambda(du) \Lambda(dv) = \\ &= \int_{\mathbb{R}^2} [A^n 1(u, v) m^{1/p}(u, v)] [p_{\eta_0^1}(u) p_{\eta_0^2}(v) m^{-1/p}(u, v)] \Lambda(du) \Lambda(dv) \leq \\ &\leq \left\| \frac{p_{\eta_0^1}(u) p_{\eta_0^2}(v)}{m(u, v)} \right\|_{L_q} \|A^n 1\|_{L_p} = \\ &= \frac{1}{(1 - q_0)^2} \left\| \frac{(g(u) - g(u) \wedge h(u))(h(v) - g(v) \wedge h(v))}{m(u, v)} \right\|_{L_q} \|A^n 1\|_{L_p}. \end{aligned} \quad (4.11)$$

To conclude the proof, it remains to combine (4.8), (4.9) and (4.11) to get (2.12). \square

Proof of proposition 1. In the proof for simplicity we omit superscript B on τ^B . Consider operator $\tilde{A}: L_\infty \rightarrow L_\infty$, where $L_\infty = L_\infty(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2), \Lambda \times \Lambda)$,

$$\tilde{A}f(x) := \mathbb{E}((1 - q(\eta_1))(1 - q(\eta_2)) \cdots (1 - q(\eta_\tau))f(\eta_\tau) | \eta_0 = x). \quad (4.12)$$

Arguing as in the proof of theorem 1 we see that the operator \tilde{A} is well defined. Furthermore, since

$$|\tilde{A}f(x)| \leq \mathbb{E}\left((1 - q(\eta_\tau))|f(\eta_\tau)| \middle| \eta_0 = x\right) \leq (1 - \varepsilon)\mathbb{E}\left(|f(\eta_\tau)| \middle| \eta_0 = x\right),$$

we have $\|\tilde{A}\| \leq 1 - \varepsilon$.

Let us define $\mathcal{H}_n := \sigma(\eta_0, \eta_1, \dots, \eta_n)$. We introduce a sequence of stopping times $\tau_1 := \tau$, $\tau_{n+1} := \inf\{t > \tau_n : \eta_t \in B\}$.

Note that for any function $f \in C_b(\mathbb{R}^2)$ we have

$$\begin{aligned} \mathbb{E} \prod_{i=0}^{\tau_n} (1 - q(\eta_i)) f(\eta_{\tau_n}) &= \mathbb{E} \mathbb{E} \left(\prod_{i=0}^{\tau_n} (1 - q(\eta_i)) f(\eta_{\tau_n}) \middle| \mathcal{H}_{\tau_{n-1}} \right) = \\ &= \mathbb{E} \left(\prod_{i=0}^{\tau_{n-1}} (1 - q(\eta_i)) \tilde{A}f(\eta_{\tau_{n-1}}) \right) = \cdots = \mathbb{E} \prod_{i=0}^{\tau_1} (1 - q(\eta_i)) \tilde{A}^{n-1} f(\eta_{\tau_1}). \end{aligned} \quad (4.13)$$

Let $\Delta(n)$ be an increasing positive deterministic function of n . We will choose appropriate $\Delta(n)$ later. Introduce the set $\kappa_n := \{\omega : \tau_{\lfloor \Delta(n) \rfloor} \leq n\}$, where by $\lfloor u \rfloor$ we denote a lower integer part of u .

Applying lemma 2 to the random processes X_n^1 and X_n^2 we get

$$\begin{aligned} \frac{1}{2} d_{TV}(X_{n+1}^1, X_{n+1}^2) &= \frac{1}{2} d_{TV}(\tilde{X}_{n+1}^1, \tilde{X}_{n+1}^2) \leq (1 - q_0) \mathbb{E} \prod_{i=0}^n (1 - q(\eta_i)) = \\ &= (1 - q_0) \mathbb{E} \mathbb{I}(\kappa_n) \prod_{i=0}^n (1 - q(\eta_i)) + (1 - q_0) \mathbb{E}(1 - \mathbb{I}(\kappa_n)) \prod_{i=0}^n (1 - q(\eta_i)) \leq \\ &\leq (1 - q_0) \mathbb{E} \tilde{A}^{\lfloor \Delta(n) \rfloor - 1} \mathbb{1}(\eta_{\tau_1}) + (1 - q_0) \mathbb{E}(1 - \mathbb{I}(\kappa_n)) \leq \\ &\leq (1 - q_0) \mathbb{E} \tilde{A}^{\lfloor \Delta(n) \rfloor - 1} \mathbb{1}(\eta_{\tau_1}) + (1 - q_0) \mathbb{P}(\tau_{\lfloor \Delta(n) \rfloor} > n). \end{aligned} \quad (4.14)$$

Evidently,

$$\mathbb{E} \tilde{A}^{\lfloor \Delta(n) \rfloor - 1} \mathbb{1}(\eta_{\tau_1}) \leq \|\tilde{A}\|^{\lfloor \Delta(n) \rfloor - 1} \leq (1 - \varepsilon)^{\lfloor \Delta(n) \rfloor - 1}. \quad (4.15)$$

