Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2016, Vol. 52, No. 4, 1799–1822
DOI: 10.1214/15-AIHP699

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Subgeometric rates of convergence in Wasserstein distance for Markov chains

Alain Durmus, Gersende Fort and Éric Moulines

LTCI, Telecom ParisTech & CNRS, 46 rue Barrault, 75634 Paris Cedex 13, France.
E-mail: alain.durmus@telecom-paristech.fr; gersende.fort@telecom-paristech.fr; eric.moulines@telecom-paristech.fr

Received 6 November 2014; revised 4 July 2015; accepted 9 July 2015

Abstract. In this paper, we provide sufficient conditions for the existence of the invariant distribution and for subgeometric rates of convergence in Wasserstein distance for general state-space Markov chains which are (possibly) not irreducible. Compared to (*Ann. Appl. Probab.* **24** (2) (2014) 526–552), our approach is based on a purely probabilistic coupling construction which allows to retrieve rates of convergence matching those previously reported for convergence in total variation in (*Bernoulli* **13** (3) (2007) 831–848).

Our results are applied to establish the subgeometric ergodicity in Wasserstein distance of non-linear autoregressive models and of the pre-conditioned Crank–Nicolson Markov chain Monte Carlo algorithm in Hilbert space.

Résumé. Dans cet article, nous donnons des conditions suffisantes pour l'existence d'une probabilité invariante et qui permettent d'établir des taux de convergence sous-géométriques en distance de Wasserstein, pour des chaînes de Markov définies sur des espaces d'états généraux et non nécessairement irréductibles. Comparée à (*Ann. Appl. Probab.* **24** (2) (2014) 526–552), notre approche est basée sur une construction par couplage purement probabiliste, ce qui permet de retrouver les taux de convergence obtenus précédement pour la variation total dans (*Bernoulli* **13** (3) (2007) 831–848).

Par application de ces résultats, nous établissons la convergence sous-géométrique en distance de Wasserstein de modèles non linéaires auto-régressifs et l'algorithme de MCMC, l'algorithme de Crank–Nicolson pré-conditionné dans les espaces de Hilbert, pour une certaine classe de mesure cible.

MSC: 60J10; 60B10; 60J05; 60J22; 65C40

Keywords: Markov chains; Wasserstein distance; Subgeometric ergodicity; Markov chain Monte Carlo in infinite dimension

1. Introduction

Convergence of general state-space Markov chains in total variation distance (or V-total variation) has been studied by many authors. There is a wealth of contributions establishing explicit rates of convergence under conditions implying geometric ergodicity; see [19, Chapter 16], [2,7,21] and the references therein. Subgeometric (or Riemanian) convergence has been more scarcely studied; [24] characterized subgeometric convergence using a sequence of drift conditions, which proved to be difficult to use in practice. [15] have shown that, for polynomial convergence rates, this sequence of drift conditions can be replaced by a single drift condition, which shares some similarities with the classical Foster–Lyapunov approach for the geometric ergodicity. This result was later extended by [12] and [9] to general subgeometric rates of convergence. Explicit convergence rates were obtained in [10,12,25] and [1].

The classical proofs of convergence in total variation distance are based either on a regenerative or a pairwise coupling construction, which requires the existence of accessible small sets and additional assumptions to control the moments of the successive return time to these sets. The existence of an accessible small set implies that the chain is irreducible.

In this paper, we establish rates of convergence for general state-space Markov chains which are (possibly) not irreducible. In such cases, Markov chains might not converge in total variation distance, but nevertheless may converge in a weaker sense; see for example [18]. We study in this paper the convergence in Wasserstein distance, which also implies the weak convergence. The use of the Wasserstein distance to obtain explicit rates of convergence has been considered by several authors, most often under conditions implying geometric ergodicity. A significant breakthrough in this domain has been achieved in [13]. The main motivation of [13] was the convergence of the solutions of stochastic delay differential equations (SDDE) to their invariant measure. Nevertheless, the techniques introduced in [13] laid the foundations of several contributions. [14] used these techniques to prove the convergence of Markov chain Monte Carlo algorithms in infinite dimensional Hilbert spaces. An application for switched and piecewise deterministic Markov processes can be found in [8]. The results of [13] were generalized by [6] which establishes conditions implying the existence and uniqueness of the invariant distribution, and the subgeometric ergodicity of Markov chains (in discrete-time) and Markov processes (in continuous-time). [6] used this result to establish subgeometric ergodicity of the solutions of SDDE. Nevertheless, when applied to the context of V-total variation, the rates obtained in [6] in discrete-time do not exactly match the rates established in [9].

In this paper, we complement and sharpen the results presented in [6] in the discrete-time setting. The approach developed in this paper is based on a coupling construction, which shares some similarities with the pairwise coupling used to prove geometric convergence in V-total variation. The arguments are therefore mostly probabilistic whereas [6] heavily relies on functional analysis techniques and methods. We provide a sufficient condition couched in terms of a single drift condition for a coupling kernel outside an appropriately defined coupling set, extending the notion of d-small set of [13]. We then show how this single drift condition implies a sequence of drift inequalities from which we deduce an upper bound of some subgeometric moment of the successive return times to the coupling set. The last step is to show that the Wasserstein distance between the distribution of the chain and the invariant probability measure is controlled by these moments. We apply our results to the convergence of some Markov chain Monte Carlo samplers with heavy tailed target distribution and to nonlinear autoregressive models whose the noise distribution can be singular with the Lebesgue measure. We also study the convergence of the preconditioned Crank–Nicolson algorithm when the target distribution has a density w.r.t. a Gaussian measure on an Hilbert space, under conditions which are weaker than [14].

The paper is organized as follows: in Section 2, the main results on the convergence of Markov chains in Wasserstein distance are presented, under different sets of assumptions. Section 3 is devoted to the applications of these results. The proofs are given in Section 4 and Section 5.

Notations

Let (E, d) be a Polish space where d is a distance bounded by 1. We denote by $\mathcal{B}(E)$ the associated Borel σ-algebra and $\mathcal{P}(E)$ the set of probability measures on $(E, \mathcal{B}(E))$. Let $\mu, \nu \in \mathcal{P}(E)$; λ is a coupling of μ and ν if λ is a probability on the product space $(E \times E, \mathcal{B}(E \times E))$, such that $\lambda(A \times E) = \mu(A)$ and $\lambda(E \times A) = \nu(A)$ for all $A \in \mathcal{B}(E)$. The set of couplings of $\mu, \nu \in \mathcal{P}(E)$ is denoted $\mathcal{C}(\mu, \nu)$. Let P be Markov kernel of $E \times \mathcal{B}(E)$; a Markov kernel Q on $(E \times E, \mathcal{B}(E \times E))$ such that, for every $x, y \in E$, $Q((x, y), \cdot)$ is a coupling of $P(x, \cdot)$ and $P(y, \cdot)$ is a *coupling kernel* for P.

The Wasserstein metric associated with d, between two probability measures $\mu, \nu \in \mathcal{P}(E)$ is defined by:

$$W_d(\mu, \nu) = \inf_{\gamma \in \mathcal{C}(\mu,\nu)} \int_{E \times E} d(x, y)\, \mathrm{d}\gamma(x, y). \tag{1}$$

When d is the trivial metric $d_0(x, y) = \mathbb{1}_{x \neq y}$, the associated Wasserstein metric is the total variation distance $W_{d_0}(\mu, \nu) = \sup_{A \in \mathcal{B}(E)} |\mu(A) - \nu(A)|$. Since d is bounded, the Monge–Kantorovich duality theorem implies (see [26, Remark 6.5]) that the lower bound in (1) is realized. In addition, W_d is a metric on $\mathcal{P}(E)$ and $\mathcal{P}(E)$ equipped with W_d is a Polish space; see [26, Theorems 6.8 and 6.16]. Finally, the convergence in W_d is equivalent to the weak convergence, since W_d is equivalent to the Prokorov metric (see e.g. [4, Theorem 6.8 and 6.9]).

Let Λ_0 be the set of measurable functions $r_0 : \mathbb{R}_+ \to [2, +\infty)$, such that r_0 is non-decreasing, $x \mapsto \log(r_0(x))/x$ is non-increasing and $\lim_{x\to\infty} \log(r_0(x))/x = 0$. Denote by Λ the set of positive functions $r : \mathbb{R}_+ \to (0, +\infty)$, such that there exists $r_0 \in \Lambda_0$ satisfying:

$$0 < \liminf_{x\to+\infty} r(x)/r_0(x) \le \limsup_{x\to+\infty} r(x)/r_0(x) < +\infty. \tag{2}$$

Finally, let $\mathbb{F}$ be the set of concave increasing functions $\phi : \mathbb{R}_+ \to \mathbb{R}_+$, continuously differentiable on $[1, +\infty)$, and satisfying $\lim_{x\to+\infty} \phi(x) = +\infty$ and $\lim_{x\to+\infty} \phi'(x) = 0$. For $\phi \in \mathbb{F}$, we denote by $\phi^{\leftarrow}$ the inverse of ϕ.

2. Main results

The key ingredient for the derivation of the convergence of a Markov kernel P on (E, d) is the existence of a coupling kernel $Q((x, y), \cdot)$ for P satisfying a strong contraction property when (x, y) belongs to a set Δ, referred to as a *coupling set*. For $\Delta \in \mathcal{B}(E \times E)$, a positive integer ℓ and $\varepsilon > 0$, consider the following assumption:

H1$(\Delta, \ell, \varepsilon)$**.**

(i) *Q is a d-weak-contraction: for every $x, y \in E$, $Qd(x, y) \le d(x, y)$.*
(ii) *$Q^\ell d(x, y) \le (1 - \varepsilon)d(x, y)$, for every $(x, y) \in \Delta$.*

A set Δ satisfying H1$(\Delta, \ell, \varepsilon)$(ii) will be referred to as a (ℓ, ε, d)-coupling set. Of course the definition of this set also depends on the choice of the coupling kernel Q, but this dependence is implicit in the notation. If $d = d_0$ and Δ is a $(1, \varepsilon)$-pseudo small set (with $\varepsilon > 0$) in the sense that

$$\inf_{(x,y)\in\Delta} \bigl[P(x, \cdot) \wedge P(y, \cdot)\bigr](E) \ge \varepsilon,$$

then H1$(\Delta, 1, \varepsilon)$ is satisfied by the pairwise coupling kernel (see [20]). Furthermore, a simple way to check that $\Delta \in \mathcal{B}(E \times E)$ is a $(1, \varepsilon, d)$-coupling set is the following. Let $\varepsilon > 0$. If for all $(x, y) \in E \times E$, $W_d(P(x, \cdot), P(y, \cdot)) \le d(x, y)$, and for all $(x, y) \in \Delta$, $W_d(P(x, \cdot), P(y, \cdot)) \le (1 - \varepsilon)d(x, y)$, then [26, Corollary 5.22] implies that there exists a Markov kernel Q on $(E \times E, \mathcal{B}(E \times E))$ satisfying H1$(\Delta, 1, \varepsilon)$.

The following theorem shows that, under H1$(\Delta, \ell, \varepsilon)$ and a condition which essentially claims that if the first moment of the hitting time to the coupling set Δ is finite, the Markov kernel P admits a unique invariant distribution.

Theorem 1. *Assume that there exist*

(i) *a coupling kernel Q for P, a set $\Delta \in \mathcal{B}(E \times E)$, $\ell \in \mathbb{N}^*$ and $\varepsilon > 0$ such that* H1$(\Delta, \ell, \varepsilon)$ *holds*,
(ii) *a measurable function $\mathrm{V} : E^2 \to [1, \infty)$ and a constant $b < \infty$ such that the following drift condition is satisfied*:

$$Q\mathrm{V}(x, y) \le \mathrm{V}(x, y) - 1 + b\mathbb{1}_\Delta(x, y), \qquad \sup_{(x,y)\in\Delta} Q^{\ell-1}\mathrm{V}(x, y) < +\infty, \tag{3}$$

(iii) *an increasing sequence of integers $\{n_k, k \in \mathbb{N}\}$ and a concave function $\psi : \mathbb{R}^+ \to \mathbb{R}^+$ such that $\lim_{v\to+\infty} \psi(v) = +\infty$ and*

$$\sup_{k\in\mathbb{N}} P^{n_k}[\psi \circ \mathrm{V}_{x_0}](x_0) < +\infty, \qquad P\mathrm{V}_{x_0}(x_0) < +\infty \quad \textit{for some } x_0 \in E, \tag{4}$$

where $\mathrm{V}_{x_0} = \mathrm{V}(x_0, \cdot)$.

Then, P admits a unique invariant distribution.

Proof. See Section 4.1. □

If we now combine H1$(\Delta, \ell, \varepsilon)$ with a condition which implies the control of the tail probabilities of the successive return times to the coupling sets (more precisely, of the moments of order larger than one of these return times) then the Wasserstein distance between $P^n(x, \cdot)$ and $P^n(y, \cdot)$ may be shown to decrease at a subgeometric rate. To control these moments, it is quite usual to consider drift conditions. In this paper, we focus on a class of drift conditions which has been first introduced in [9]. For $\Delta \in \mathcal{B}(E \times E)$, a function $\phi \in \mathbb{F}$, a measurable function $V : E \to [1, +\infty)$, consider the following assumption:

H2(Δ, ϕ, V).

(i) *There exists a constant* $b < \infty$ *such that for all* $x, y \in E$:

$$PV(x) + PV(y) \leq V(x) + V(y) - \phi\big(V(x) + V(y)\big) + b\mathbb{1}_\Delta(x, y). \tag{5}$$

(ii) $\sup_{(x,y)\in\Delta}\{V(x) + V(y)\} < +\infty$.

Not surprisingly, this condition implies that the return time to the coupling set Δ possesses a first moment. This property combined with Theorem 1 yields

Corollary 2. *Assume that there exist a coupling kernel* Q *for* P, $\Delta \in \mathcal{B}(E \times E)$, $\ell \in \mathbb{N}^*$, $\varepsilon > 0$, $\phi \in \mathbb{F}$ *and* $V : E \to [1, \infty)$ *such that* H1$(\Delta, \ell, \varepsilon)$–H2$(\Delta, \phi, V)$ *are satisfied. Then,* P *admits a unique invariant probability measure* π *and* $\int_E \phi \circ V(x)\pi(\mathrm{d}x) < \infty$.

Proof. See Section 4.2. □

We now derive expressions of the rate of convergence and make explicit the dependence upon the initial condition of the chain. For $\phi \in \mathbb{F}$, set

$$H_\phi(t) = \int_1^t \frac{1}{\phi(s)}\,\mathrm{d}s. \tag{6}$$

Since for $t \geq 1$, $\phi(t) \leq \phi(1) + \phi'(1)(t - 1)$, the function H_ϕ is monotone increasing to infinity, twice continuously differentiable and concave. Its inverse, denoted $H_\phi^\leftarrow$, is well defined on $\mathbb{R}_+$, is twice continuously differentiable and convex (see e.g. [9, Section 2.1]).

Theorem 3. *Assume that there exist a coupling kernel* Q *for* P, $\Delta \in \mathcal{B}(E \times E)$, $\ell \in \mathbb{N}^*$, $\varepsilon > 0$, $\phi \in \mathbb{F}$ *and* $V : E \to [1, \infty)$ *such that* H1$(\Delta, \ell, \varepsilon)$–H2$(\Delta, \phi, V)$ *are satisfied. Let* π *be the invariant probability of* P.

(i) *There exist constants* $\{C_i\}_{i=1}^3$ *such that for all* $x \in E$ *and all* $n \geq 1$

$$\begin{aligned} W_d\big(P^n(x, \cdot), \pi\big) \leq{}& C_1 V(x)/H_\phi^\leftarrow(n/2) + C_2/\phi\big(H_\phi^\leftarrow(n/2)\big) \\ &+ C_3/H_\phi^\leftarrow\big(-\log(1-\varepsilon)n/\big\{2\big(\log\big(H_\phi^\leftarrow(n)\big) - \log(1-\varepsilon)\big)\big\}\big). \end{aligned}$$

(ii) *For all* $\delta \in (0, 1)$, *there exists a constant* C_δ *such that for all* $x \in E$ *and all* $n \geq 1$

$$W_d\big(P^n(x, \cdot), \pi\big) \leq C_\delta V(x)/\phi\big(\big\{H_\phi^\leftarrow(n)\big\}^\delta\big).$$

The values of the constants C_i, *for* $i = 1, 2, 3$, *and* C_δ *are given explicitly in the proof, and depend on* $\Delta, \ell, \varepsilon, \phi, V, b$.