Let us estimate the second term in (4.14). It follows from Chebyshev inequality that

$$\begin{aligned} \mathbb{P}(\tau_{\lfloor \Delta(n) \rfloor} > n) &\leq e^{-\lambda n} \mathbb{E} e^{\lambda \tau_{\lfloor \Delta(n) \rfloor}} = e^{-\lambda n} \mathbb{E} e^{\lambda \tau_{\lfloor \Delta(n) \rfloor - 1}} \mathbb{E} \left(e^{\lambda(\tau_{\lfloor \Delta(n) \rfloor} - \tau_{\lfloor \Delta(n) \rfloor - 1})} \middle| \mathcal{H}_{\tau_{\lfloor \Delta(n) \rfloor - 1}} \right) = \\ &= e^{-\lambda n} \mathbb{E} e^{\lambda \tau_{\lfloor \Delta(n) \rfloor - 1}} \mathbb{E} \left(e^{\lambda(\tau_{\lfloor \Delta(n) \rfloor} - \tau_{\lfloor \Delta(n) \rfloor - 1})} \middle| \eta_{\tau_{\lfloor \Delta(n) \rfloor - 1}} \right). \end{aligned} \quad (4.16)$$

Now, since $\eta_{\tau_{\lfloor \Delta(n) \rfloor - 1}} \in B$, condition 2) of the theorem implies

$$\mathbb{E} \left(e^{\lambda(\tau_{\lfloor \Delta(n) \rfloor} - \tau_{\lfloor \Delta(n) \rfloor - 1})} \middle| \eta_{\tau_{\lfloor \Delta(n) \rfloor - 1}} = x \right) = \mathbb{E}_x e^{\lambda \tau_1} < M.$$

Combining this with (4.16) we see that

$$\mathbb{P}(\tau_{\lfloor \Delta(n) \rfloor} > n) \leq e^{-\lambda n} M \mathbb{E} e^{\lambda \tau_{\lfloor \Delta(n) \rfloor - 1}} \leq e^{-\lambda n} M^{\lfloor \Delta(n) \rfloor - 1} \mathbb{E} e^{\lambda \tau_1} \leq e^{-\lambda n + \Delta(n) \ln M} M^{-1} \mathbb{E} e^{\lambda \tau_1}.$$

Hence, it follows from estimate above, (4.14) and (4.15) that

$$\frac{1}{2} d_{TV}(X_{n+1}^1, X_{n+1}^2) \leq (1 - \varepsilon)^{-2} e^{\ln(1-\varepsilon)\Delta(n)} + e^{-\lambda n + \Delta(n) \ln M} M^{-1} \mathbb{E} e^{\lambda \tau_1}. \quad (4.17)$$

Thus, we proved that the last inequality is satisfied for all $\Delta(n) > 0$. We choose $\Delta(n)$ which minimizes the right-hand side of the inequality for sufficiently large n . Obviously, such $\Delta(n)$ is the solution of the following equation:

$$-\ln(1 - \varepsilon)\Delta(n) = \lambda n - \Delta(n) \ln M.$$

Consequently,

$$\Delta(n) = \frac{\lambda n}{\ln M - \ln(1 - \varepsilon)}. \quad (4.18)$$

To complete the proof it remains to substitute (4.18) into (4.17). \square

Proof of remark 2. Let us show how the proof of proposition 1 should be modified if condition 2 is satisfied only for $x \in B \setminus K(1)$.

Again, let $\Delta(n)$ be an increasing positive deterministic function of n , which will be chosen later. We consider the following three sets: $\iota_n := \{\omega : \exists k \leq n \ q(\eta_k) = 1\}$, $\kappa_n := \{\omega : \tau_{\lfloor \Delta(n) \rfloor} \leq n\} \setminus \iota_n$ and $\zeta_n := \Omega \setminus (\iota_n \cup \kappa_n)$. Using (4.13) we get (cf. (4.14))

$$\begin{aligned} \frac{1}{2} d_{TV}(X_{n+1}^1, X_{n+1}^2) &\leq (1 - q_0) \mathbb{E} \prod_{i=0}^n (1 - q(\eta_i)) = \\ &= (1 - q_0) \mathbb{E} (\mathbb{I}(\iota_n) + \mathbb{I}(\kappa_n) + \mathbb{I}(\zeta_n)) \prod_{i=0}^n (1 - q(\eta_i)) \leq \\ &\leq (1 - q_0) \mathbb{E} \tilde{A}^{\lfloor \Delta(n) \rfloor - 1} \mathbb{1}(\eta_{\tau_1}) + (1 - q_0) \mathbb{E} \mathbb{I}(\zeta_n) \leq \\ &\leq (1 - q_0) \mathbb{E} \tilde{A}^{\lfloor \Delta(n) \rfloor - 1} \mathbb{1}(\eta_{\tau_1}) + (1 - q_0) \mathbb{P}(\tau_{\lfloor \Delta(n) \rfloor} > n, \Omega \setminus \iota_n). \end{aligned} \quad (4.14')$$