Proof. See Section 4.3 □

We summarize in Table 1 the rates of convergence obtained (for a given $x \in E$) from Theorem 3 for usual concave functions ϕ: logarithmic rates $\phi(t) = (1 + \log t)^\kappa$ for some $\kappa > 0$; polynomial rates $\phi(t) = t^\kappa$ for some $\kappa \in (0, 1)$; subexponential rates $\phi(t) = t/(1 + \log t)^\kappa$ for some $\kappa > 0$. Note that since $\phi \in \mathbb{F}$, the first term in the RHS of the bound in (i) is not the leading term (for fixed x, when $n \to \infty$). In the case ϕ is logarithmic or polynomial, the leading term in the RHS is the second one so that the rate of decay is given by $1/\phi(H_\phi^\leftarrow(n/2))$. For the logarithmic and polynomial cases, the best rates are given by Theorem 3(i) and for the subexponential case, by Theorem 3(ii).

In practice, it is often easier to establish a drift inequality on E rather than on $E \times E$ as in H2(Δ, ϕ, V). Theorem 4 relates the following single drift condition to the drift H2. For a function $\phi \in \mathbb{F}$, a measurable function $V : E \to [1, +\infty)$ and a constant $b \geq 0$, consider the following assumption

Table 1
Rates of convergence when ϕ increases at a logarithmic rate, a polynomial rate and a subexponential rate, obtained from Theorem 3 and from [6, Theorem 2.1] and [9, Section 2.3]

Order of the rates of convergence in	$\phi(x)=(1+\log(x))^{\kappa}$ for $\kappa>0$	$\phi(x)=x^{\kappa}$ for $\kappa\in(0,1)$ set $\varsigma=\kappa/(1-\kappa)$	$\phi(x)=x/(1+\log(x))^{\kappa}$ for $\kappa>0$ set $\varsigma=1/(1+\kappa)$
Theorem 3	$1/\log^{\kappa}(n)$	$1/n^{\varsigma}$	$\exp(-\delta((1+\kappa)n)^{\varsigma})$ for all $\delta\in(0,1)$
[9]	$1/\log^{\kappa}(n)$	$1/n^{\varsigma}$	$n^{\kappa\varsigma}\exp(-((1+\kappa)n)^{\varsigma})$
[6] for all $\delta\in(0,1)$	$1/\log^{\delta\kappa}(n)$	$1/n^{\delta\varsigma}$	$\exists C>0$ $\exp(-Cn^{\varsigma})$

H3 (ϕ, V, b). $\phi(0)=0$ *and for all* $x\in E$,

$$PV(x)\le V(x)-\phi\circ V(x)+b. \tag{7}$$

Theorem 4. *Let* $\phi\in\mathbb{F}$, *a measurable function* $V: E\to[1,+\infty)$ *and a constant* $b\ge 0$ *such that* H3(ϕ, V, b) *holds. Then* H2$(\{V\le \upsilon\}^2, c\phi, V)$ *is satisfied for any* $\upsilon>\phi^{\leftarrow}(2b)$ *and with* $c=1-2b/\phi(\upsilon)$.

The proof is postponed in Section 4.4. Note that we can assume without loss of generality that $t\mapsto\phi(t)$ is concave increasing and continuously differentiable only for large t; see Lemma 21.

Our assumptions and results can be compared to [6] which also establish convergence in Wasserstein distance at a subgeometric rate under the single drift condition H3(ϕ, V, b) and the following assumptions

- B(i) For all $x, y\in E$, $W_d(P(x,\cdot), P(y,\cdot))\le d(x,y)$.
- B(ii) There exists $\eta>0$ such that the level set $\Delta=\{(x,y): V(x)+V(y)\le\phi^{\leftarrow}(2b)+\eta\}$ is d-small for P: there exists $\varepsilon>0$ such that for any $x,y\in\Delta$, $W_d(P(x,\cdot),P(y,\cdot))\le(1-\varepsilon)d(x,y)$.

Under these conditions, [6, Theorem 2.1] shows the existence and uniqueness of the stationary distribution π and provides rates of convergence to stationarity in the Wasserstein distance W_d; expressions for these rates are provided in the last row of Table 1 for various choices of functions ϕ. It can be seen that our results always improve the rates of convergence when compared to those of [6].

Let us compare the assumptions of Theorem 3 to (B). It follows from [26, Corollary 5.22] that under B(i) and B(ii), there exists a coupling kernel for P (which is the coupling kernel realizing the lower bound in the Monge–Kantorovitch duality theorem) such that H1$(\Delta, 1, \varepsilon)$ holds. Since Theorem 4 establishes that a single drift condition of the form H3 implies a drift condition of the form H2, the assumptions of [6, Theorem 2.1] essentially differ from the assumptions of Theorem 3 through the coupling set assumption: [6, Theorem 2.1] only covers coupling sets of order 1 when our result covers coupling sets of order ℓ, for any $\ell\ge 1$. This is an unnatural and sometimes annoying restriction since in practical examples the order ℓ is most likely to be large (see e.g. the examples in Section 3). Note that the strategy consisting in applying a result for a coupling set of order 1 to the ℓ-iterated kernel is not equivalent to applying a result for a coupling set of order ℓ to the one iterated kernel; we provide an illustration of this claim in Section 3.1. Checking H1$(\Delta,\ell,\varepsilon)$ is easier than checking (B) since allowing the coupling set to be of any order provides far more flexibility.

Our results can also be compared to the explicit rates in [9] derived for convergence in total variation distance. In [9], it is assumed that P is phi-irreducible, aperiodic, that the drift condition H3 holds and that the level sets $\{V\le\upsilon\}$ are small in the usual sense, i.e. for some $\ell\in\mathbb{N}^*$, $\varepsilon\in(0,1)$ and a probability ν that may depend upon the level set, $P^{\ell}(x,A)\ge\varepsilon\nu(A)$, for all $x\in\{V\le\upsilon\}$ and $A\in\mathcal{B}(E)$. Under these assumptions, [9, Proposition 2.5] shows that for any $x\in E$, $\lim_{n\to\infty}\phi(H_{\phi}^{\leftarrow}(n))W_{d_0}(P^n(x,\cdot),\pi)=0$, where W_{d_0} is the total variation distance. Table 1 displays the rate r_{ϕ} obtained in [9] (see penultimate row) and the rates given by Theorem 3 (row 2): our results coincide with [9] for the polynomial and logarithmic cases and the logarithm of the rate differs by a constant (which can be chosen arbitrarily close to one in our case) in the subexponential case. Nevertheless, we would like to stress that our conditions do not require ϕ-irreducibility and therefore apply in more general contexts.

3. Application

3.1. *A symmetric random walk Metropolis algorithm*

Let $E \stackrel{\text{def}}{=} \{k/4; k \in \mathbb{Z}\}$ endowed with the trivial distance d_0, thus (E, d_0) is a Polish space. Consider a symmetric random walk Metropolis (SRWM) algorithm on E for an heavy tailed target distribution π given by

$$\pi(x) \propto 1/\big(1+|x|\big)^{1+h}, \quad \text{for all } x \in E, \tag{8}$$

where $h \in (0, 1/2)$. Starting at $x \in E$, the Metropolis algorithm proposes at each iteration, a candidate y from a random walk with a symmetric increment distribution q on E. The move is accepted with probability $\alpha(x, y) = 1 \wedge (\pi(y)/\pi(x))$. The Markov kernel associated with the SRWM algorithm is given, for all $x \in E$ and $A \subset E$, by

$$P(x, A) = \sum_{y, x+y \in A} \alpha(x, x+y)q(y) + \delta_x(A) \sum_{y \in E} \big(1 - \alpha(x, x+y)\big)q(y).$$

Assume that q is the uniform distribution on $\{-1/4, 0, 1/4\}$. It is easily checked that P is irreducible and aperiodic. In the following, we prove that [6, Theorem 2.1] cannot be applied to this case, contrary to Theorem 3.

We first prove that P cannot be geometrically ergodic. The proof essentially follows from [16, Theorem 2.2], where the authors established necessary and sufficient conditions for the geometric and the polynomial ergodicity of random walk type Markov chains on $\mathbb{R}$.

Proposition 5. *P is not geometrically ergodic.*

Proof. The proof is by contradiction: we assume that P is geometrically ergodic. Since it is also P irreducible and aperiodic, the stationary distribution π is unique and geometrically regular: for any set A such that $\pi(A) > 0$, there exists $L > 1$ such that $\mathbb{E}_\pi[L^{\tau_A}] = \sum_{x \in E} \pi(x)\mathbb{E}_x[L^{\tau_A}] < \infty$, where τ_A is the return time to A. Choose $M > 0$, $A = \{x \in E, |x| \le M\}$. Since for $|x| \ge M$, $\tau_A \ge 4(|x| - M)$ $\mathbb{P}_x$-a.s. the regularity of π claims that there exists $L > 1$ such that $\sum_{x \in \mathbb{Z}} L^{|x|}\pi(x) < \infty$. This clearly yields to a contradiction. $\square$

We then show that the Markov kernel P satisfies a sub-geometric drift condition. For $s \ge 0$, set $V_s(x) = 1 \vee |x|^s$.

Proposition 6. *For all $s \in (2, 2+h)$, there exist $b, c > 0$ such that for all $x \in E$*

$$PV_s(x) \le V_s(x) - cV_s(x)^{(s-2)/s} + b. \tag{9}$$

Proof. We have for all $x \ge 5/4$,

$$PV_s(x) - V_s(x) = \big(x^s/3\big)\Big(\Big(\big(1-(4x)^{-1}\big)^s - 1\Big) - \big(1 - 1/(5+4x)\big)^{1+h}\Big(1 - \big(1+(4x)^{-1}\big)^s\Big)\Big).$$

Since $(1-(4x)^{-1})^s - 1 = -s/(4x) - (1-s)s/(32x^2) + o(x^{-2})$ and $(1 - 1/(5+4x))^{1+h} = 1 - (1+h)/(4x) + (10 + 11h + h^2)/(32x^2) + o(x^{-2})$ as $x \to +\infty$, then $PV_s(x) - V_s(x) = x^{s-2}s(s-h-2)/48 + o(x^{s-2})$. The same expansion remains valid as $x \to -\infty$ upon replacing x by $-x$. $\square$

Using this result, [9, Proposition 2.5] shows that for any $x \in E$, $P^n(x, \cdot)$ converges to π in total variation norm, at the rates $n^{\tilde{h}}$ for all $\tilde{h} \in (0, h/2)$.

We can also apply Theorem 4 and Theorem 3(i). For any $s \in (2, 2+h)$, H3(ϕ_s, V_s, b) is satisfied with $\phi_s(x) = cx^{(s-2)/s}$, $V_s(x) = 1 \vee |x|^s$ and $b < +\infty$. For $x, y \in E$ and $A, B \subset E$, consider the following kernel:

$$Q\big((x, y), (A \times B)\big) = P(x, A)P(y, B)\mathbb{1}_{\{x \ne y\}} + P(x, A \cap B)\mathbb{1}_{\{x = y\}}.$$

Clearly, Q is a coupling kernel for P. Let us prove that for any $M > 0$ and any $\ell \geq 4M$, there exists $\varepsilon > 0$ such that H1$(\Delta, \ell, \varepsilon)$ holds with $\Delta = \{|x| \vee |y| \leq M\}$. We have $Qd_0(x, y) \leq d_0(x, y)$ for every $x \neq y \in E$ and by definition of Q, $Qd_0(x, x) = 0$ for every $x \in E$. Let $M > 0$, $\ell \geq 4M$. For any $x, y \in \{|x| \vee |y| \leq M\}$ such that $|x| < |y|$

$$\widetilde{\mathbb{P}}_{x,y}[X_\ell = Y_\ell] \geq \widetilde{\mathbb{P}}_{x,y}[X_{4|y|} = Y_{4|y|}] \geq \widetilde{\mathbb{P}}_{x,y}\big[\tau_0^X = 4|x|, X_{4|x|+1} = 0, \dots, X_{4|y|} = 0, \tau_0^Y = 4|y|\big],$$

where $\tau_0^X = \inf\{n \geq 1, X_n = 0\}$ and $\tau_0^Y = \inf\{n \geq 1, Y_n = 0\}$. Since

$$\widetilde{\mathbb{P}}_{x,y}\big[\tau_0^X = 4|x|\big] \geq (1/3)^{4|x|}, \qquad \widetilde{\mathbb{P}}_{x,y}\big[\tau_0^Y = 4|y|\big] \geq (1/3)^{4|y|},$$
$$\widetilde{\mathbb{P}}_{x,y}[X_{4|x|+1} = 0, \dots, X_{4|y|} = 0] \geq (1/3)^{4(|y|-|x|)},$$

it follows that $Q^\ell d_0(x, y) = 1 - \widetilde{\mathbb{P}}_{x,y}[X_\ell = Y_\ell] \leq 1 - (1/3)^{8|y|} \leq 1 - (1/3)^{8M} d_0(x, y)$. This inequality remains valid when $x = y$. This concludes the proof of H1$(\Delta, \ell, \varepsilon)$. By Theorem 4, the kernel P is subgeometrically ergodic in total variation distance at the rates $n^{\tilde{h}}$, for $\tilde{h} \in (0, h/2)$.

In this example, [6, Theorem 2.1] cannot be applied. Indeed, on one hand, for any $M > 0$ the set $\Delta_M = \{|x| \vee |y| \leq M\}$ is a $(1, \varepsilon, d_0)$-coupling set for P^ℓ iff $l \geq 4M$. This property is a consequence of the above discussion (for the converse implication) and of the equality $W_{d_0}(P^\ell(x, \cdot), P^\ell(y, \cdot)) = 1$ if $|x - y| > \ell/2$ (for the direct implication). On the other hand, in order to check B(ii) for some ℓ-iterated kernel P^ℓ, we have to prove that there exists $\eta > 0$ such that $\Delta_\star = \{(x, y) \in E^2; V_s(x) + V_s(y) \leq (2b\ell/c)^{s/(s-2)} + \eta\}$ is a $(1, \varepsilon, d_0)$-coupling set for P^ℓ – the constants b, c are given by Proposition 6. Unfortunately, since $b/c \geq 1$ (apply the drift inequality (9) with $x = 0$), and $1/(s - 2) \geq 2$, we get

$$\big\{(x, y) \in E, |x| \vee |y| \leq 4\ell^2\big\} \subset \big\{x, y \in E; |x| \vee |y| \leq (2b\ell/c)^{1/(s-2)}\big\} \subset \Delta_\star.$$

Therefore whatever ℓ, $\Delta_\star$ is not a $(1, \varepsilon, d_0)$ coupling set for P^ℓ.

3.2. *Non linear autoregressive model*

In this section, we consider the functional autoregressive process $\{X_n, n \in \mathbb{N}\}$ on $E = \mathbb{R}^p$, given by $X_{n+1} = g(X_n) + Z_{n+1}$ Denote by $\|\cdot\|$ the Euclidean norm on E and $\mathrm{B}(x, M)$ the ball of radius $M \geq 0$ and centered at $x \in \mathbb{R}^p$, associated with this norm. Consider the following assumptions:

AR1. *$\{Z_n, n \in \mathbb{N}^*\}$ is an independent and identically distributed (i.i.d.) zero-mean $\mathbb{R}^p$-valued sequence, independent of X_0, and satisfying $\int \exp(\beta_0 \|z\|^{\kappa_0}) \mu(\mathrm{d}z) < +\infty$, where μ is the distribution of Z_1 for some $\beta_0 > 0$ and $\kappa_0 \in (0, 1]$.*

AR2. *For all $M > 0$, $g : \mathbb{R}^p \to \mathbb{R}^p$ is C_M-Lipschitz on $\mathrm{B}(0, M)$ with respect to $\|\cdot\|$ where $C_M \in (0, 1)$. Furthermore, there exist positive constants r, M_0, and $\rho \in [0, 2)$, such that $\|g(x)\| \leq \|x\|(1 - r\|x\|^{-\rho})$ if $\|x\| \geq M_0$.*

A simple example of function g satisfying AR2 is $x \mapsto x \cdot \max(1/2, 1 - 1/\|x\|^\rho)$ with $\rho \in [0, 2)$. Denote by P the Markov kernel defined by the process $(X_n)_n$. Proposition 7 establishes H3(ϕ, V, b) in the case where $\rho > \kappa_0$, and a geometric drift condition in the other case.

Proposition 7 ([9, Theorem 3.3]). *Assume* AR1 *and* AR2.

(i) *If $\rho > \kappa_0$, there exist $\beta \in (0, \beta_0)$ and $b, c > 0$ such that* H3(ϕ, V, b) *holds with*

$$\phi(x) := cx\big(1 + \log(x)\big)^{1-\rho/(\kappa_0 \wedge (2-\rho))} \quad \textit{and} \quad V(x) := \exp\big(\beta\|x\|^{\kappa_0 \wedge (2-\rho)}\big).$$

(ii) *If $\rho \leq \kappa_0$, then there exist $b < +\infty$ and $\zeta \in (0, 1)$ such that for all $x \in \mathbb{R}^p$, $PV(x) \leq \zeta V(x) + b$ where $V(x) = \exp(\beta\|x\|^{\kappa_0})$ with $\beta \in (0, \beta_0)$.*

Proof. The proof of Proposition 7 is along the same lines as [9, Theorem 3.3] and is omitted.[1] □

Consider the coupling kernel Q defined for all $x, y \in E$ and $A \in \mathcal{B}(E \times E)$ by

$$Q\big((x, y), A\big) = \int \mathbb{1}_A\big(g(x) + z, g(y) + z\big)\mu(\mathrm{d}z). \tag{10}$$

For $\eta > 0$, define $d_\eta(x, y) \stackrel{\text{def}}{=} 1 \wedge \eta^{-1}\|x - y\|$.

Proposition 8. *Assume* AR1 *and* AR2. *For any* $M > 0$, *there exist* $\varepsilon, \eta > 0$ *such that* $\mathrm{B}(0, M) \times \mathrm{B}(0, M)$ *is a* $(1, \varepsilon, d_\eta)$*-coupling set.*

Proof. Since $d_\eta(x, y) = \|x - y\|/\eta$ for any $x, y \in \mathrm{B}(0, M)$ and $\eta = 2M$, we get under AR2,

$$\mathbb{E}\big[d_\eta\big(g(x) + Z_1, g(y) + Z_1\big)\big] \leq \eta^{-1}\big\|g(x) - g(y)\big\| \wedge 1 \leq C_M \eta^{-1}\|x - y\| \leq C_M d_\eta(x, y). \tag{11}$$

Finally, since AR2 implies that g is 1-Lipschitz on $\mathbb{R}^p$, (11) shows that $\mathbb{E}[d_\eta(g(x) + Z_1, g(y) + Z_1)] \leq d_\eta(x, y)$ for all $x, y \in \mathbb{R}^p$. □

For all $\eta, \eta' > 0$, d_η and $d_{\eta'}$ are Lipschitz equivalent, i.e., there exists $C > 0$ such that for all $x, y \in \mathbb{R}^p$, $C^{-1} d_\eta(x, y) \leq d_{\eta'}(x, y) \leq C d_\eta(x, y)$, which implies (see (1)) that W_{d_η} and $W_{d_{\eta'}}$ are Lipschitz equivalent.

Theorem 9. *Assume* AR1 *and* AR2 *hold. Then* P *admits a unique invariant distribution* π.

(i) *If* $\rho > \kappa_0$, *there exist two constants* C_1 *and* C_2 *such that for all* $x \in \mathbb{R}^p$ *and* $n \in \mathbb{N}^*$

$$W_{d_1}\big(P^n(x, \cdot), \pi\big) \leq C_1 V(x) \exp\big(-C_2 n^{\varsigma}\big),$$

where $\varsigma = (\kappa_0 \wedge (2 - \rho))/\rho$.

(ii) *If* $\rho \leq \kappa_0$, *then there exist* $\tilde{\zeta} \in (0, 1)$ *and a constant* C *such that for all* $x \in \mathbb{R}^p$ *and* $n \in \mathbb{N}^*$

$$W_{d_1}\big(P^n(x, \cdot), \pi\big) \leq C V(x) \tilde{\zeta}^n.$$

Proof. By application of Corollary 2, Theorem 3 and Theorem 4, we deduce (i) from Proposition 7(i) and Proposition 8. By an application of [13, Theorem 4.8, Corollary 4.11], we deduce (ii) from Proposition 7(ii) and Proposition 8. □

Perhaps surprisingly, we cannot relax the condition $\kappa_0 \in (0, 1]$, to obtain geometric convergence for $1 < \rho \leq \kappa_0$. Indeed, [22, Theorem 3.2(a)] provides an example where AR1 and AR2 are satisfied for $\kappa_0 = 2$ and $\rho \in (1, 2)$, but the chain fails to be geometrically ergodic (for the total variation distance).

3.3. *The preconditioned Crank–Nicolson algorithm*

In this section, we consider the preconditioned Crank–Nicolson algorithm (Algorithm 1) introduced in [3] and analyzed in [14] for sampling in a separable Hilbert space $(\mathcal{H}, \|\cdot\|)$ a distribution with density $\pi \propto \exp(-g)$ with respect to a zero-mean Gaussian measure γ with covariance operator C; see [5]. This algorithm is studied in [14] under conditions which imply the geometric convergence in Wasserstein distance.

We consider the convergence of the Crank–Nicolson algorithm under the weaker condition CN1 below for which the results in [14] cannot be applied. We will show that subgeometric convergence can nevertheless be obtained.

[1] We point out that in [9], it is additionally required that the distribution of Z_1 has a nontrivial absolutely continuous component which is bounded away from zero in a neighborhood of the origin. However, this condition is only required to establish the ϕ-irreducibility of the Markov chain, which is not needed here.

Algorithm 1: Preconditioned Crank–Nicolson algorithm

```
Data: ρ ∈ [0, 1)
Result: {X_n, n ∈ ℕ}
begin
    Initialize X_0
    for n ≥ 0 do
        Generate Z_{n+1} ~ γ
        Generate U_{n+1} ~ U([0, 1])
        if U_{n+1} ≤ α(X_n, ρX_n + √(1 − ρ²) Z_{n+1}) = 1 ∧ exp(g(X_n) − g(ρX_n + √(1 − ρ²) Z_{n+1})) then
            X_{n+1} = ρX_n + √(1 − ρ²) Z_{n+1}
        else
            X_{n+1} = X_n
```

CN 1. *The function* $g : \mathcal{H} \to \mathbb{R}$ *is* β*-Hölder for some* $\beta \in (0, 1]$ *i.e., there exists* C_g*, such that for all* $x, y \in \mathcal{H}$*,* $|g(x) - g(y)| \le C_g \|x - y\|^\beta$*.*

Examples of densities satisfying CN1 are $g(x) = -\|x\|^\beta$ with $\beta \in (0, 1]$. The following Theorem implies that under CN1, $\exp(-g)$ is γ-integrable (see [5, Theorem 2.8.5]).

Theorem 10 (Fernique's theorem). *There exist* $\theta \in \mathbb{R}_+^*$ *and a constant* C_θ *such that* $\int_{\mathcal{H}} \exp(\theta\|\xi\|^2)\,\mathrm{d}\gamma(\xi) \le C_\theta$*.*

The Crank–Nicolson kernel P_{cn} has been shown to be geometrically ergodic by [14] under the assumptions that g is globally Lipschitz and that there exist positive constants C, M_1, M_2 such that for $x \in \mathcal{H}$ with $\|x\| \ge M_1$, $\inf_{z \in \overline{\mathrm{B}}(\rho x, M_2)} \exp(g(x) - g(z)) \ge C$ (see [14, Assumption 2.10–2.11]), where we denote by $\mathrm{B}(x, M)$ the open ball centered at $x \in \mathcal{H}$ and of radius $M > 0$ associated with $\|\cdot\|$, and by $\overline{\mathrm{B}}(x, M)$ its closure. Such an assumption implies that the acceptance ratio $\alpha(x, \rho x + \sqrt{1-\rho^2}\xi)$ is bounded from below as $\|x\| \to \infty$ uniformly on $\xi \in \overline{\mathrm{B}}(0, M_2/\sqrt{1-\rho^2})$. In CN1, this condition is weakened in order to address situations in which the acceptance-rejection ratio vanishes when $\|x\| \to \infty$: this happens when $\lim_{\|x\| \to +\infty}\{g(\rho x) - g(x)\} = +\infty$. We first check that H3(ϕ, V, b) is satisfied with

$$V(x) = \exp\big(s\|x\|^2\big), \tag{12}$$

where $s = (1-\rho)^2\theta/16$ and θ is given by Theorem 10.

Proposition 11. *Assume* CN1*, and let* $\rho \in [0, 1)$*. Then there exist* $b \in \mathbb{R}_+$ *and* $c \in (0, 1)$ *such that for all* $x \in \mathcal{H}$

$$P_{\mathrm{cn}}V(x) \le V(x) - \phi \circ V(x) + b,$$

where $\phi \in \mathbb{F}$ *and* $\phi(t) \sim_{t \to \infty} ct\exp(-\{\log(t)/\kappa\}^{\beta/2})$*, with* $\kappa = \theta C_g^{-2/\beta}/36$.

Proof. The proof is postponed to Section 5.1. □

We now deal with showing H1. To that goal, we introduce the distance $d_\eta(x, y) = 1 \wedge \eta^{-1}\|x - y\|^\beta$, for any $\eta > 0$, and for $x, y \in E$ the basic coupling Q_{cn} between $P_{\mathrm{cn}}(x, \cdot)$ and $P_{\mathrm{cn}}(y, \cdot)$: the same Gaussian variable Ξ and the same uniform variable U are generated to build X_1 and Y_1, with initial conditions x, y. Define $\Lambda_{(x,y)}(z) = (\rho x + \sqrt{1-\rho^2}z, \rho y + \sqrt{1-\rho^2}z)$ and $\widetilde{\gamma}_{(x,y)}$ the pushforward of γ by $\Lambda_{(x,y)}$. Then an explicit form of Q_{cn} is given, for

$A \in \mathcal{B}(\mathcal{H} \times \mathcal{H})$, by:

$$\begin{aligned} Q_{\mathrm{cn}}\big((x,y),A\big) = {} & \int_A \alpha(x,v) \wedge \alpha(y,t)\,\mathrm{d}\widetilde{\gamma}_{(x,y)}(v,t) + \int_{\mathcal{H}\times\mathcal{H}} \big(\alpha(y,t)-\alpha(x,v)\big)_+ \mathbb{1}_A(x,t)\,\mathrm{d}\widetilde{\gamma}_{(x,y)}(v,t) \\ & + \int_{\mathcal{H}\times\mathcal{H}} \big(\alpha(x,v)-\alpha(y,t)\big)_+ \mathbb{1}_A(v,y)\,\mathrm{d}\widetilde{\gamma}_{(x,y)}(v,t) \\ & + \delta_{(x,y)}(A) \int_{\mathcal{H}\times\mathcal{H}} \big(1-\alpha(x,v) \vee \alpha(y,t)\big)\,\mathrm{d}\widetilde{\gamma}_{(x,y)}(v,t), \end{aligned} \tag{13}$$

where for $u \in \mathbb{R}$, $(u)_+ = \max(u,0)$. The following Proposition shows that H1 is satisfied.

Proposition 12. *Assume* CN1. *There exists $\eta > 0$ such that, Q_{cn} is a d_η-weak contraction and for every $u > 1$, there exist $\ell \geq 1$ and $\varepsilon > 0$ such that $\{V \leq u\}^2$ is a $(\ell, \varepsilon, d_\eta)$-coupling set.*

Proof. See Section 5.2. □

Note that for all $\eta > 0$, d_η is Lipschitz equivalent to d_1, therefore W_{d_η} and W_{d_1} are Lipschitz equivalent. As a consequence of Proposition 11, Proposition 12, Theorem 3 and Theorem 4, we have

Theorem 13. *Let P_{cn} be the kernel of the preconditioned Crank–Nicolson algorithm with target density $\mathrm{d}\pi \propto \exp(-g)\,\mathrm{d}\gamma$ and design parameter $\rho \in [0,1)$. Assume* CN1. *Then P_{cn} admits π as a unique invariant probability measure and there exist C_1, C_2 such that for all $n \in \mathbb{N}^*$ and $x \in \mathcal{H}$*

$$W_{d_1}\big(P^n_{\mathrm{cn}}(x,\cdot),\pi\big) \leq C_1 V(x) \exp\big(-\kappa\big(\log(n) - C_2 \log\big(\log(n)\big)\big)^{2/\beta}\big),$$

where V is given by (12), *$d_1(x,y) = \|x-y\|^\beta \wedge 1$ and $\kappa = \theta C_g^{-2/\beta}/36$ for θ given by Theorem* 10.

Theorem 13 covers the case of the independent sampler (case $\rho = 0$). Both the rate of convergence through the constant C_2 and the control in the initial value x throught C_1 and the function V depends on ρ.

4. Proofs of Section 2

In this section, C is a constant which may take different values upon each appearance.

For $\Delta \in \mathcal{B}(E \times E)$, $\ell \in \mathbb{N}^*$ and a canonical Markov chain on the space $((E\times E)^{\mathbb{N}}, (\mathcal{B}(E) \otimes \mathcal{B}(E))^{\otimes \mathbb{N}})$, denote by $T_0 = \inf\{n \geq \ell, (X_n, Y_n) \in \Delta\}$ the first return time to Δ after $\ell - 1$ steps. Then, define recursively for $j \geq 1$,

$$T_j = T_0 \circ \theta^{T_{j-1}} + T_{j-1} = T_0 + \sum_{k=0}^{j-1} T_0 \circ \theta^{T_k}, \tag{14}$$

where θ is the shift operator.

Let Q be a coupling kernel for P. Hereafter, $\{(X_n, Y_n), n \in \mathbb{N}\}$ is the canonical Markov chain on the space $((E\times E)^{\mathbb{N}}, (\mathcal{B}(E) \otimes \mathcal{B}(E))^{\otimes \mathbb{N}})$ with Markov kernel Q. We denote by $\widetilde{\mathbb{P}}_{x,y}$ and $\widetilde{\mathbb{E}}_{x,y}$ the associated canonical probability and expectation, respectively, when the initial distribution of the Markov chain is the Dirac mass at (x,y).

For any $n \in \mathbb{N}^\star$ and $x, y \in E$, the n-iterated kernel $Q^n((x,y),\cdot)$ is a coupling of $(P^n(x,\cdot), P^n(y,\cdot))$; hence $W_d(P^n(x,\cdot), P^n(y,\cdot)) \leq \widetilde{\mathbb{E}}_{x,y}[d(X_n,Y_n)]$. Define the filtration $\{\widetilde{\mathcal{F}}_n, n \geq 0\}$ by $\widetilde{\mathcal{F}}_n = \sigma((X_k,Y_k), k \leq n)$.

Before proceeding to the actual derivation of the proofs, we present a roadmap of them. The key step for our results is given by the following inequality: for any $x, y \in E$ and $n, m \geq 1$,

$$W_d\big(P^n(x,\cdot), P^m(y,\cdot)\big) \leq B(n,m)\big(V(x) + V(y)\big), \tag{15}$$

with $\lim_{n,m\to+\infty} B(n,m) = 0$. Under the assumptions of Theorem 1, this inequality will imply that P admits at most one invariant probability. In addition, by applying (15) with $n \leftarrow n+m$, and $y \leftarrow x$, we show that $\{P^n(x,\cdot), n \in \mathbb{N}\}$ is a Cauchy sequence in $(\mathcal{P}(E), W_d)$ and therefore converges in W_d to some probability measure π_x which is shown to be invariant for P. Since P admits one invariant probability measure, then π_x does not depend on x (see Section 4.1). The proof of Corollary 2 consists in verifying that the assumptions of Theorem 1 are satisfied.