The first term in (4.14') is estimated like in inequality (4.15). To estimate the second term let us notice (cf. (4.16))

$$\begin{aligned} \mathbb{P}(\tau_{\lfloor \Delta(n) \rfloor} > n, \Omega \setminus \iota_n) &= \sum_{k=1}^{\lfloor \Delta(n) \rfloor} \mathbb{P}(\tau_k > n, \tau_{k-1} \leq n, \Omega \setminus \iota_n) \leq \\ &\leq e^{-\lambda n} \sum_{k=1}^{\lfloor \Delta(n) \rfloor} \mathbb{E} e^{\lambda \tau_k} \mathbb{I}(\tau_{k-1} \leq n, \Omega \setminus \iota_{\tau_{k-1}}). \end{aligned} \quad (4.16')$$

Obviously,

$$\mathbb{E} e^{\lambda \tau_k} \mathbb{I}(\tau_{k-1} \leq n, \Omega \setminus \iota_{\tau_{k-1}}) = \mathbb{E} e^{\lambda \tau_{k-1}} \mathbb{I}(\tau_{k-1} \leq n, \Omega \setminus \iota_{\tau_{k-1}}) \mathbb{E} (e^{\lambda(\tau_k - \tau_{k-1})} | \eta_{\tau_{k-1}}).$$

We see that if $\omega \notin \iota_{\tau_{k-1}}$ then $\eta_{\tau_{k-1}}(\omega) \in B \setminus K(1)$. Therefore, modified condition 2 implies

$$\mathbb{I}(\tau_{k-1} \leq n, \Omega \setminus \iota_{\tau_{k-1}}) \mathbb{E} (e^{\lambda(\tau_k - \tau_{k-1})} | \eta_{\tau_{k-1}}) < M \mathbb{I}(\tau_{k-2} \leq n, \Omega \setminus \iota_{\tau_{k-2}}).$$

Hence

$$\begin{aligned} \mathbb{E}e^{\lambda\tau_k} \mathbb{I}(\tau_{k-1} \leq n, \Omega \setminus \iota_{\tau_{k-1}}) &\leq M \mathbb{E}e^{\lambda\tau_{k-1}} \mathbb{I}(\tau_{k-2} \leq n, \Omega \setminus \iota_{\tau_{k-2}}) \leq \\ &\leq M^{k-1} \mathbb{E}e^{\lambda\tau_1} \mathbb{I}(q(\eta_0) < 1). \end{aligned}$$

Consequently, we have

$$\begin{aligned} \mathbb{P}(\tau_{\lfloor \Delta(n) \rfloor} > n, \Omega \setminus \iota_n) &\leq e^{-\lambda n} \sum_{k=1}^{\lfloor \Delta(n) \rfloor} M^{k-1} \mathbb{E}e^{\lambda\tau_1} \mathbb{I}(q(\eta_0) < 1) \leq \\ &\leq e^{-\lambda n + \Delta(n) \ln M} (M-1)^{-1} \mathbb{E}e^{\lambda\tau_1} \mathbb{I}(q(\eta_0) < 1). \end{aligned}$$

The end of the proof is similar to the end of the proof of proposition 1. \square

Proof of theorem 2. To prove the theorem we slightly modify the coupling construction used in lemma 2.

Let us define

$$q'(u, v) := \begin{cases} 1, & \text{if } q(u, v) = 1 \text{ and } (u, v) \in B; \\ \varepsilon, & \text{if } \varepsilon \leq q(u, v) < 1 \text{ and } (u, v) \in B; \\ 0, & \text{otherwise,} \end{cases}$$

Introduce $l(u, v) := q'(u, v)/q(u, v)$. Consider the Markov process $(\eta_n^1, \eta_n^2, \xi_n', \zeta_n')$ and set

$$\eta_0^1 := X_0^1, \eta_0^2 := X_0^2, \xi_0' := 0, \zeta_0' := 1.$$

Assume that the process $(\eta_n^1, \eta_n^2, \xi_n', \zeta_n')$ has the transition probability density $\varphi'(x, y)$ with respect to the measure Λ , where $\varphi'(x, y)$ is defined by the following formula (cf. (2.4) and (2.5))

$$\varphi'(x, y) := \varphi'_1(x, y_1)\varphi'_2(x, y_2)\varphi'_3(x, y_3)\varphi'_4(x, y_4), \quad (4.19)$$

here $x = (x^1, x^2, x^3, x^4)$, $y = (y^1, y^2, y^3, y^4)$ and if $0 < q'(x^1, x^2) < 1$

$$\begin{aligned} \varphi'_1(x, u) &:= (1 - q'(x^1, x^2))^{-1} (p(x^1, u) - [p(x^1, u) \wedge p(x^2, u)]l(x^1, x^2)), \\ \varphi'_2(x, u) &:= (1 - q'(x^1, x^2))^{-1} (p(x^2, u) - [p(x^1, u) \wedge p(x^2, u)]l(x^1, x^2)), \\ \varphi'_3(x, u) &:= \mathbb{I}(x^4 = 1)q'(x^1, x^2)^{-1}[p(x^1, u) \wedge p(x^2, u)]l(x^1, x^2) + \mathbb{I}(x^4 = 0)p(x^3, u), \\ \varphi'_4(x, u) &:= \mathbb{I}(x^4 = 1) (\delta_1(u)(1 - q'(x^1, x^2)) + \delta_0(u)q'(x^1, x^2)) + \mathbb{I}(x^4 = 0)\delta_0(u). \end{aligned}$$

If $q'(x^1, x^2) = 0$ then we set $\varphi'_3(x, u) := \mathbb{I}(x^4 = 1) \mathbb{I}(x^1 < u < x^1 + 1) + \mathbb{I}(x^4 = 0)p(x^3, u)$ and if $q'(x^1, x^2) = 1$ then we set $\varphi'_1(x, u) = \varphi'_2(x, u) := \mathbb{I}(x^1 < u < x^1 + 1)$.

One can easily verify that the statements of lemma 2 and proposition 1 with appropriate modifications are correct for the new coupling. Let us show that conditions of modified proposition 1 are satisfied.

The main advantage of the modified coupling is that the transition probability density of the process $\eta_i' = (\eta_i^1, \eta_i^2)$ can be easily estimated. Indeed, if $x = (x^1, x^2) \in B \setminus K(1)$ we obviously have

$$\varphi'_1(x, u)\varphi'_2(x, v) \leq \frac{p(x^1, u)p(x^2, v)}{(1 - \varepsilon)^2}$$

and if $x \notin B$ then

$$\varphi'_1(x, u)\varphi'_2(x, v) = p(x^1, u)p(x^2, v).$$

Therefore for all $x \notin K(1)$ we have

$$\begin{aligned}
\mathbb{P}_x(\tau > n) &= \mathbb{P}(\eta'_1 \notin B, \dots, \eta'_n \notin B) = \\
&= \int_{\mathbb{R}^2 \setminus B} \cdots \int_{\mathbb{R}^2 \setminus B} \varphi'_1(x, x_1^1) \varphi'_2(x, x_1^2) \cdots \varphi'_1(x_{n-1}, x_n^1) \varphi'_2(x_{n-1}, x_n^2) \Lambda(dx_1) \cdots \Lambda(dx_n) \leq \\
&\leq \frac{1}{(1-\varepsilon)^2} \int_{\mathbb{R}^2 \setminus B} \cdots \int_{\mathbb{R}^2 \setminus B} p(x^1, x_1^1) p(x^2, x_1^2) \cdots p(x_{n-1}^1, x_n^1) p(x_{n-1}^2, x_n^2) \Lambda(dx_1) \cdots \Lambda(dx_n) = \\
&= \frac{1}{(1-\varepsilon)^2} \mathbb{P}_x(T > n). \tag{4.20}
\end{aligned}$$

Consequently, for $x \in B \setminus K(1)$ we obtain

$$\begin{aligned}
\mathbb{E}_x e^{\lambda\tau} &= \sum_{n=1}^{\infty} e^{\lambda n} \mathbb{P}_x(\tau = n) = 1 + (e^\lambda - 1) \sum_{n=0}^{\infty} e^{\lambda n} \mathbb{P}_x(\tau > n) \leq \\
&\leq \frac{1}{(1-\varepsilon)^2} (1 + (e^\lambda - 1) \sum_{n=0}^{\infty} e^{\lambda n} \mathbb{P}_x(T > n)) < \frac{M}{(1-\varepsilon)^2}.
\end{aligned}$$

Similarly, one can prove that $\mathbb{E} e^{\lambda\tau} \mathbb{I}(q(\eta_0) < 1) \leq C \mathbb{E} e^{\lambda T}$ for some $C > 0$. Thus, taking into account remark 2, we see that conditions of proposition 1 are satisfied. \square

Proof of theorem 3. The proof is similar to the proof of theorem 2 and proposition 1.

We consider the same coupling construction as in the proof of theorem 2. We claim that conditions 1 and 2 implies that $\mathbb{E} (\tau^B)^\lambda < \infty$ and

$$\mathbb{E}_x (\tau^B)^\lambda < M(1-\varepsilon)^{-2} \tag{4.21}$$

for all $x \in B \setminus K(1)$. Indeed, both inequalities immediately follow from (4.20).

Now let us prove (2.16). First we assume that (4.21) holds for $x \in B$. As in the proof of proposition 1 we consider a set $\kappa_n := \{\omega : \tau_{[\Delta(n)]} \leq n\}$, where $\Delta(n)$ is an increasing positive deterministic function of n . By a similar argument,

$$\frac{1}{2} d_{TV}(X_{n+1}^1, X_{n+1}^2) \leq (1-\varepsilon)^{[\Delta(n)]-1} + \mathbb{P}(\tau_{[\Delta(n)]} > n).$$

To estimate the second term in the inequality above we use Chebyshev inequality.

$$\begin{aligned}
\mathbb{P}(\tau_{[\Delta(n)]} > n) &\leq n^{-\lambda} \mathbb{E} \tau_{[\Delta(n)]}^\lambda = n^{-\lambda} \mathbb{E} \left(\sum_{k=1}^{[\Delta(n)]} (\tau_k - \tau_{k-1}) \right)^\lambda \leq \\
&\leq n^{-\lambda} [\Delta(n)]^{\lambda-1} \mathbb{E} \sum_{k=1}^{[\Delta(n)]} (\tau_k - \tau_{k-1})^\lambda \leq n^{-\lambda} [\Delta(n)]^\lambda M_1,
\end{aligned}$$

where $M_1 := \max(M(1-\varepsilon)^{-2}, \mathbb{E} (\tau^B)^\lambda)$. Therefore it remains to take $\Delta(n) = n^{1-\lambda_1/\lambda}$ and use an argument similar to that in the proof of remark 2. \square