The proof of Theorem 3 also follows from (15), but an explicit expression of B is required (see Lemma 18). Taking $n = m$ and integrating this inequality w.r.t. the unique invariant distribution π will conclude the proof.

Let us now explain the computation of the upper bound (15). The contraction property of Q (see H1(i)) combined with the Markov property of $\{(X_n, Y_n), n \in \mathbb{N}\}$ imply that $\{d(X_n, Y_n), n \in \mathbb{N}\}$ is a supermartingale with respect to the filtration $\widetilde{\mathcal{F}}_n$; this property yields $\widetilde{\mathbb{E}}_{x,y}[d(X_n, Y_n)] \leq (1-\varepsilon)^{m-1} + \widetilde{\mathbb{P}}_{x,y}[T_m \geq n]$ for any $n, m \geq 0$. By the Markov inequality, for any increasing rate function R, it holds

$$\widetilde{\mathbb{E}}_{x,y}\big[d(X_n, Y_n)\big] \leq (1-\varepsilon)^{m-1} + \frac{\widetilde{\mathbb{E}}_{x,y}[R(T_m)]}{R(n)}. \tag{16}$$

The last step of the proof is to compute an upper bound for the moment $\widetilde{\mathbb{E}}_{x,y}[R(T_m)]$. Then m is chosen in order to balance the two terms in the RHS of (16).

We preface the proof of our results by the following result.

Proposition 14. *Assume that there exists a coupling kernel Q for P, $\Delta \in \mathcal{B}(E \times E)$, $\ell \in \mathbb{N}^*$ and $\varepsilon > 0$ such that* H1$(\Delta, \ell, \varepsilon)$ *holds. Then, for all $x, y \in E$, and $n \geq 0, m \geq 0$:*

$$\widetilde{\mathbb{E}}_{x,y}\big[d(X_n, Y_n)\big] \leq (1-\varepsilon)^m + \widetilde{\mathbb{P}}_{x,y}[T_m \geq n].$$

Proof. Set $Z_n = d(X_n, Y_n)$; under H1$(\Delta, \ell, \varepsilon)$, $\{(Z_n, \widetilde{\mathcal{F}}_n)\}_{n \geq 0}$ is a bounded non-negative supermartingale and for all $(x, y) \in \Delta$, $\widetilde{\mathbb{E}}_{x,y}[Z_\ell] \leq (1-\varepsilon) d(x, y)$. Denote by Z_∞ its $\widetilde{\mathbb{P}}_{x,y}$-a.s limit. By the optional stopping theorem, we have for every $m \geq 0$: $\widetilde{\mathbb{E}}_{x,y}[Z_{T_{m+1}} | \widetilde{\mathcal{F}}_{T_m+\ell}] \leq Z_{T_m+\ell}$. On the other hand, by the strong Markov property, $\widetilde{\mathbb{E}}_{x,y}[Z_{T_m+\ell} | \widetilde{\mathcal{F}}_{T_m}] \leq (1-\varepsilon) Z_{T_m}$. By combining these two relations, we get: $\widetilde{\mathbb{E}}_{x,y}[Z_{T_{m+1}} | \widetilde{\mathcal{F}}_{T_m}] \leq (1-\varepsilon) Z_{T_m}$. Since Z_n is upper bounded by 1, the proof follows from [17, Lemma 3.1]. □

4.1. *Proof of Theorem 1*

By Proposition 14 and the Markov inequality for all $m \geq 0$, we get

$$\widetilde{\mathbb{E}}_{x,y}\big[d(X_n, Y_n)\big] \leq (1-\varepsilon)^m + n^{-1} \widetilde{\mathbb{E}}_{x,y}[T_m]. \tag{17}$$

Using (14) and the strong Markov property, we obtain $\widetilde{\mathbb{E}}_{x,y}[T_m] = \widetilde{\mathbb{E}}_{x,y}[T_0] + \widetilde{\mathbb{E}}_{x,y}[\sum_{k=0}^{m-1} \widetilde{\mathbb{E}}_{X_{T_k}, Y_{T_k}}[T_0]]$. Using [19, Proposition 11.3.3] and the Markov property we have that

$$\widetilde{\mathbb{E}}_{x,y}[T_0] \leq Q^{\ell-1} \mathrm{V}(x, y) + b + \ell - 1,$$

which implies that $\widetilde{\mathbb{E}}_{x,y}[T_m] \leq m \sup_{(x,y) \in \Delta} Q^{\ell-1} \mathrm{V}(x, y) + Q^{\ell-1} \mathrm{V}(x, y) + (m+1)(b + \ell - 1)$, where the constant b is defined in (3). Plugging this inequality into (17) and taking $m = \lceil -\log(n)/\log(1-\varepsilon) \rceil$ implies that there exists $C < \infty$ satisfying

$$Q^n d(x, y) = \widetilde{\mathbb{E}}_{x,y}\big[d(X_n, Y_n)\big] \leq C\big(\log(n)/n\big) Q^{\ell-1} \mathrm{V}(x, y) \leq C\big(\log(n)/n\big) \mathrm{V}(x, y), \tag{18}$$

where we have used that $Q^{\ell-1} \mathrm{V}(x, y) \leq \mathrm{V}(x, y) + b(\ell - 1)$ (the constant C takes different values upon each appearance).

Uniqueness of the invariant probability

The proof is by contradiction. Assume that there exist two invariant distributions π and ν, and let $\lambda \in \mathcal{C}(\pi, \nu)$. According to Lemma 23(i), we have for every integer n,

$$W_d(\pi, \nu) = W_d\big(\pi P^n, \nu P^n\big) \leq \int_{E \times E} Q^n d(x, y) \lambda(\mathrm{d}x, \mathrm{d}y).$$

We prove that the RHS converges to zero by application of the dominated convergence theorem. It follows from (18) that for all $x, y \in E$ and $n \geq 0$, $g_n(x, y) \stackrel{\text{def}}{=} Q^n d(x, y) \leq C \mathrm{V}(x, y) \log(n)/n$ for some $C < \infty$. Therefore, the sequence of functions $\{g_n, n \in \mathbb{N}\}$ converges pointwise to 0. Since $d \leq 1$, $g_n(x, y) \leq 1$. Hence, by the Lebesgue theorem, $\int_{E\times E} g_n(x, y)\lambda(\mathrm{d}x, \mathrm{d}y) \underset{n\to+\infty}{\longrightarrow} 0$ showing that $W_d(\pi, \nu) = 0$, or equivalently $\nu = \pi$ since W_d is a distance on $\mathcal{P}(E)$.

Existence of an invariant measure

Let $x_0 \in E$. We first show that there exists $\{m_k, k \in \mathbb{N}\}$ such that $\{P^{m_k}(x_0, \cdot), k \in \mathbb{N}\}$ is a Cauchy sequence for W_d. Let $n, k \in \mathbb{N}^*$ and choose $M \geq 1$. By Lemma 23(i):

$$W_d\big(P^n(x_0, \cdot), P^{n+n_k}(x_0, \cdot)\big) \leq \inf_{\lambda \in \mathcal{C}(\delta_{x_0}, P^{n_k}(x_0,\cdot))} \left\{ \int_{E\times E} \mathbb{1}_{\{\mathrm{V}(z,t)\geq M\}} Q^n d(z, t)\lambda(\mathrm{d}z, \mathrm{d}t) + \int_{E\times E} \mathbb{1}_{\{\mathrm{V}(z,t)< M\}} Q^n d(z, t)\lambda(\mathrm{d}z, \mathrm{d}t) \right\}. \tag{19}$$

We consider separately the two terms. Set $M_\psi = \sup_k P^{n_k}[\psi \circ \mathrm{V}_{x_0}](x_0)$. Let $\lambda \in \mathcal{C}(\delta_{x_0}, P^{n_k}(x_0, \cdot))$. Since d is bounded by 1, we get

$$\begin{aligned} \int_{E\times E} \mathbb{1}_{\{\mathrm{V}(z,t)\geq M\}} Q^n d(z, t)\lambda(\mathrm{d}z, \mathrm{d}t) &\leq \int_{E\times E} \mathbb{1}_{\{\mathrm{V}(z,t)\geq M\}}\lambda(\mathrm{d}z, \mathrm{d}t) \leq P^{n_k}\big(x_0, \{\mathrm{V}_{x_0} \geq M\}\big) \\ &\leq P^{n_k}\big(x_0, \{\psi \circ \mathrm{V}_{x_0} \geq \psi(M)\}\big) \leq P^{n_k}[\psi \circ \mathrm{V}_{x_0}](x_0)/\psi(M) \\ &\leq M_\psi/\psi(M), \end{aligned} \tag{20}$$

where we have used (4) and the Markov inequality. In addition by (18), there exists $C > 0$ such that:

$$\int_{E\times E} \mathbb{1}_{\{\mathrm{V}(z,t)<M\}} Q^n d(z, t)\lambda(\mathrm{d}z, \mathrm{d}t) \leq C\big(\log(n)/n\big) \int_{E\times E} \mathbb{1}_{\{\mathrm{V}(z,t)<M\}} \mathrm{V}(z, t)\lambda(\mathrm{d}z, \mathrm{d}t).$$

Furthermore, $x \mapsto \psi(x)/x$ is non-increasing so that $\mathrm{V}(z, t) \leq M\psi(\mathrm{V}(z, t))/\psi(M)$ on $\{\mathrm{V}(z, t) \leq M\}$. This inequality and (4) imply

$$\int_{E\times E} \mathbb{1}_{\{\mathrm{V}(z,t)<M\}} Q^n d(z, t)\lambda(\mathrm{d}z, \mathrm{d}t) \leq C\big(\log(n)/n\big) M_\psi M/\psi(M). \tag{21}$$

Plugging (20) and (21) in (19), we have for every $M > 0$, $n, k \in \mathbb{N}^*$

$$W_d\big(P^n(x_0, \cdot), P^{n+n_k}(x_0, \cdot)\big) \leq \frac{M_\psi}{\psi(M)} + C\big(\log(n)/n\big)\big(M_\psi M/\psi(M)\big).$$

Setting $M = n/\log(n)$, we get that for all $n, k \in \mathbb{N}^*$

$$W_d\big(P^n(x_0, \cdot), P^{n+n_k}(x_0, \cdot)\big) \leq C/\psi\big(n/\log(n)\big). \tag{22}$$

Since $\lim_{x\to+\infty} \psi(x) = +\infty$ and $\lim_{k\to+\infty} n_k = +\infty$ there exists $\{u_k, k \in \mathbb{N}\}$ such that $u_0 = 1$ and for $k \geq 1$, $u_k = \inf\{n_l | l \in \mathbb{N}; \psi(n_l/\log(n_l)) \geq 2^k\}$. Set $m_k = \sum_{i=0}^k u_i$. Since for all $k \in \mathbb{N}$, $m_{k+1} = m_k + u_{k+1}$, by (22), $W_d(P^{m_k}(x_0, \cdot), P^{m_{k+1}}(x_0, \cdot)) \leq C2^{-k}$, which implies that the series $\sum_k W_d(P^{m_k}(x_0, \cdot), P^{m_{k+1}}(x_0, \cdot))$ converges and $\{P^{m_k}(x_0, \cdot), k \in \mathbb{N}\}$ is a Cauchy sequence in $(\mathcal{P}(E), W_d)$.

Since $(\mathcal{P}(E), W_d)$ is Polish, there exists $\pi \in \mathcal{P}(E)$ such that $\lim_{k\to+\infty} W_d(P^{m_k}(x_0, \cdot), \pi) = 0$. The second step is to prove that π is invariant. Since $\lim_{k\to+\infty} W_d(P^{m_k}(x_0, \cdot), \pi) = 0$, by the triangular inequality it holds

$$W_d(\pi, \pi P) \leq \lim_{k\to+\infty} W_d\big(P^{m_k}(x_0, \cdot), \delta_{x_0} P P^{m_k}\big) + \lim_{k\to+\infty} W_d\big(\delta_{x_0} P^{m_k} P, \pi P\big). \tag{23}$$

By Lemma 23(i) and (18), there exists C such that for any $k \geq 1$,

$$W_d\big(P^{m_k}(x_0,\cdot),\delta_{x_0}P^{m_k+1}\big) \leq \inf_{\lambda\in\mathcal{C}(\delta_{x_0},\delta_{x_0}P)} \int_{E\times E} Q^{m_k} d(z,t)\,\mathrm{d}\lambda(z,t)$$
$$\leq C\big(\log(m_k)/m_k\big) \inf_{\lambda\in\mathcal{C}(\delta_{x_0},\delta_{x_0}P)} \int_{E\times E} \mathrm{V}(z,t)\lambda(\mathrm{d}z,\mathrm{d}t)$$
$$\leq C\big(\log(m_k)/m_k\big) P\mathrm{V}_{x_0}(x_0).$$

By definition, $\lim_k m_k = +\infty$ so that by (4), the RHS converges to 0 when $k \to +\infty$. In addition, by Lemma 23(ii), $W_d(\delta_{x_0}P^{m_k}P, \pi P) \leq W_d(P^{m_k}(x_0,\cdot),\pi)$, and this RHS converges to 0 by definition of π. Plugging these results in (23) yields $W_d(\pi,\pi P) = 0$, and therefore $\pi P = \pi$.

4.2. *Proof of Corollary* 2

We prove that the assumptions of Theorem 1 are satisfied. Set $\mathrm{V}(x,y) = 1 + (V(x)+V(y))/\phi(2)$. Since Q is a coupling for P, it holds

$$Q\mathrm{V}(x,y) = 1 + \big(1/\phi(2)\big)\big(PV(x)+PV(y)\big) \leq \mathrm{V}(x,y) - \frac{\phi(V(x)+V(y))}{\phi(2)} + \big(b/\phi(2)\big)\mathbb{1}_\Delta(x,y).$$

This yields the drift inequality (3) upon noting that ϕ is increasing and $V \geq 1$ so that $\phi(V(x)+V(y))/\phi(2) \geq 1$. By iterating this inequality, we have for any ℓ,

$$\sup_{(x,y)\in\Delta}\big\{Q^{\ell-1}\mathrm{V}(x,y)\big\} \leq \sup_{(x,y)\in\Delta}\big\{\mathrm{V}(x,y)\big\} + b(\ell-1)/\phi(2),$$

and the RHS is finite since by assumption, $\sup_{(x,y)\in\Delta}\{V(x)+V(y)\} < \infty$.

Let $x_0 \in E$. Under H2(Δ,ϕ,V), $PV(x) \leq PV(x) + PV(x_0) \leq V(x) - \phi\circ V(x) + b + V(x_0)$ where we have used that $\phi(V(x)+V(x_0)) \geq \phi(V(x))$. This implies that for every $n \in \mathbb{N}^*$, $n^{-1}\sum_{k=0}^{n-1} P^k(\phi\circ V)(x) \leq b + V(x_0) + V(x)/n$. For any x, we have $P\mathrm{V}_x(x) < \infty$. Finally, since $\phi \in \mathbb{F}$, we can set $\psi = \phi$. Let us define the increasing sequence $\{n_k, k\in\mathbb{N}\}$. Set $M_\phi > b + V(x_0)$; there exists an increasing sequence $\{n_k, k\in\mathbb{N}\}$ such that $\lim_k n_k = +\infty$ and

$$P^{n_k}(\phi\circ V)(x_0) \leq M_\phi, \quad \text{for all } k\in\mathbb{N}. \tag{24}$$

Finally, [6, Lemma 4.1] implies $\int_E \phi\circ V(x)\pi(\mathrm{d}x) < \infty$.