4.3 Other proofs

Proof of example 1. The proof is a straightforward application of lemma 2 and inequality (2.6). \square

Proof of example 2. Let us verify that conditions of theorem 2 are satisfied. Let $B = K(\varepsilon)$. Obviously, for all x we have

$$P_x(T^{K(\varepsilon)} > 1) \leq 1 - \delta^2.$$

Hence,

$$P_x(T^{K(\varepsilon)} > n) \leq (1 - \delta^2)^n.$$

Therefore both conditions of theorem 2 hold. \square

Proof of example 3. Let us verify that conditions of theorem 2 are satisfied. We set $B = (-K; K) \times (-K; K)$. The main idea of the proof is that after the first component of the bivariate process (X_n^1, X_n^2) reaches the interval $(-K; K)$ it stays there for m more steps with a positive probability (at least δ^m). On the other hand, the probability that the second component does not visit $(-K; K)$ during that time is small for large m . Therefore, there is a positive probability that both X_n^1 and X_n^2 reach $(-K; K)$ simultaneously. Moreover, if the second component still does not reach $(-K; K)$, we repeat the attempt: wait until the first component hits $(-K; K)$ and wait another m units of time with a hope that the second component will also visit $(-K; K)$.

Now let us give a formal proof. Without loss of generality we can assume that X_n^1 and X_n^2 are independent. Let $m > 0$ and $0 < \lambda_1 < \lambda$. We will choose appropriate m and λ_1 later. Our goal is to prove that

$$\sup_{\substack{|x^1| < K \\ |x^2| < K}} E_x e^{\lambda_1 T^{(-K; K) \times (-K; K)}} < \infty. \quad (4.22)$$

In this proof for simplicity by $\{X_{a,b}^i \notin D\}$ we denote the set $\{X_a^i \notin D, X_{a+1}^i \notin D, \dots, X_b^i \notin D\}$, where a and b are integers, $i = 1, 2$ and $D \in \mathcal{B}(\mathbb{R})$.

Step 1. Let $|u| < K$. Condition 4 of the example and Chebyshev inequality imply

$$P_u\left(X_{1,n}^1 \notin (-K; K)\right) \leq M e^{-\lambda n}.$$

Therefore the probability that X_t^1 does not reach interval $(-K; K)$ during m consecutive steps $k < t \leq k + m$ can be estimated as follows:

$$\begin{aligned} P_u\left(X_{k+1, k+m}^1 \notin (-K; K)\right) &= \sum_{i=0}^k P_u\left(X_i^1 \in (-K; K), X_{i+1, k+m}^1 \notin (-K; K)\right) \leq \\ &\leq \sum_{i=0}^k M e^{-\lambda(k+m-i)} < M e^{-\lambda m} (1 - e^{-\lambda m})^{-1} =: \alpha. \end{aligned} \quad (4.23)$$

Hence for the probability that X_t^1 does not belong to $(-K; K)$ on moments $k_j < t \leq k_j + m$, where $j = 1 \dots n$ and $k_1 < k_1 + m < k_2 < \dots < k_n$ we obtain

$$P_u\left(\bigcap_{j=1}^n X_{k_j+1, k_j+m}^1 \notin (-K; K)\right) \leq 2^{n-1} \alpha^n.$$

Indeed, it is sufficient to consider all possible cases of behavior of the process X_t^1 between these intervals. Namely, it can either belongs to $(-K; K)$ or not. The number of all possible cases is 2^{n-1} and the probability of each of the case is less than α^n .

Step 2. Introduce the stopping times

$$\begin{aligned} S_{1,m} &:= \{\inf t > 0 : X_{t-m+1}^1 \in (-K; K), \dots, X_t^1 \in (-K; K)\}, \\ S_{n,m} &:= \{\inf t > S_{n-1,m} + m - 1 : X_{t-m+1}^1 \in (-K; K), \dots, X_t^1 \in (-K; K)\}. \end{aligned}$$

Hence $S_{1,m}$ is the first moment when the process X_t^1 belongs to $(-K, K)$ during m consecutive time steps. Then for the expectation in (4.22) we obtain

$$\mathbb{E}_x e^{\lambda_1 T} \leq \sum_{n=1}^{\infty} \mathbb{E}_x e^{\lambda_1 S_{n,m}} \mathbb{I}(S_{n-1,m} < T \leq S_{n,m}) \leq \sum_{n=0}^{\infty} \mathbb{E}_x \mathbb{I}(T > S_{n,m}) (e^{\lambda_1 S_{n+1,m}} - e^{\lambda_1 S_{n,m}}) + 1.$$

Since X_n^1 and X_n^2 are independent we have

$$\mathbb{E}_x \mathbb{I}(T > S_{n,m}) (e^{\lambda_1 S_{n+1,m}} - e^{\lambda_1 S_{n,m}}) \leq \alpha^n 2^{n-1} (M_1 - 1) M_1^n,$$

where $M_1 := \sup_{|u| < K} \mathbb{E}_u \exp(\lambda_1 S_{1,m})$. Consequently, it remains to prove that $M_1 < 1/2\alpha$.

Indeed, if this is the case, then

$$\mathbb{E}_x e^{\lambda_1 T} \leq \sum_{n=0}^{\infty} \alpha^n 2^{n-1} (M_1 - 1) M_1^n + 1 = 1 + \frac{M_1 - 1}{2 - 4\alpha M_1},$$

and (4.22) holds.

Step 3. Now let us take m such that $\alpha < 1/4$. We claim that it is possible to find λ_1 such that $M_1 < 2$.

Consider the stopping time $\tilde{S}_{1,m+1}$

$$\tilde{S}_{1,m+1} := \begin{cases} m, & \text{if } X_1^1 \in (-K, K), \dots, X_m^1 \in (-K, K); \\ \{\inf t > 0 : X_{t-m}^1 \in (-K; K), \dots, X_t^1 \in (-K; K)\}, & \text{otherwise.} \end{cases}$$

Obviously, $S_{1,m} \leq \tilde{S}_{1,m+1}$. Introduce also time of the first return into $(-K, K)$.

$$R := \{\inf t > 0 : X_t^1 \in (-K; K) \text{ and } \exists s < t : X_s^1 \notin (-K; K)\}.$$

It is clear, that $R > 1$. Therefore for all $n > m$ we have

$$\begin{aligned} \mathbb{P}_u(S_{1,m} > n) &\leq \mathbb{P}_u(\tilde{S}_{1,m+1} > n) = \\ &= \mathbb{E}_u \mathbb{I}(\exists s \leq m : X_s^1 \notin (-K; K)) \mathbb{E}(\mathbb{I}(\tilde{S}_{1,m+1} > n) | \mathcal{F}_R) \leq \\ &\leq (1 - \delta^m) \mathbb{P}_u(\tilde{S}_{1,m+1} > n - 1). \end{aligned}$$

Consequently, for all $|u| < K$ we have $\mathbb{P}_u(S_{1,m} > n) \leq (1 - \delta^m)^{n-m}$. This implies that for sufficiently small λ_1 we get $M_1 < 2$. \square

Proof of theorem 4. Let us denote by X_n^u a Markov process with the transition probability density $p(u, v)$ and with the initial distribution $X_0^u = u$. It follows from theorem 1 that for any $\varepsilon > 0$ there exist $N > 0$ such that if $n > N$, then for all u

$$d_{TV}(X_{n+1}^u, X_{n+1}) \leq 2e^{-n(|\ln r(A)| - \varepsilon)}.$$

Therefore, the statement of the theorem immediately follows from [8, Theorem 19.1.2]. \square

Proof of theorem 5. First, let us notice that in the second least upper bound in (3.4), it is sufficient to consider only sets K of the form $\{\omega : X_{t_1} \in B_1, \dots, X_{t_m} \in B_m\}$, where $t + n \leq t_1 < \dots < t_m$ are positive integers and $B_1, \dots, B_m \in \mathcal{B}(\mathbb{R})$.

By X_n^u we denote a Markov process with the transition probability density $p(u, v)$ and with the initial distributions $X_0^u = u$. It follows from the definition of process X_n^π that $X_n^\pi \stackrel{d}{=} X_m^\pi$ for any positive integers n, m . Then we have

$$\begin{aligned} & \left| \mathbb{P}(X_{t_1} \in B_1, \dots, X_{t_n} \in B_n | X_t = u) - \mathbb{P}(X_{t_1} \in B_1, \dots, X_{t_n} \in B_n) \right| \leq \\ & = \left| \mathbb{P}(X_{t_1-t}^u \in B_1, \dots, X_{t_n-t}^u \in B_n) - \mathbb{P}(X_{t_1-t}^\pi \in B_1, \dots, X_{t_n-t}^\pi \in B_n) \right| + \\ & + \left| \mathbb{P}(X_{t_1}^\pi \in B_1, \dots, X_{t_n}^\pi \in B_n) - \mathbb{P}(X_{t_1} \in B_1, \dots, X_{t_n} \in B_n) \right|. \end{aligned} \quad (4.24)$$

Introduce λ_2 such that $\lambda_1 < \lambda_2 < \lambda - 1$. Let us apply theorem 3 to the Markov processes X_n^u and X_n^π with λ_2 in place of λ . Then it follows from (2.16) that for all $u \in \mathbb{R}$ and for all $k \in \mathbb{Z}_+$

$$d_{TV}(X_k^\pi, X_k^u) \leq Cg(u)k^{-\lambda_1},$$

where $g(u) := \mathbb{E}_{(u, X_0^\pi)} (T^B)^{\lambda_2}$.