4.3. *Proof of Theorem* 3

We preface the proof by some preliminary technical results. By using Proposition 14, for every $x,y\in E$ and $m\geq 0$, $\widetilde{\mathbb{E}}_{x,y}[d(X_n,Y_n)] \leq (1-\varepsilon)^m + \widetilde{\mathbb{P}}_{x,y}[T_m > n]$. The crux of the proof is to obtain estimates of tails of the successive return times to Δ. Following [24], we start by considering a sequence of drift conditions on the product space $E\times E$. For $\Delta \in \mathcal{B}(E\times E)$, $\ell\in\mathbb{N}^*$, a sequence of measurable functions $\{\mathrm{V}_n, n\in\mathbb{N}\}$, $\mathrm{V}_n : E\times E \to \mathbb{R}_+$, a function $r\in\Lambda$ and a constant $b<\infty$, let us consider the following assumption:

A$(\Delta,\ell,\mathrm{V}_n,r,b)$. *For all $x,y\in E$:*

$$Q\mathrm{V}_{n+1}(x,y) \leq \mathrm{V}_n(x,y) - r(n) + br(n)\mathbb{1}_\Delta(x,y) \quad \textit{and} \quad \sup_{(x,y)\in\Delta} Q^{\ell-1}\mathrm{V}_0(x,y) < \infty.$$

Under A$(\Delta,\ell,\mathrm{V}_n,r,b)$, we first obtain bounds on the moments $\widetilde{\mathbb{E}}_{x,y}[R(T_0)]$ for $x,y\in E$ (see Proposition 15), where

$$R(t) = 1 + \int_0^t r(s)\,\mathrm{d}s, \quad t\geq 0. \tag{25}$$

We will then deduce bounds for $\widetilde{\mathbb{P}}_{x,y}[T_m \geq n]$ (see Lemma 17). Set

$$c_{1,r} = \sup_{k\in\mathbb{N}^*} R(k) \Big/ \sum_{i=0}^{k-1} r(i), \qquad c_{2,r} = \sup_{m,n\in\mathbb{N}} R(m+n)/\bigl\{R(m)R(n)\bigr\}. \tag{26}$$

It follows from Lemma 24 that these constants are finite.

Proposition 15. *Assume that there exist a coupling kernel Q for P, $\Delta \in \mathcal{B}(E\times E)$, $\ell \in \mathbb{N}^*$, a sequence of measurable functions $\{\mathrm{V}_n, n\in\mathbb{N}\}$, $\mathrm{V}_n : E\times E \to \mathbb{R}_+$, a function $r \in \Lambda$ and a constant $b<\infty$ such that $\mathrm{A}(\Delta, \ell, \mathrm{V}_n, r, b)$ is satisfied. Then, for any $x, y\in E$,*

$$\widetilde{\mathbb{E}}_{x,y}\bigl[R(T_0)\bigr] \leq c_{1,r}c_{2,r}R(\ell-1)\bigl\{Q^{\ell-1}\mathrm{V}_0(x,y) + br(0)\bigr\}, \tag{27}$$

and $\sup_{(z,t)\in\Delta} \widetilde{\mathbb{E}}_{z,t}[R(T_0)]$ is finite.

Proof. By [19, Proposition 11.3.2], $\widetilde{\mathbb{E}}_{x,y}[\sum_{k=0}^{\tau_\Delta - 1} r(k)] \leq \mathrm{V}_0(x,y) + br(0)$, where τ_Δ is the return time to Δ. Since $R(k) \leq c_{1,r}\sum_{p=0}^{k-1} r(p)$, the previous inequality provides a bound on $\widetilde{\mathbb{E}}_{x,y}[R(\tau_\Delta)]$. The conclusion follows from the Markov property upon noting that $R(T_0) \leq c_{2,r}R(\ell-1)R(\tau_\Delta \circ \theta^{\ell-1})$. □

Combining the strong Markov property, (14) and Proposition 15, it is easily seen that $\widetilde{\mathbb{E}}_{x,y}[T_m] < \infty$ for any $m\geq 0$ and $x, y\in E$. This yields the following result.

Corollary 16. *Assume that there exist a coupling kernel Q for P, $\Delta \in \mathcal{B}(E\times E)$, $\ell \in \mathbb{N}^*$, a sequence of measurable functions $\{\mathrm{V}_n, n\in\mathbb{N}\}$, $\mathrm{V}_n : E\times E \to \mathbb{R}_+$, a function $r \in \Lambda$ and a constant $b<\infty$ such that $\mathrm{A}(\Delta, \ell, \mathrm{V}_n, r, b)$ is satisfied. Then, for all $j\geq 0$ and $(x,y)\in E\times E$, $\widetilde{\mathbb{P}}_{x,y}[T_j < \infty] = 1$.*

For $r\in\Lambda$, there exists $r_0 \in \Lambda_0$ such that $c_{3,r} = 1 \vee \sup_{t\geq 0} r(t)/r_0(t) < \infty$ and $c_{4,r} = 1\vee \sup_{t\geq 0} r_0(t)/r(t) < \infty$. Denote $c_{5,r} = \sup_{t,u\in\mathbb{R}_+} r(t+u)/\{r(t)r(u)\}$ and define for $\kappa > 0$, the real M_κ such that for all $t\geq M_\kappa$, $r(t)\leq \kappa R(t)$. M_κ is well defined by Lemma 24(iii).

Lemma 17. *Assume that there exist a coupling kernel Q for P, $\Delta \in \mathcal{B}(E\times E)$, $\ell \in \mathbb{N}^*$, a sequence of measurable functions $\{\mathrm{V}_n, n\in\mathbb{N}\}$, $\mathrm{V}_n : E\times E \to \mathbb{R}_+$, a function $r \in \Lambda$ and constants $\varepsilon > 0$, $b<\infty$ such that $\mathrm{H1}(\Delta,\ell,\varepsilon)$ and $\mathrm{A}(\Delta, \ell, \mathrm{V}_n, r, b)$ are satisfied. Then,*

(i) *for all $x, y \in E$ and for all $n\in\mathbb{N}$, $m\in\mathbb{N}^*$,*

$$\widetilde{\mathbb{P}}_{x,y}[T_m \geq n] \leq \bigl\{a_1 Q^{\ell-1}\mathrm{V}_0(x,y) + a_2\bigr\}/R(n/2) + a_3/R\bigl(n/(2m)\bigr),$$

(ii) *for all $\kappa > 0$, for all $x, y\in E$ and for all $n, m\in\mathbb{N}$,*

$$\widetilde{\mathbb{P}}_{x,y}[T_m \geq n] \leq (1+b_1\kappa)^m\bigl\{\kappa^{-1}r(M_\kappa) + a_1 Q^{\ell-1}\mathrm{V}_0(x,y) + a_2\bigr\}/R(n).$$

The constants $\{a_i\}_{i=1}^3$, b_1 can be directly obtained from the proof.

Proof. Since $r\in\Lambda$, there exists $r_0\in\Lambda_0$ such that $c_{3,r} + c_{4,r} < \infty$. Denote by R_0 the function (25) associated with r_0.

$$\begin{aligned}
\widetilde{\mathbb{P}}_{x,y}[T_m \geq n] &\leq \widetilde{\mathbb{P}}_{x,y}[T_0 \geq n/2] + \widetilde{\mathbb{P}}_{x,y}[T_m - T_0 \geq n/2] \\
&\leq \widetilde{\mathbb{E}}_{x,y}\bigl[R(T_0)\bigr]/R(n/2) + \widetilde{\mathbb{E}}_{x,y}\bigl[R_0\bigl((T_m - T_0)/m\bigr)\bigr]/R_0\bigl(n/(2m)\bigr) \\
&\leq \bigl\{a_1 Q^{\ell-1}\mathrm{V}_0(x,y) + a_2\bigr\}/R(n/2) + c_{3,r}\widetilde{\mathbb{E}}_{x,y}\bigl[R_0\bigl((T_m - T_0)/m\bigr)\bigr]/R\bigl(n/(2m)\bigr),
\end{aligned} \tag{28}$$

where we used Proposition 15 in the last inequality, and $a_1 = c_{1,r}c_{2,r}R(\ell-1)$; $a_2 = a_1 br(0)$. Since R_0 is convex (see Lemma 24), we have by (14):

$$\widetilde{\mathbb{E}}_{x,y}\big[R_0\big((T_m - T_0)/m\big)\big] \le c_{4,r}m^{-1}\widetilde{\mathbb{E}}_{x,y}\left[\sum_{k=0}^{m-1} R\big(T_0 \circ \theta^{T_k}\big)\right].$$

Using Corollary 16 and the strong Markov property, for any $x, y \in E$ and $m \ge 1$,

$$\widetilde{\mathbb{E}}_{x,y}\big[R_0\big((T_m - T_0)/m\big)\big] \le c_{4,r}C_\Delta, \quad \text{with } C_\Delta = \sup_{(x,y)\in\Delta} \widetilde{\mathbb{E}}_{x,y}\big[R(T_0)\big]. \tag{29}$$

Plugging (29) in (28) implies (i) with $a_3 = c_{3,r}c_{4,r}C_\Delta$. We now consider (ii). Again by the Markov inequality, since R is increasing,

$$\widetilde{\mathbb{P}}_{x,y}[T_m \ge n] \le R^{-1}(n)\widetilde{\mathbb{E}}_{x,y}\big[R(T_m)\big]. \tag{30}$$

If $m = 0$, the result follows from Proposition 15. If $m \ge 1$, using the definitions of T_m and R, given respectively in (14) and (25), and since for all $t, u \in \mathbb{R}_+$, $R(t+u) \le R(t) + c_{5,r}R(u)r(t)$, we get

$$\widetilde{\mathbb{E}}_{x,y}\big[R(T_m)\big] \le \widetilde{\mathbb{E}}_{x,y}\big[R(T_{m-1})\big] + c_{5,r}\widetilde{\mathbb{E}}_{x,y}\big[r(T_{m-1})R\big(T_0 \circ \theta^{T_{m-1}}\big)\big].$$

Thus, by the strong Markov property

$$\widetilde{\mathbb{E}}_{x,y}\big[R(T_m)\big] \le \widetilde{\mathbb{E}}_{x,y}\big[R(T_{m-1})\big] + c_{5,r}C_\Delta\widetilde{\mathbb{E}}_{x,y}\big[r(T_{m-1})\big]. \tag{31}$$

Let $\kappa > 0$. Since by definition, for all $t \ge M_\kappa$, $r(t) \le \kappa R(t)$, $\widetilde{\mathbb{E}}_{x,y}[r(T_{m-1})] \le r(M_\kappa) + \kappa\widetilde{\mathbb{E}}_{x,y}[R(T_{m-1})]$, so that (31) becomes

$$\widetilde{\mathbb{E}}_{x,y}\big[R(T_m)\big] \le (1 + c_{5,r}C_\Delta\kappa)\widetilde{\mathbb{E}}_{x,y}\big[R(T_{m-1})\big] + c_{5,r}C_\Delta r(M_\kappa).$$

By a straightforward induction we get,

$$\widetilde{\mathbb{E}}_{x,y}\big[R(T_m)\big] \le (1 + c_{5,r}C_\Delta\kappa)^m\big(\widetilde{\mathbb{E}}_{x,y}\big[R(T_0)\big] + r(M_\kappa)/\kappa\big).$$

Plugging this result in (30) and using Proposition 15 concludes the proof. Note that $b_1 = c_{5,r}C_\Delta$ and $a_2 = c_{1,r}c_{2,r}R(\ell-1)br(0)$. □

Lemma 18. *Assume that there exist a coupling kernel Q for P, $\Delta \in \mathcal{B}(E \times E)$, $\ell \in \mathbb{N}^*$, a sequence of measurable functions $\{\mathrm{V}_n, n \in \mathbb{N}\}$, $\mathrm{V}_n : E \times E \to \mathbb{R}_+$, a function $r \in \Lambda$ and constants $\varepsilon > 0$, $b < \infty$ such that* H1$(\Delta, \ell, \varepsilon)$ *and* A$(\Delta, \ell, \mathrm{V}_n, r, b)$ *are satisfied. Then,*

(i) *for all $x, y \in E$ and $n \in \mathbb{N}$,*

$$\widetilde{\mathbb{E}}_{x,y}\big[d(X_n, Y_n)\big] \le 1/R(n) + \big\{a_1 Q^{\ell-1}\mathrm{V}_0(x,y) + a_2\big\}/R(n/2) + a_3 v_n^{-1},$$

where $v_n \stackrel{\text{def}}{=} R(-n\log(1-\varepsilon)/\{2(\log(R(n)) - \log(1-\varepsilon))\})$,

(ii) *for all $\delta \in (0,1)$, $x, y \in E$ and $n \in \mathbb{N}$,*

$$\widetilde{\mathbb{E}}_{x,y}\big[d(X_n, Y_n)\big] \le \big(1 + (1 + b_1\kappa)\big\{\kappa^{-1}r(M_\kappa) + a_1 Q^{\ell-1}\mathrm{V}_0(x,y) + b_2\big\}\big)/R^\delta(n),$$

where $\kappa = ((1-\varepsilon)^{-(1-\delta)/\delta} - 1)/b_1$.

The constants a_i, b_j are given by Lemma 17.

Proof. By Proposition 14 and Lemma 17(i), there exist $\{a_i\}_{i=1}^3$ such that for all x, y in E and for all $n \geq 0$ and $m \geq 0$

$$\begin{aligned}\widetilde{\mathbb{E}}_{x,y}\big[d(X_n, Y_n)\big] &\leq (1-\varepsilon)^m + \widetilde{\mathbb{P}}_{x,y}[T_m \geq n] \\ &\leq (1-\varepsilon)^m + \big\{a_1 Q^{\ell-1}\mathrm{V}_0(x,y) + a_2\big\}/R(n/2) + a_3/R\big(n/(2m)\big).\end{aligned}$$

We get the first inequality by choosing $m = \lceil -\log(R(n))/\log(1-\varepsilon)\rceil$. Let us prove (ii). Fix $\delta \in (0,1)$ and choose the smallest integer m such that $(1-\varepsilon)^m \leq R(n)^{-\delta}$ (i.e. $m = \lceil -\delta \log R(n)/\log(1-\varepsilon)\rceil$). Apply Lemma 17(ii), with $\kappa > 0$ such that $(1 + b_1\kappa) = (1-\varepsilon)^{-((1-\delta)/\delta)}$; hence, upon noting that $R(n)^{-\delta} \leq (1-\varepsilon)^{m-1}$, it holds

$$(1+b_1\kappa)^m = (1+b_1\kappa)\big\{(1-\varepsilon)^{m-1}\big\}^{-((1-\delta)/\delta)} \leq (1+b_1\kappa)\big\{R(n)^{-\delta}\big\}^{-((1-\delta)/\delta)} = (1+b_1\kappa)R(n)^{1-\delta}. \qquad \square$$

We now prove that H2(Δ, ϕ, V) implies A. For a function $\phi \in \mathbb{F}$ and a measurable function $V : E \to [1,\infty)$, set

$$r_\phi(t) = \big(H_\phi^{\leftarrow}\big)'(t) = \phi\big(H_\phi^{\leftarrow}(t)\big), \tag{32}$$

where H_ϕ is defined in (6) and $H_\phi^{\leftarrow}$ denotes its inverse; and define for $k \geq 0$, $H_k : [1,\infty) \to \mathbb{R}_+$ and $\mathrm{V}_k : E \times E \to \mathbb{R}_+$ by

$$H_k(u) = \int_0^{H_\phi(u)} r_\phi(t+k)\,\mathrm{d}t = H_\phi^{\leftarrow}\big(H_\phi(u)+k\big) - H_\phi^{\leftarrow}(k), \tag{33}$$

$$\mathrm{V}_k(x,y) = H_k\big(V(x)+V(y)\big). \tag{34}$$

Note that V_k is measurable, H_k is twice continuously differentiable on $[1,\infty)$ and that $H_0(x) \leq x$ so $\mathrm{V}_0(x,y) \leq V(x)+V(y)$. The proof of the following lemma is adapted from [9, Proposition 2.1].

Lemma 19. *Assume that there exist $\Delta \in \mathcal{B}(E \times E)$, a function $\phi \in \mathbb{F}$ and a measurable function $V : E \to [1,\infty)$ such that* H2(Δ,ϕ,V) *is satisfied. For any $x, y \in E$ and any coupling $\lambda \in \mathcal{C}(P(x,\cdot), P(y,\cdot))$ we have*:

$$\int_{E\times E} \mathrm{V}_{k+1}(z,t)\,\mathrm{d}\lambda(z,t) \leq \mathrm{V}_k(x,y) - r_\phi(k) + \frac{b}{r_\phi(0)} r_\phi(k+1)\mathbb{1}_\Delta(x,y),$$

where r_ϕ and V_k are defined in (32) *and* (34) *respectively.*

Proof. Set $\mathrm{V}(x,y) = V(x)+V(y)$. By [9, Proposition 2.1] H_{k+1} is concave, which implies that for all $u \geq 1$ and $t \in \mathbb{R}$ such that $t+u \geq 1$, we have

$$H_{k+1}(t+u) - H_{k+1}(u) \leq H'_{k+1}(u)t. \tag{35}$$

In addition, according to the proof of [9, Proposition 2.1], for every $u \geq 1$ it holds:

$$H_{k+1}(u) - \phi(u)H'_{k+1}(u) \leq H_k(u) - r_\phi(k). \tag{36}$$

Therefore, the Jensen inequality and (5) imply

$$\begin{aligned}\int_{E\times E}\mathrm{V}_{k+1}(z,t)\,\mathrm{d}\lambda(z,t) &\leq H_{k+1}\bigg(\int_{E\times E}\mathrm{V}(z,t)\,\mathrm{d}\lambda(z,t)\bigg) \\ &\leq H_{k+1}\big(\mathrm{V}(x,y) - \phi\circ\mathrm{V}(x,y) + b\mathbb{1}_\Delta(x,y)\big).\end{aligned}$$

Using (35), (36) and the inequality $H'_{k+1}(\mathrm{V}(x,y)) \leq H'_{k+1}(1)$ we get that

$$\begin{aligned}\int_{E\times E}\mathrm{V}_{k+1}(z,t)\,\mathrm{d}\lambda(z,t) &\leq H_{k+1}\big(\mathrm{V}(x,y)\big) - \phi\circ\mathrm{V}(x,y)H'_{k+1}\big(\mathrm{V}(x,y)\big) + bH'_{k+1}(1)\mathbb{1}_\Delta(x,y) \\ &\leq H_k\big(\mathrm{V}(x,y)\big) - r_\phi(k) + bH'_{k+1}(1)\mathbb{1}_\Delta(x,y).\end{aligned}$$

The proof is concluded upon noting that $H'_{k+1}(1) = r_\phi(k+1)/r_\phi(0)$. $\square$

Proposition 20. *Assume that there exist a coupling kernel Q for P, $\Delta \in \mathcal{B}(E \times E)$, a function $\phi \in \mathbb{F}$ and a measurable function $V : E \to [1, \infty)$ such that* H2(Δ, ϕ, V) *is satisfied. Then for any $\ell \geq 0$,* A$(\Delta, \ell, \mathrm{V}_n, r_\phi, \{\sup_{p \geq 0} r_\phi(p+1)/r_\phi(p)\}b/r_\phi(0))$ *holds with $\mathrm{V}_n(x, y) = H_n(V(x) + V(y))$ where r_ϕ and H_n are given by* (32) *and* (33) *respectively.*

Proof. By [9, Lemma 2.3], $r_\phi \in \Lambda$. Then, it follows from Lemma 19 and Lemma 24(i) that for all $x, y \in E$,

$$Q\mathrm{V}_{k+1}(x, y) \leq \mathrm{V}_k(x, y) - r_\phi(k) + b\Big\{\sup_{p \geq 0} r_\phi(p+1)/r_\phi(p)\Big\} r_\phi(k)\mathbb{1}_\Delta(x, y)/r_\phi(0).$$

Finally, since Q^ℓ is a coupling kernel for P^ℓ, we have by iterating the inequality (5)

$$Q^\ell \mathrm{V}_0(x, y) \leq P^\ell V(x) + P^\ell V(y) \leq V(x) + V(y) + \ell b.$$

Therefore under H2(Δ, ϕ, V), $\sup_{(x,y)\in\Delta} Q^{\ell-1}\mathrm{V}_0(x, y) < +\infty$. □

Proof of Theorem 3. (i) Using Proposition 20, Lemma 18 applies with $R(t) = 1 + \int_0^t r_\phi(s)\,\mathrm{d}s$ for $t \in \mathbb{R}_+$. Note that we have $R = H_\phi^{\leftarrow}$.

Set $M_V > 0$ such that $\pi(V \leq M_V) \geq 1/2$; such a constant exists since $\pi(E) = 1$ and $E = \bigcup_{k \in \mathbb{N}}\{V \leq k\}$. Set $M > M_V$ and define the probability π_M by $\pi_M(\cdot) = \pi(\cdot \cap \{V \leq M\})/\pi(\{V \leq M\})$. Since π is invariant for P, $W_d(P^n(x, \cdot), \pi) = W_d(P^n(x, \cdot), \pi P^n)$ and the triangle inequality implies:

$$W_d\big(P^n(x, \cdot), \pi\big) \leq W_d\big(P^n(x, \cdot), \pi_M P^n\big) + W_d\big(\pi_M P^n, \pi P^n\big), \quad \text{for all } n \geq 1. \tag{37}$$

Consider the first term in the RHS of (37). By Lemma 23(i), for all $x \in E$ and $n \geq 1$:

$$W_d\big(P^n(x, \cdot), \pi_M P^n\big) \leq \inf_{\lambda \in \mathcal{C}(\delta_x, \pi_M)} \int_{E \times E} Q^n d(z, t)\,\mathrm{d}\lambda(z, t).$$

Let $v_n = R(-n\log(1-\varepsilon)/\{2(\log(R(n)) - \log(1-\varepsilon))\})$. By Lemma 18(i) and since $R = H_\phi^{\leftarrow}$ is increasing, for all $x \in E$ and $n \geq 1$

$$\begin{aligned} &R(n/2)W_d\big(P^n(x, \cdot), \pi_M P^n\big) \\ &\quad \leq R(n/2)/R(n) + a_1 \inf_{\lambda \in \mathcal{C}(\delta_x, \pi_M)} \int_{E \times E} \big(P^{\ell-1}V(z) + P^{\ell-1}V(t)\big)\,\mathrm{d}\lambda(z, t) + a_2 + a_3 R(n/2)/v_n \\ &\quad \leq a_1\bigg(V(x) + \int_E V(t)\,\mathrm{d}\pi_M(t) + b(\ell-1)\bigg) + a_2 + 1 + a_3 R(n/2)/v_n, \end{aligned} \tag{38}$$

where in the last inequality, we used

$$P^k V(x) \leq V(x) + bk/2, \tag{39}$$

which is obtained by iterating the drift inequality (5) and applying it with $x = y$. Since $x \mapsto \phi(x)/x$ is non-increasing, $V(t) \leq M\phi(V(t))/\phi(M)$ on $\{V \leq M\}$, we have

$$\int_E V(t)\,\mathrm{d}\pi_M(t) \leq 2\pi(\phi \circ V)M/\phi(M). \tag{40}$$

Note that by Corollary 2, $M_\phi = \int_E \phi \circ V(t)\,\mathrm{d}\pi(t) < \infty$. Combining (38) and (40) yield

$$W_d\big(P^n(x, \cdot), \pi_M P^n\big) \leq \big\{a_1\big(V(x) + 2M_\phi M/\phi(M) + b(\ell-1)\big) + a_2 + 1\big\}/R(n/2) + a_3/v_n. \tag{41}$$

Consider the second term in the RHS of (37). Since d is bounded by 1, $W_d(\mu, \nu) \leq W_{d_0}(\mu, \nu)$ (where W_{d_0} is the total variation distance) and Lemma 23(ii) implies $W_d(\pi_M P^n, \pi P^n) \leq W_d(\pi_M, \pi) \leq W_{d_0}(\pi_M, \pi)$. For every $A \in \mathcal{B}(E)$, we get

$$\big|\pi_M(A) - \pi(A)\big| = \big|\pi_M(A)\big(1 - \pi\big(\{V \leq M\}\big)\big) + \pi_M(A)\pi(V \leq M) - \pi(A)\big| \leq 2\pi\big(\{V > M\}\big),$$

showing that

$$W_d\big(\pi_M P^n, \pi P^n\big) \le 2\pi\big(\{V > M\}\big) = 2\pi\big(\big\{\phi(V) > \phi(M)\big\}\big) \le 2M_\phi/\phi(M). \tag{42}$$

Since $R(n/2) > M_V$ for all n large enough, we can now choose $M = R(n/2)$ in (41) and (42). This yields

$$W_d\big(P^n(x,\cdot), \pi\big) \le \big\{a_1\big(V(x) + b(\ell-1)\big) + a_2 + 1\big\}/H_\phi^{\leftarrow}(n/2) + 2M_\phi(a_1+1)/\phi\big(R(n/2)\big) + a_3/\upsilon_n.$$

(ii) The proof is along the same lines, using Lemma 18(ii) instead of Lemma 18(i). Finally, we end up with the following inequality for n large enough:

$$\begin{aligned} W_d\big(P^n(x,\cdot), \pi\big) \le{}& \big(1 + (1+b_1\kappa)\big\{\kappa^{-1} r_\phi(M_\kappa) + a_1\big(V(x) + b(\ell-1)\big) + b_2\big\}\big)/\big\{R^\delta(n)\big\} \\ &+ 2M_\phi\big((1+b_1\kappa)a_1 + 1\big)/\big\{\phi\big(R^\delta(n)\big)\big\}, \end{aligned}$$

where $\kappa = ((1-\varepsilon)^{-(1-\delta)/\delta} - 1)/b_1$. □

4.4. *Proof of Theorem* 4

Note that since $c = 1 - 2b/\phi(\upsilon)$ and $\upsilon > \phi^{\leftarrow}(2b)$, we get $c \in (0,1)$. Set $\mathbf{C} = \{V \le \upsilon\}$. By (7),

$$PV(x) + PV(y) \le V(x) + V(y) - c\phi\big(V(x) + V(y)\big) + 2b\mathbb{1}_{\mathbf{C}\times\mathbf{C}}(x,y) + \Omega(x,y),$$

where $\Omega(x,y) = c\phi(V(x) + V(y)) - \phi(V(x)) - \phi(V(y)) + 2b\mathbb{1}_{(\mathbf{C}\times\mathbf{C})^c}(x,y)$. We show that for every $x, y \in E$, $\Omega(x,y) \le 0$. Since ϕ is sub-additive (note that $\phi(0) = 0$), for all $x, y \in E$

$$\Omega(x,y) \le -(1-c)\big(\phi\big(V(x)\big) + \phi\big(V(y)\big)\big) + 2b\mathbb{1}_{(\mathbf{C}\times\mathbf{C})^c}(x,y).$$

On $(\mathbf{C}\times\mathbf{C})^c$, $\phi(V(x)) + \phi(V(y)) \ge \phi(\upsilon)$. The definition of c implies that $\Omega(x,y) \le 0$.

5. Proofs of Section 3.3

Lemma 21. *Let $M > 0$. Assume that there exists an increasing continuously differentiable concave function $\phi : [M,\infty) \to \mathbb{R}_+$, such that $\lim_{x\to\infty}\phi'(x) = 0$ and satisfying, on $\{V \ge M\}$, $PV(x) \le V(x) - \phi\circ V(x) + b$. Then, there exist $\tilde\phi \in \mathbb{F}$ and $\tilde b$ such that, $PV \le V - \tilde\phi\circ V + \tilde b$ on E, $\phi(v) = \tilde\phi(v)$ for all v large enough, and $\tilde\phi(0) = 0$.*

Proof. Observe indeed that the function $\tilde\phi$ defined by

$$\tilde\phi(t) = \begin{cases} (2\phi'(M) - \frac{\phi(M)}{M})t + \frac{2(\phi(M) - M\phi'(M))}{\sqrt{M}}\sqrt{t} & \text{for } 0 \le t < M, \\ \phi(t) & \text{for } t \ge M, \end{cases}$$

is concave increasing and continuously differentiable on $[1, +\infty)$, $\tilde\phi(0) = 0$, $\lim_{v\to\infty}\tilde\phi(v) = \infty$ and $\lim_{v\to\infty}\tilde\phi'(v) = 0$. The drift inequality (5) implies that for all $x \in E$

$$PV(x) \le V(x) - \tilde\phi\big(V(x)\big) + \tilde b,$$

with $\tilde b = b + \sup_{\{t\le M\}}\{\tilde\phi(t) - \phi(t)\}$. □

5.1. *Proof of Proposition* 11

For notational simplicity, let $P = P_{\mathrm{cn}}$. By definition of P, $V(X_1) \le V(X_0) \vee V(\rho X_0 + \sqrt{1-\rho^2}Z_1)$. Since $\|x+y\|^2 \le 2\|x\|^2 + 2\|y\|^2$, we get

$$\sup_{x\in\mathrm{B}(0,1)} PV(x) \le \sup_{x\in\mathrm{B}(0,1)} \int_{\mathcal{H}} \exp\big(2s\big(\|x\|^2 + \big(1-\rho^2\big)\|z\|^2\big)\big)\,\mathrm{d}\gamma(z), \tag{43}$$

and Theorem 10 implies that the RHS is finite.

Now, let $x \notin \mathrm{B}(0,1)$ and set $w(x) = (1-\rho)\|x\|/2$. Define the events $\mathcal{I} = \{\|Z_1\| \le w(X_0)/\sqrt{1-\rho^2}\}$, $\mathcal{A} = \{\alpha(X_0, \rho X_0 + \sqrt{1-\rho^2} Z_1) \ge U\}$, and $\mathcal{R} = \{\alpha(X_0, \rho X_0 + \sqrt{1-\rho^2} Z_1) < U\}$, where $U \sim \mathcal{U}([0,1])$, $Z_1 \sim \gamma$, and U and Z_1 are independent. With these definitions, we get,

$$P V(x) = \mathbb{E}_x\big[V(X_1)\mathbb{1}_{\mathcal{I}^c}\big] + \mathbb{E}_x\big[V(X_1)\mathbb{1}_{\mathcal{I}}(\mathbb{1}_{\mathcal{A}} + \mathbb{1}_{\mathcal{R}})\big]. \tag{44}$$

For the first term in the RHS, using again $V(X_1) \le V(X_0) \vee V(\rho X_0 + \sqrt{1-\rho^2} Z_1)$ and $\|x+y\|^2 \le 2\|x\|^2 + 2\|y\|^2$, we get

$$\begin{aligned}
\mathbb{E}_x\big[V(X_1)\mathbb{1}_{\mathcal{I}^c}\big] &\le \exp\big(2s\|x\|^2\big) \int_{\sqrt{1-\rho^2}\|z\| \ge w(x)} \exp\big(2s\big(1-\rho^2\big)\|z\|^2\big)\,\mathrm{d}\gamma(z) \\
&\le \exp\big(2s\|x\|^2 - (\theta/2)w(x)^2\big) \int_{\mathcal{H}} \exp\big((\theta/2 + 2s)\big(1-\rho^2\big)\|z\|^2\big)\,\mathrm{d}\gamma(z) \\
&\le \int_{\mathcal{H}} \exp\big((5/8)\big(1-\rho^2\big)\theta\|z\|^2\big)\,\mathrm{d}\gamma(z),
\end{aligned}$$

where the definition of s and w are used for the last inequality. Hence by Theorem 10, there exists a constant $b < \infty$ such that

$$\sup_{x \in \mathcal{H}} \mathbb{E}_x\big[V(X_1)\mathbb{1}_{\mathcal{I}^c}\big] \le b. \tag{45}$$

Consider the second term in the RHS of (44). On the event $\mathcal{A} \cap \mathcal{I}$, the move is accepted and $\|X_1 - \rho X_0\| \le w(X_0)$. On $\mathcal{R}$, the move is rejected and $X_1 = X_0$. Hence,

$$\mathbb{E}_x\big[V(X_1)\mathbb{1}_{\mathcal{I}}(\mathbb{1}_{\mathcal{A}} + \mathbb{1}_{\mathcal{R}})\big] \le \Big\{\sup_{z \in \mathrm{B}(\rho x, w(x))} V(z)\Big\}\mathbb{P}_x[\mathcal{I} \cap \mathcal{A}] + V(x)\mathbb{P}_x[\mathcal{I} \cap \mathcal{R}].$$