Hence, taking into account remark 1, we obtain

$$\begin{aligned} & \left| \mathbb{P}(X_{t_1-t}^u \in B_1, \dots, X_{t_n-t}^u \in B_n) - \mathbb{P}(X_{t_1-t}^\pi \in B_1, \dots, X_{t_n-t}^\pi \in B_n) \right| \leq \\ & \leq \frac{1}{2} d_{TV}(X_{t_1-t}^\pi, X_{t_1-t}^u) \leq \frac{1}{2} Cg(u)(t_1 - t)^{-\lambda_1} \leq \frac{1}{2} Cg(u)n^{-\lambda_1}. \end{aligned} \quad (4.25)$$

Similarly,

$$\begin{aligned} \left| \mathbb{P}(X_{t_1} \in B_1, \dots, X_{t_n} \in B_n) - \mathbb{P}(X_{t_1}^\pi \in B_1, \dots, X_{t_n}^\pi \in B_n) \right| & \leq \frac{1}{2} d_{TV}(X_{t_1}^\pi, X_{t_1}) \leq \\ & \leq \frac{1}{2} Cgn^{-\lambda_1}, \end{aligned} \quad (4.26)$$

where $g = \mathbb{E}_{(X_0, X_0^\pi)} (T^B)^{\lambda_2}$.

Thus, using (3.4), (4.24), (4.25) and (4.26) we get

$$\beta(n) \leq \frac{1}{2} C \sup_{t \geq 0} \mathbb{E}(g(X_t) + g)n^{-\lambda_1}.$$

Hence, it remains to prove, that

$$\sup_{t \geq 0} \mathbb{E}g(X_t) < \infty.$$

We claim that this is the case. Indeed, if we introduce a first hit time after time n

$$T^{(n)} := \inf\{t > n : (X_t, X_t^\pi) \in B\},$$

then $\mathbb{E}g(X_t) = \mathbb{E}(T^{(t)} - t)^{\lambda_2}$. Moreover, it follows from conditions 2) and 4) of the theorem that for all n and for all $x \in B$

$$\begin{aligned} \mathbb{P}_x(T^B > n) & \leq n^{-\lambda} \mathbb{E}_x (T^B)^\lambda \leq Mn^{-\lambda}, \\ \mathbb{P}(T^B > n) & \leq n^{-\lambda} \mathbb{E}_{(X_0, X_0^\pi)} (T^B)^\lambda. \end{aligned}$$

Hence,

$$\begin{aligned}
\mathbb{P}(T^{(t)} - t > n) &= \mathbb{P}(T^{(t)} > n + t) = \\
&= \sum_{k=1}^t \mathbb{P}((X_k, X_k^\pi) \in B, (X_{k+1}, X_{k+1}^\pi) \notin B, \dots, (X_{t+n}, X_{t+n}^\pi) \notin B) + \\
&+ \mathbb{P}((X_1, X_1^\pi) \notin B, \dots, (X_{t+n}, X_{t+n}^\pi) \notin B) = \\
&= \sum_{k=1}^t \mathbb{E} \left[\mathbb{I}((X_k, X_k^\pi) \in B) \mathbb{P}_{(X_k, X_k^\pi)}(T^B > t + n - k) \right] + \mathbb{P}(T^B > t + n) \leq \\
&\leq M_1 \sum_{k=0}^t (t + n - k)^{-\lambda} \leq M_1 \sum_{k=n}^{\infty} k^{-\lambda} \leq M_2 n^{-\lambda+1},
\end{aligned}$$

for some $M_1 > 0$ and $M_2 > 0$. Note that the final estimate of probability $\mathbb{P}(T^{(t)} - t > n)$ does not depend on t . Consequently, since $\lambda_2 < \lambda - 1$ we have

$$\mathbb{E}(T^{(t)} - t)^{\lambda_2} < C_1$$

for some $C_1 > 0$ for all t .

This completes the proof of theorem 5. \square

Proof of theorem 6. First, let us prove the central limit theorem (CLT) for process X_n^π . This process is stationary, therefore CLT follows from CLT for stationary processes with mixing ([8, Theorem 18.5.3]). Indeed, it sufficient to check that $\sum_{n=1}^{\infty} \alpha(n)^{\frac{\delta}{2+\delta}} < \infty$, where by $\alpha(n)$ we denote α -mixing coefficient of sequence X_n^π (see, e.g., [8]).

Let us apply theorem 5 to the process X_n^π with $\lambda_1 > \frac{2+\delta}{\delta}$. Since $\alpha(n) \leq \beta(n)$, we get

$$\sum_{n=1}^{\infty} \alpha(n)^{\frac{\delta}{2+\delta}} \leq \sum_{n=1}^{\infty} \beta(n)^{\frac{\delta}{2+\delta}} < \infty,$$

where the last series converges because $\beta(n) = O(n^{-\lambda_1})$.