For $z \in \mathrm{B}(\rho x, w(x))$, by the triangle inequality, $V(z) \le \exp(s(1+\rho)^2\|x\|^2/4)$. Therefore for any $x \notin \mathrm{B}(0,1)$ since $\rho \in [0,1)$, $\sup_{z \in \mathrm{B}(\rho x, w(x))} V(z) \le \zeta V(x)$, with $\zeta = \exp\{((1+\rho)^2/4 - 1)s\} < 1$. This yields

$$\begin{aligned}
\mathbb{E}_x\big[V(X_1)\mathbb{1}_{\mathcal{I}}(\mathbb{1}_{\mathcal{A}} + \mathbb{1}_{\mathcal{R}})\big] &\le \zeta V(x)\mathbb{P}_x[\mathcal{I} \cap \mathcal{A}] + V(x)\mathbb{P}_x[\mathcal{I} \cap \mathcal{R}] \\
&\le V(x)\mathbb{P}_x[\mathcal{I}] - (1-\zeta)V(x)\mathbb{P}_x[\mathcal{A} \cap \mathcal{I}].
\end{aligned}$$

Since U_1 and Z_1 are independent, we get

$$\mathbb{P}_x[\mathcal{A} \cap \mathcal{I}] = \mathbb{E}_x\big[\big(1 \wedge \mathrm{e}^{g(x) - g(\rho x + \sqrt{1-\rho^2} Z_1)}\big)\mathbb{1}_{\mathcal{I}}\big].$$

By definition of the set $\mathcal{I}$ and using the inequality $\inf_{z \in \overline{\mathrm{B}}(\rho x, w(x))} \exp(g(x) - g(z)) \ge \exp(-C_g(1-\rho)^\beta(3/2)^\beta\|x\|^\beta)$, we get $\mathbb{P}_x[\mathcal{A} \cap \mathcal{I}] \ge \exp(-\{\ln V(x)/\kappa\}^{\beta/2})\mathbb{P}_x[\mathcal{I}]$, with $\kappa = \theta C_g^{-2/\beta}/36$. Hence, for any $x \notin \mathrm{B}(0,1)$,

$$\mathbb{E}_x\big[V(X_1)\mathbb{1}_{\mathcal{I}}(\mathbb{1}_{\mathcal{A}} + \mathbb{1}_{\mathcal{R}})\big] \le V(x) - (1-\zeta)V(x)\exp\big(-\kappa^{-\beta/2}\log^{\beta/2} V(x)\big). \tag{46}$$

Combining (43), (45) and (46) in (44), it follows that there exists $\tilde{b} > 0$ such that, for every $x \in \mathcal{H}$,

$$P V(x) \le V(x) - (1-\zeta)V(x)\exp\big(-\kappa^{-\beta/2}\log^{\beta/2} V(x)\big) + \tilde{b}.$$

The proof follows from Lemma 21.

5.2. *Proof of Proposition* 12

We preface the proof of Proposition 12 by a Lemma.

Lemma 22. *Assume* CN1. *There exists* $\eta \in (0,1)$ *satisfying the following assertions*

(i) *For all* $L > 0$, *there exists* $k(Q_{\mathrm{cn}}, L, \eta) < 1$ *such that, for all* $x, y \in \mathrm{B}(0, L)$ *satisfying* $d_\eta(x,y) < 1$, $Q_{\mathrm{cn}} d_\eta(x,y) \le k(Q_{\mathrm{cn}}, L, \eta) d_\eta(x,y)$.

(ii) *For all* $x, y \in \mathcal{H}$, $Q_{\mathrm{cn}} d_\eta(x,y) \le d_\eta(x,y)$.

Proof. Let $\eta \in (0,1)$; for ease of notation, we simply write Q for Q_{cn}. Let $L > 0$ and choose $x, y \in \mathrm{B}(0,L)$ satisfying $d_\eta(x,y) < 1$. Let (X_1, Y_1) be the basic coupling between $P(x,\cdot)$ and $P(y,\cdot)$; let Z_1, U_1 be the Gaussian variable and the uniform variable used for the basic coupling. Set $\mathcal{I} = \{\sqrt{1-\rho^2}\|Z_1\| \le 1\}$, $\mathcal{A} = \{\Psi_\wedge(X_0, Y_0, Z_1) > U_1\}$, $\mathcal{R} = \{\Psi_\vee(X_0, Y_0, Z_1) < U_1\}$, where

$$\Psi_\wedge(x,y,z) = \alpha\bigl(x, \rho x + \sqrt{1-\rho^2} z\bigr) \wedge \alpha\bigl(y, \rho y + \sqrt{1-\rho^2} z\bigr), \tag{47}$$

$$\Psi_\vee(x,y,z) = \alpha\bigl(x, \rho x + \sqrt{1-\rho^2} z\bigr) \vee \alpha\bigl(y, \rho y + \sqrt{1-\rho^2} z\bigr). \tag{48}$$

On the event $\mathcal{A}$, the moves are both accepted so that $X_1 = \rho X_0 + \sqrt{1-\rho^2} Z_1$ and $Y_1 = \rho X_0 + \sqrt{1-\rho^2} Z_1$; On the event $\mathcal{R}$, the moves are both rejected so that $X_1 = X_0$ and $Y_1 = Y_0$. It holds,

$$Q d_\eta(x,y) \le \widetilde{\mathbb{E}}_{x,y}\bigl[d_\eta(X_1,Y_1)\bigr] \le \widetilde{\mathbb{E}}_{x,y}\bigl[d_\eta(X_1,Y_1)\mathbb{1}_{\mathcal{A}\cup\mathcal{R}}\bigr] + \widetilde{\mathbb{P}}_{x,y}\bigl[(\mathcal{A}\cup\mathcal{R})^c\bigr], \tag{49}$$

where we have used d_η is bounded by 1. Since $d_\eta(X_1,Y_1) = \rho^\beta d_\eta(X_0,Y_0)$, on $\mathcal{A}$, and $d_\eta(X_1,Y_1) = d_\eta(X_0,Y_0)$, on $\mathcal{R}$, we get $\widetilde{\mathbb{E}}_{x,y}[d_\eta(X_1,Y_1)(\mathbb{1}_{\mathcal{A}\cup\mathcal{R}})] \le \rho^\beta d_\eta(x,y)\widetilde{\mathbb{P}}_{x,y}[\mathcal{A}] + d_\eta(x,y)\widetilde{\mathbb{P}}_{x,y}[\mathcal{R}]$. Since $\widetilde{\mathbb{P}}_{x,y}[\mathcal{A}] + \widetilde{\mathbb{P}}_{x,y}[\mathcal{R}] \le 1$, we have

$$\begin{aligned}\widetilde{\mathbb{E}}_{x,y}\bigl[d_\eta(X_1,Y_1)(\mathbb{1}_{\mathcal{A}\cup\mathcal{R}})\bigr] &\le d_\eta(x,y) - \bigl(1-\rho^\beta\bigr) d_\eta(x,y)\widetilde{\mathbb{P}}_{x,y}[\mathcal{A}] \\ &\le d_\eta(x,y) - \bigl(1-\rho^\beta\bigr) d_\eta(x,y)\widetilde{\mathbb{P}}_{x,y}[\mathcal{A}\cap\mathcal{I}].\end{aligned} \tag{50}$$

Set $\Theta(x,y,z) = |\alpha(x, \rho x + \sqrt{1-\rho^2} z) - \alpha(y, \rho y + \sqrt{1-\rho^2} z)|$. Since Z_1 and U_1 are independent, it follows that $\widetilde{\mathbb{P}}_{x,y}[(\mathcal{A}\cup\mathcal{R})^c] \le \int_{\mathcal{H}} \Theta(x,y,z)\,\mathrm{d}\gamma(z)$ Plugging this identity and (50) in (49) yields

$$Q d_\eta(x,y) \le d_\eta(x,y) - \bigl(1-\rho^\beta\bigr) d_\eta(x,y)\widetilde{\mathbb{P}}_{x,y}[\mathcal{A}\cap\mathcal{I}] + \int_{\mathcal{H}} \Theta(x,y,z)\,\mathrm{d}\gamma(z). \tag{51}$$

Let us now define $h : \mathcal{H} \to \mathbb{R}$ by

$$h(z) = g(z) - g(\rho z). \tag{52}$$

We bound from below $\widetilde{\mathbb{P}}_{x,y}[\mathcal{A}\cap\mathcal{I}]$. Since U_1 is independent of Z_1, it follows that

$$\widetilde{\mathbb{P}}_{x,y}[\mathcal{A}\cap\mathcal{I}] \ge \widetilde{\mathbb{E}}_{x,y}\bigl[\Psi_\wedge(X_0,Y_0,Z_1)\mathbb{1}_{\mathcal{I}}\bigr].$$

By CN1, for all $z \in \mathcal{H}$ such that $\sqrt{1-\rho^2}\|z\| \le 1$, it holds for $\varpi \in \mathcal{H}$, $g(\varpi) - g(\rho\varpi + \sqrt{1-\rho^2} z) \ge h(\varpi) - C_g$. Then,

$$\Psi_\wedge(x,y,z) \ge 1 \wedge \bigl(\mathrm{e}^{-C_g}\mathrm{e}^{h(x)}\bigr) \wedge \bigl(\mathrm{e}^{-C_g}\mathrm{e}^{h(y)}\bigr) \ge \mathrm{e}^{-C_g}\bigl[1 \wedge \mathrm{e}^{h(x)\wedge h(y)}\bigr].$$

Therefore,

$$\widetilde{\mathbb{P}}_{x,y}[\mathcal{A}\cap\mathcal{I}] \ge \mathrm{e}^{-C_g}\bigl[1 \wedge \mathrm{e}^{h(x)\wedge h(y)}\bigr]\widetilde{\mathbb{P}}_{x,y}[\mathcal{I}]. \tag{53}$$

We now upper bound the integral term in (51). For $x, y \in \mathcal{H}$, define the partition of $\mathcal{H}$,

$$\mathcal{K}_1(x,y) = \bigl\{z \in \mathcal{H} : \alpha\bigl(x, \rho x + \sqrt{1-\rho^2} z\bigr) = \alpha\bigl(y, \rho y + \sqrt{1-\rho^2} z\bigr) = 1\bigr\},$$

$$\mathcal{K}_2(x,y) = \bigl\{z \in \mathcal{H} : \alpha\bigl(x, \rho x + \sqrt{1-\rho^2} z\bigr) = 1 > \alpha\bigl(y, \rho y + \sqrt{1-\rho^2} z\bigr)\bigr\},$$

$$\mathcal{K}_3(x,y)=\{z\in\mathcal{H}:\alpha(y,\rho y+\sqrt{1-\rho^2}z)=1>\alpha(x,\rho x+\sqrt{1-\rho^2}z)\},$$

$$\mathcal{K}_4(x,y)=\{z\in\mathcal{H}:\alpha(y,\rho y+\sqrt{1-\rho^2}z)<1 \text{ and } \alpha(x,\rho x+\sqrt{1-\rho^2}z)<1\}.$$

Since on $\mathcal{K}_1(x,y)$, $\Theta(x,y,z)=0$,

$$\int_{\mathcal{H}}\Theta(x,y,z)\,\mathrm{d}\gamma(z)=\sum_{j=2}^{4}\int_{\mathcal{K}_j(x,y)}\Theta(x,y,z)\,\mathrm{d}\gamma(z). \tag{54}$$

For any $a,b>0$, we have $|a-b|=(a\vee b)[1-((a/b)\wedge(b/a))]$. Upon noting that $1-\mathrm{e}^{-t}\le t$ for any $t\ge 0$, we have

$$\Theta(x,y,z)\le\Psi_\vee(x,y,z)\big|g(y)-g(x)-g(\rho y+\sqrt{1-\rho^2}z)+g(\rho x+\sqrt{1-\rho^2}z)\big|\mathbb{1}_{\bigcup_{i=2}^4\mathcal{K}_i(x,y)}(z).$$

By CN1, this yields, for $x,y\in\mathcal{H}$ such that $d_\eta(x,y)<1$,

$$\Theta(x,y,z)\le 2C_g\|y-x\|^\beta\Psi_\vee(x,y,z)\le 2C_g\eta d_\eta(x,y)\Psi_\vee(x,y,z). \tag{55}$$

On $\mathcal{K}_2(x,y)$, $g(x)>g(\rho x+\sqrt{1-\rho^2}z)$ and, together with the definition (52), this implies that $h(x)\ge g(\rho x+\sqrt{1-\rho^2}z)-g(\rho x)$. Therefore, since under CN1, $h(x)\ge -C_g(1-\rho^2)^{\beta/2}\|z\|^\beta$ we get

$$\begin{aligned}\int_{\mathcal{K}_2(x,y)}\Theta(x,y,z)\,\mathrm{d}\gamma(z)&\le 2C_g\eta d_\eta(x,y)\int_{\mathcal{K}_2(x,y)}\mathrm{d}\gamma(z)\\&\le 2C_g\eta d_\eta(x,y)\left\{\left[\mathrm{e}^{h(x)}\int_{\mathcal{K}_2(x,y)}\mathrm{e}^{C_g(1-\rho^2)^{\beta/2}\|z\|^\beta}\,\mathrm{d}\gamma(z)\right]\wedge 1\right\}\\&\le C_I\eta d_\eta(x,y)\{\mathrm{e}^{h(x)}\wedge 1\},\end{aligned} \tag{56}$$

for a constant C_I, which is finite according to Theorem 10. By symmetry, on $\mathcal{K}_3(x,y)$,

$$\int_{\mathcal{K}_3(x,y)}\Theta(x,y,z)\,\mathrm{d}\gamma(z)\le C_I\eta d_\eta(x,y)\{\mathrm{e}^{h(y)}\wedge 1\}. \tag{57}$$

On $\mathcal{K}_4(x,y)$, using CN1,

$$\alpha(x,\rho x+\sqrt{1-\rho^2}z)=\mathrm{e}^{g(x)-g(\rho x+\sqrt{1-\rho^2}z)}\wedge 1\le(\mathrm{e}^{h(x)}\mathrm{e}^{C_g(1-\rho^2)^{\beta/2}\|z\|^\beta})\wedge 1;$$

and by symmetry, we obtain a similar upper bound for $\alpha(y,\rho y+\sqrt{1-\rho^2}z)$. Since $\mathrm{e}^{C_g(1-\rho^2)^{\beta/2}\|z\|^\beta}\ge 1$, these two inequalities imply $\Psi_\vee(x,y,z)\le\mathrm{e}^{C_g(1-\rho^2)^{\beta/2}\|z\|^\beta}(\mathrm{e}^{h(x)\vee h(y)}\wedge 1)$. Hence, using again (55) and Theorem 10, there exists $C_I<+\infty$ such that

$$\int_{\mathcal{K}_4(x,y)}\Theta(x,y,z)\,\mathrm{d}\gamma(z)\le C_I\eta d_\eta(x,y)[\mathrm{e}^{h(x)\vee h(y)}\wedge 1]. \tag{58}$$

Plugging (56), (57), (58) into (54), we finally obtain

$$\int_{\mathcal{H}}\Theta(x,y,z)\,\mathrm{d}\gamma(z)\le 3C_I\eta d_\eta(x,y)[\mathrm{e}^{h(x)\vee h(y)}\wedge 1].$$

Finally, under CN1, for every $x,y\in\mathcal{H}$ such that $d_\eta(x,y)<1$, $|h(x)-h(y)|\le 2C_g\|x-y\|^\beta\le 2C_g\eta^\beta$. Therefore $\mathrm{e}^{h(x)\vee h(y)}\wedge 1\le\mathrm{e}^{2C_g\eta^\beta}[\mathrm{e}^{h(x)\wedge h(y)}\wedge 1]$ and

$$\int_{\mathcal{H}}\Theta(x,y,z)\,\mathrm{d}\gamma(z)\le 3C_I\mathrm{e}^{2C_g\eta^\beta}\eta d_\eta(x,y)[\mathrm{e}^{h(x)\wedge h(y)}\wedge 1]. \tag{59}$$

Plugging (53) and (59) in (51) yields

$$Qd_\eta(x,y) \le d_\eta(x,y)\big(1 - \big\{\big(1-\rho^\beta\big)\mathrm{e}^{-C_g}\widetilde{\mathbb{P}}_{x,y}[\mathcal{I}] - 3C_I\mathrm{e}^{2C_g\eta^\beta}\eta\big\}\big[\mathrm{e}^{h(x)\wedge h(y)} \wedge 1\big]\big).$$

Note that $M = \widetilde{\mathbb{P}}_{x,y}[\mathcal{I}]$ is a positive quantity that does not depend on x, y. Therefore, we may choose η sufficiently small so that, for every $x, y \in \mathcal{H}$ satisfying $d_\eta(x,y) < 1$,

$$Qd_\eta(x,y) \le d_\eta(x,y)\big(1 - (1/2)\big(1-\rho^\beta\big)\mathrm{e}^{-C_g}M\big[\mathrm{e}^{h(x)\wedge h(y)} \wedge 1\big]\big), \tag{60}$$

which implies Lemma 22(i) upon noting that, under the stated assumptions, $\inf_{\mathrm{B}(0,L)} h > -\infty$.

We now consider (ii). For every $x, y \in \mathcal{H}$, $d_\eta(x,y) \le 1$, which implies that $Qd_\eta(x,y) \le 1$. For every $x, y \in \mathcal{H}$ such that $d_\eta(x,y) = 1$, $Qd_\eta(x,y) \le 1 = d_\eta(x,y)$. If $d_\eta(x,y) < 1$, (60) shows that $Qd_\eta(x,y) \le d_\eta(x,y)$. □

Proof of Proposition 12. Let $\{(X_n, Y_n), n \in \mathbb{N}\}$ be a Markov chain with Markov kernel Q given by (13). We denote for all $n \in \mathbb{N}^*$, Z_n and U_n, respectively the common Gaussian variable and uniform variable, used in the definition (X_n, Y_n). Note that by definition the variables $\{Z_n, U_n; n \in \mathbb{N}\}$ are independent.

Since $\{x : V(x) \le u\} = \{x : \|x\| \le (s\log(u))^{1/2}\}$, for $u \ge 1$, we only prove that for all $L > 0$, there exist $\ell \in \mathbb{N}^*$ and $\varepsilon > 0$ such that $\overline{\mathrm{B}}(0,L)^2$ is a $(\ell, \varepsilon, d_\eta)$-coupling set. By Lemma 22(i), for any $L > 0$, there exists $k(Q,L,\eta) \in (0,1)$ such that for any $x, y \in \overline{\mathrm{B}}(0,L)$ satisfying $d_\eta(x,y) < 1$, $Qd_\eta(x,y) \le k(Q,L,\eta)d_\eta(x,y)$. Then by Lemma 22(ii), for every $n \in \mathbb{N}^*$,

$$Q^n d_\eta(x,y) \le Q^{n-1}d_\eta(x,y) \le \cdots \le k(Q,L,\eta)d_\eta(x,y). \tag{61}$$

Consider now the case $d_\eta(x,y) = 1$. Let $n \in \mathbb{N}^*$ and denote for all $1 \le i \le n$, $\mathcal{A}_i = \{U_i \le \Psi_\wedge(X_{i-1}, Y_{i-1}, Z_i)\}$ and $\tilde{\mathcal{A}}_i(n) = \bigcap_{1\le j\le i}(\{\sqrt{1-\rho^2}\|Z_j\| \le L/n\} \cap \mathcal{A}_j)$ where $\Psi_\wedge$ is defined in (47).

On the event $\tilde{\mathcal{A}}_i(n)$, $X_j = \rho X_{j-1} + \sqrt{1-\rho^2}Z_j$ and $Y_j = \rho Y_{j-1} + \sqrt{1-\rho^2}Z_j$ for all $1 \le j \le i$. Then, since $d_\eta(x,y) \le \eta^{-1}\|x-y\|^\beta$, on $\tilde{\mathcal{A}}_n(n)$ it holds $d_\eta(X_n, Y_n) \le \eta^{-1}\rho^{\beta n}\|X_0 - Y_0\|^\beta$. This inequality and $d_\eta(x,y) \le 1$ yield

$$\begin{aligned} Q^n d_\eta(x,y) &= \widetilde{\mathbb{E}}_{x,y}\big[d_\eta(X_n,Y_n)(\mathbb{1}_{\tilde{\mathcal{A}}_n(n)} + \mathbb{1}_{(\tilde{\mathcal{A}}_n(n))^c})\big] \le \eta^{-1}\rho^{\beta n}\|x-y\|^\beta\widetilde{\mathbb{P}}_{x,y}\big[\tilde{\mathcal{A}}_n(n)\big] + \widetilde{\mathbb{P}}_{x,y}\big[\big(\tilde{\mathcal{A}}_n(n)\big)^c\big] \\ &\le \eta^{-1}\rho^{\beta n}(2L)^\beta\widetilde{\mathbb{P}}_{x,y}\big[\tilde{\mathcal{A}}_n(n)\big] + \widetilde{\mathbb{P}}_{x,y}\big[\big(\tilde{\mathcal{A}}_n(n)\big)^c\big] \le 1 + \big(\eta^{-1}\rho^{\beta n}(2L)^\beta - 1\big)\widetilde{\mathbb{P}}_{x,y}\big[\tilde{\mathcal{A}}_n(n)\big]. \end{aligned} \tag{62}$$

As $\rho \in [0,1)$, there exists ℓ such that, $\eta^{-1}\rho^{\beta\ell}(2L)^\beta < 1$. It remains to lower bound $\widetilde{\mathbb{P}}_{x,y}[\tilde{\mathcal{A}}_\ell(\ell)]$ by a positive constant to conclude. Since the random variables $\{(Z_i, U_i); i \in \mathbb{N}^*\}$ are independent, we get

$$\begin{aligned} \widetilde{\mathbb{P}}_{x,y}\big[\tilde{\mathcal{A}}_\ell(\ell)\big] &= \widetilde{\mathbb{P}}_{x,y}\big[\tilde{\mathcal{A}}_{\ell-1}(\ell) \cap \big\{\sqrt{1-\rho^2}\|Z_\ell\| \le L/\ell\big\}\big] \\ &\quad \times \widetilde{\mathbb{E}}_{x,y}\big[\Psi_\wedge(X_{\ell-1}, Y_{\ell-1}, Z_\ell)|\tilde{\mathcal{A}}_{\ell-1}(\ell) \cap \big\{\sqrt{1-\rho^2}\|Z_\ell\| \le L/\ell\big\}\big]. \end{aligned}$$

For all $1 \le i \le \ell$, on the event $\bigcap_{j\le i}\{\sqrt{1-\rho^2}\|Z_j\| \le L/\ell\}$, it holds

$$\Psi_\wedge(X_{i-1}, Y_{i-1}, Z_i) \ge \exp\Big(-\sup_{z\in\mathrm{B}(0,2L)} g(z) + \inf_{z\in\mathrm{B}(0,2L)} g(z)\Big) = \delta,$$

where $\delta \in (0,1)$. Therefore, since Z_ℓ is independent of $\tilde{\mathcal{A}}_{\ell-1}(\ell)$, we have

$$\widetilde{\mathbb{P}}_{x,y}\big[\tilde{\mathcal{A}}_\ell(\ell)\big] \ge \delta\widetilde{\mathbb{P}}_{x,y}\big[\tilde{\mathcal{A}}_{\ell-1}(\ell)\big]\widetilde{\mathbb{P}}_{x,y}\big[\sqrt{1-\rho^2}\|Z_\ell\| \le L/\ell\big].$$

An immediate induction leads to $\widetilde{\mathbb{P}}_{x,y}[\tilde{\mathcal{A}}_\ell(\ell)] \ge (\widetilde{\mathbb{P}}_{x,y}[\sqrt{1-\rho^2}\|Z_1\| \le L/\ell])^\ell\delta^\ell$. Plugging this result in (62) and (61) implies there exists $\zeta \in (0,1)$ such that for all $x, y \in \overline{\mathrm{B}}(0,L)$, $Q^\ell d_\eta(x,y) \le \zeta d_\eta(x,y)$. □

Appendix A: Wasserstein distance: some useful properties

Let (E,d) be a Polish space, with d bounded by 1. Then, for all $\mu,\nu\in\mathcal{P}(E)$: $W_d(\mu,\nu)\le W_{d_0}(\mu,\nu)$ since for all $x,y\in E$, $d(x,y)\le d_0(x,y)$. Hence when d is bounded by 1, the convergence in total variation distance implies the convergence in the Wasserstein metric W_d.

Lemma 23. *Let (E,d) be a Polish space, with d bounded by* 1, *and let P be a Markov kernel on $(E,\mathcal{B}(E))$. Let Q be a coupling kernel for P.*

(i) *Then, for all probability measures $\mu,\nu\in\mathcal{P}(E)$ and $n\in\mathbb{N}^*$,*

$$W_d\bigl(\mu P^n,\nu P^n\bigr)\le\inf_{\lambda\in\mathcal{C}(\mu,\nu)}\int_{E\times E}Q^n d(z,t)\,\mathrm{d}\lambda(z,t).$$

(ii) *If in addition Q is a d-weak-contraction, then for all $x,y\in E$, $W_d(P(x,\cdot),P(y,\cdot))\le d(x,y)$ and for all probability measures $\mu,\nu\in\mathcal{P}(E)$,*

$$W_d(\mu P,\nu P)\le W_d(\mu,\nu).$$

Proof. (i) For every $\lambda\in\mathcal{C}(\mu,\nu)$, λQ^n is a coupling of μP^n and νP^n. This yields the result. Consider now (ii). Using (i), we get

$$W_d(\mu P,\nu P)\le\inf_{\lambda\in\mathcal{C}(\mu,\nu)}\int_{E\times E}Qd(z,t)\,\mathrm{d}\lambda(z,t)\le\inf_{\lambda\in\mathcal{C}(\mu,\nu)}\int_{E\times E}d(z,t)\,\mathrm{d}\lambda(z,t)\le W_d(\mu,\nu).$$ □

Appendix B: Subgeometric functions and sequences

Lemma 24. *Let $r\in\Lambda_0$ and R be given by* (25).

(i) *For all $t,v\in\mathbb{R}_+$, $r(t+v)\le r(t)r(v)$.*
(ii) *R is differentiable, convex and increasing to $+\infty$.*
(iii) $\lim_{t\to\infty}r(t)/R(t)=0$.
(iv) *There exists a constant C such that for any $t,v\in\mathbb{R}_+$, $R(t+v)\le CR(t)R(v)$.*
(v) $\sup_k R(k)/\sum_{i=0}^{k-1}r(i)<\infty$.

Proof. (i) follows from [23, Lemma 1]. Consider now (ii). By definition, r is non-decreasing, thus is bounded on every compact set; then, R is continuous. Moreover, it is differentiable and its derivative is r, which is non-decreasing. Then R is convex. In addition $r(0)\ge 2$, thus R is increasing to $+\infty$. (iii) Set $u(t)\stackrel{\text{def}}{=}\log(r(t))/t$. Since $r\in\Lambda_0$, the function u is non increasing, which implies that, for every $h\in(0,1)$,

$$\log\bigl(1+\bigl\{r(t+h)-r(t)\bigr\}/r(t)\bigr)=\log\bigl(r(t+h)/r(t)\bigr)=t\bigl(u(t+h)-u(t)\bigr)+hu(t+h)\le hu(t+h).$$

Since $\lim_{t\to+\infty}u(t)=0$, for all $\varepsilon>0$, there exists $T\in\mathbb{R}_+$ such that for all $t\ge T$ and $h\in(0,1)$, $(r(t+h)-r(t))\le\varepsilon hr(t)$. Therefore for all $t\ge T$ and $h\in(0,1)$, $(R(t+h)-R(t))/(hR(t))\le\varepsilon+r(T+1)/R(t)$. Taking $h\to 0$ it follows $r(t)/R(t)\le\varepsilon+r(T+1)/R(t)$, for all $t\ge T$. The proof is concluded by (ii). (iv) follows from (i) and (iii). Finally, for (v), the upper bound follows from (iv) and $R(k-1)\le 1+\sum_{i=0}^{k-1}r(i)$. □

Acknowledgements

We thank the anonymous referees for detailed reports and very helpful comments.

References

[1] C. Andrieu, G. Fort and M. Vihola. Quantitative convergence rates for subgeometric Markov chains. *J. Appl. Probab.* **52** (2) (2015) 391–404. MR3372082

[2] P. H. Baxendale. Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Probab.* **15** (1B) (2005) 700–738. MR2114987

[3] A. Beskos, G. Roberts, A. Stuart and J. Voss. MCMC methods for diffusion bridges. *Stoch. Dyn.* **8** (3) (2008) 319–350. MR2444507

[4] P. Billingsley. *Convergence of Probability Measures*, 2nd edition. *Wiley Series in Probability and Statistics: Probability and Statistics*. Wiley, New York, 1999. MR1700749

[5] V. I. Bogachev. *Gaussian Measures*. *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1998. MR1642391

[6] O. Butkovsky. Subgeometric rates of convergence of Markov processes in the Wasserstein metric. *Ann. Appl. Probab.* **24** (2) (2014) 526–552. MR3178490

[7] O. A. Butkovsky and A. Y. Veretennikov. On asymptotics for Vaserstein coupling of Markov chains. *Stochastic Process. Appl.* **123** (9) (2013) 3518–3541. MR3071388

[8] B. Cloez and M. Hairer. Exponential ergodicity for Markov processes with random switching. *Bernoulli* **21** (1) (2015) 505–536. MR3322329

[9] R. Douc, G. Fort, E. Moulines and P. Soulier. Practical drift conditions for subgeometric rates of convergence. *Ann. Appl. Probab.* **14** (3) (2004) 1353–1377. MR2071426

[10] R. Douc, A. Guillin and E. Moulines. Bounds on regeneration times and limit theorems for subgeometric Markov chains. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** (2) (2008) 239–257. MR2446322

[11] R. Douc, E. Moulines and P. Soulier. Computable convergence rates for sub-geometric ergodic Markov chains. *Bernoulli* **13** (3) (2007) 831–848. MR2348753

[12] G. Fort and E. Moulines. Polynomial ergodicity of Markov transition kernels. *Stochastic Process. Appl.* **103** (1) (2003) 57–99. MR1947960

[13] M. Hairer, J. C. Mattingly and M. Scheutzow. Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations. *Probab. Theory Related Fields* **149** (1–2) (2011) 223–259. MR2773030

[14] M. Hairer, A. M. Stuart and S. J. Vollmer. Spectral gaps for Metropolis–Hastings algorithms in infinite dimensions. *Ann. Appl. Probab.* **24** (2014) 2455–2490. MR3262508

[15] S. F. Jarner and G. O. Roberts. Polynomial convergence rates of Markov chains. *Ann. Appl. Probab.* **12** (1) (2002) 224–247. MR1890063

[16] S. F. Jarner and R. L. Tweedie. Necessary conditions for geometric and polynomial ergodicity of random-walk-type. *Bernoulli* **9** (4) (2003) 559–578. MR1996270

[17] S. F. Jarner and R. L. Tweedie. Locally contracting iterated functions and stability of Markov chains. *J. Appl. Probab.* **38** (2) (2001) 494–507. MR1834756

[18] N. Madras and D. Sezer. Quantitative bounds for Markov chain convergence: Wasserstein and total variation distances. *Bernoulli* **16** (3) (2010) 882–908. MR2730652

[19] S. Meyn and R. Tweedie. *Markov Chains and Stochastic Stability*, 2nd edition. Cambridge University Press, New York, 2009. MR2509253

[20] G. O. Roberts and J. S. Rosenthal. Small and pseudo-small sets for Markov chains. *Stoch. Models* **17** (2) (2001) 121–145. MR1853437

[21] G. O. Roberts and J. S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probab. Surv.* **1** (2004) 20–71. MR2095565

[22] G. O. Roberts and R. L. Tweedie. Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83** (1) (1996) 95–110. MR1399158

[23] C. Stone and S. Wainger. One-sided error estimates in renewal theory. *J. Anal. Math.* **20** (1) (1967) 325–352. MR0217898

[24] P. Tuominen and R. L. Tweedie. Subgeometric rates of convergence of f-ergodic Markov chains. *Adv. in Appl. Probab.* **26** (3) (1994) 775–798. MR1285459

[25] A. Y. Veretennikov. On polynomial mixing bounds for stochastic differential equations. *Stochastic Process. Appl.* **70** (1) (1997) 115–127. MR1472961

[26] C. Villani. *Optimal Transport: Old and New*. *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, 2009. MR2459454