Now, let us establish CLT in the general case. Let $u \in \mathbb{R}$ and let us fix N_0 . We have for $n > N_0$

$$\begin{aligned}
&\left| \mathbb{P}\left(\frac{\sum_{i=1}^n X_i - n\mathbb{E}X_1^\pi}{\sqrt{n}} \leq u\right) - \mathbb{P}\left(\frac{\sum_{i=1}^n X_i^\pi - n\mathbb{E}X_1^\pi}{\sqrt{n}} \leq u\right) \right| \leq \\
&\leq \left| \mathbb{P}\left(\frac{\sum_{i=1}^n X_i - n\mathbb{E}X_1^\pi}{\sqrt{n}} \leq u\right) - \mathbb{P}\left(\frac{\sum_{i=N_0}^n X_i - n\mathbb{E}X_1^\pi}{\sqrt{n}} \leq u\right) \right| + \\
&+ \left| \mathbb{P}\left(\frac{\sum_{i=N_0}^n X_i - n\mathbb{E}X_1^\pi}{\sqrt{n}} \leq u\right) - \mathbb{P}\left(\frac{\sum_{i=N_0}^n X_i - n\mathbb{E}X_1^\pi}{\sqrt{n}} \leq u\right) \right| + \\
&+ \left| \mathbb{P}\left(\frac{\sum_{i=N_0}^n X_i^\pi - n\mathbb{E}X_1^\pi}{\sqrt{n}} \leq u\right) - \mathbb{P}\left(\frac{\sum_{i=1}^n X_i^\pi - n\mathbb{E}X_1^\pi}{\sqrt{n}} \leq u\right) \right|.
\end{aligned}$$

The first and the last term in the right-hand side of the inequality above tend to zero, as $n \rightarrow \infty$. Lemma 2 yields that the middle term is less than $d_{TV}(X_{N_0}, X_{N_0}^\pi)$ and also tends to zero, as $N_0 \rightarrow \infty$. Therefore for all u

$$\lim_{n \rightarrow \infty} \left| \mathbb{P}\left(\frac{\sum_{i=1}^n X_i - n\mathbb{E}X_1^\pi}{\sqrt{n}} \leq u\right) - \mathbb{P}\left(\frac{\sum_{i=1}^n X_i^\pi - n\mathbb{E}X_1^\pi}{\sqrt{n}} \leq u\right) \right| = 0.$$

Theorem 6 is proved. \square

References

- [1] A.A. Borovkov, *Ergodicity and Stability of Stochastic Processes*, Wiley, Chichester, 1998.
- [2] P. Diaconis, D. Stroock, Geometric bounds for eigenvalues of Markov chains, *Annal. Appl. Prob.*, 1 (1991) 36-61.
- [3] R.L. Dobrushin, Central limit theorem for non-stationary Markov chains, I and II, *Theory Prob. Appl.*, 1 (1956) 65–80 & 329–83.
- [4] W. Doeblin, Exposé de la théorie des chaînes simples constantes de Markov á un nombre fini d'états. *Rev. Math. Union Interbalkanique*, 2 (1938) 77–105.
- [5] J.L. Doob, *Stochastic processes*, Wiley, New York, 1953.
- [6] E.B. Dynkin, *Markov processes*, Springer, Berlin, 1965.
- [7] D. Griffeath, A maximal coupling for Markov chains, *Z. Wahrsch. Verw. Gebiete*, 31, 2 (1975) 95-106.
- [8] I.A. Ibragimov, Yu.V. Linnik, *Independent and Stationary Sequences of Random Variables*, Wolters-Noordhoff Publishing, Groningen, 1971.
- [9] V.V. Kalashnikov, The property of γ -reflexivity for Markov sequences, *Soviet. Math. Dokl*, 14 (1973) 1869–1873.
- [10] J. Lamperti, Criteria for the recurrence or transience of stochastic process, I, *J. Math. Anal. Appl.*, 1 (1960) 314-330.
- [11] T. Lindvall, *Lectures on the coupling method*, Dover Publications, New York, 2002.
- [12] A.A. Markov, Extension of the law of large numbers to dependent quantities (in Russian), *Izv. Fiz.-Matem. Obsch. Kazan Univ.*, 15 (1906) 135–156.
- [13] A.A. Markov, *Izbrannie Trudy (Selected Works)*, AN SSSR, Leningrad, 1951.
- [14] E. Nummelin, *General irreducible Markov chains and non-negative operators*, Cambridge University Press, Cambridge, 1984.
- [15] J.W. Pitman, On Coupling of Markov Chains, *Z. Wahrsch. Verw. Gebiete*, 35 (1976) 315–322.
- [16] S. Orey, *Lecture notes on limit theorems for Markov chain transition probabilities*, Van Nostrand Reinhold, London, 1971.
- [17] E. Seneta, *Non-Negative Matrices and Markov Chains*, Springer-Verlag, New York, 1981.
- [18] R.L. Tweedie, Criteria for Ergodicity, Exponential Ergodicity and Strong Ergodicity of Markov Processes, *J. Appl. Prob.*, 18 (1981) 122–130.
- [19] L.N. Vaserstein, Markov processes on countable product spaces describing large systems of automata. *Prob. Pered. Inf.*, 3 (1969) 64–72.

- [20] A.Yu. Veretennikov, On polynomial mixing and convergence rate for stochastic difference and differential equations, *Probab. Theory Appl.*, 44, 2 (1999) 317–321.
- [21] A.Yu. Veretennikov, On lower bounds for mixing coefficients of Markov diffusions, In: Yu. Kabanov, R. Liptser, J. Stoyanov (Eds.), *From Stochastic Calculus to Mathematical Finance; The Shiryaev Festschrift*, Springer, Berlin, 2006, pp. 623–633.