The Annals of Applied Probability
2014, Vol. 24, No. 2, 526–552
DOI: 10.1214/13-AAP922

SUBGEOMETRIC RATES OF CONVERGENCE OF MARKOV PROCESSES IN THE WASSERSTEIN METRIC

BY OLEG BUTKOVSKY[1]

Lomonosov Moscow State University and Technion—Israel Institute of Technology

We establish subgeometric bounds on convergence rate of general Markov processes in the Wasserstein metric. In the discrete time setting we prove that the Lyapunov drift condition and the existence of a "good" d-small set imply subgeometric convergence to the invariant measure. In the continuous time setting we obtain the same convergence rate provided that there exists a "good" d-small set and the Douc–Fort–Guillin supermartingale condition holds. As an application of our results, we prove that the Veretennikov–Khasminskii condition is sufficient for subexponential convergence of strong solutions of stochastic delay differential equations.

1. Introduction. In this paper, we study rate of convergence of Markov processes to an invariant measure in the Wasserstein metric. We establish subgeometric bounds on the convergence rate, thus generalizing the results of [4, 5, 11]. We apply the obtained estimates to prove subgeometric ergodicity of strong solutions of stochastic differential delay equations (SDDEs) under Veretennikov–Khasminskii-type conditions. This extends the corresponding results [4, 15, 25, 26] for stochastic differential equations (without delay).

There are quite a few works which deal with convergence of Harris recurrent Markov chains in total variation; see, for example, the monograph [16] and the references therein. Less is known about convergence of Markov chains that are not Harris recurrent. Recall [12] that if a Markov chain has a unique invariant measure, then either (a) the chain is positive Harris recurrent in an absorbing set and the invariant measure is nonsingular, or (b) the invariant measure is singular and there are no Harris sets. It is quite clear that in case (b) the marginal distributions of the Markov chain do not converge in total variation, whereas they might converge weakly (and, hence, in the Wasserstein metric). Thus, for non-Harris chains [case (b)] it is natural to study convergence in the Wasserstein metric (rather than in the total variation metric).

Received November 2012; revised January 2013.

[1] Supported in part by Russian Foundation for Basic Research Grant 10-01-00397-a, Israel Science Foundation Grant 497/10 and a Technion fellowship.

MSC2010 subject classifications. 60J05, 60J25, 34K50.

Key words and phrases. Markov processes, Wasserstein metric, stochastic delay equations, subgeometric convergence, Lyapunov functions.

Many interesting Markov processes fall into case (b). For instance, following [11], consider SDDE

$$dX(t) = -cX(t)\,dt + g\big(X(t-1)\big)\,dW(t), \qquad t > 0,$$

where $c > 0$, W is a one-dimensional Brownian motion and g is a strictly increasing positive bounded continuous function. One can show that the strong solution of this equation has a unique invariant measure and converges to it weakly, but not in total variation. On the other hand, the Wasserstein distance between $X(t)$ and the invariant measure decays exponentially to zero as $t \to \infty$. Section 3 contains further examples of processes belonging to case (b).

Many methods of estimation of convergence rates in the total variation metric assume that a Markov process is ψ-irreducible and are based on the analysis of small sets. Probably, one of the first results in this area is due to Dobrushin [3], who proved that if the whole state space is small, then a Markov chain is exponentially ergodic. Later Popov [20] and Nummelin and Tuominen [17] replaced the global Dobrushin condition with a combination of a local Dobrushin condition (existence of a "good" small set) and the Lyapunov drift condition (LDC). This result was further extended by Jarner and Roberts [13] and Douc and coauthors [5], who established polynomial and general subgeometric estimates of convergence rate, correspondingly. Similar results for continuous time Markov processes (under an additional assumption that the state space is locally compact) are due to Fort and Roberts [7] and Douc, Fort and Guillin [4]. The latter work provides subgeometric estimates of the convergence rate under condition that a certain functional of a Markov process is a supermartingale. Let us also mention the recent paper of Hairer and Mattingly [10], which contains a new simple proof of the exponential ergodicity of a Markov process under LDC and the local Dobrushin condition.

Thus, many techniques rely on the irreducibility of a Markov process, the existence of a "good" small set, and (for continuous time processes) the local compactness of the state space. However, if the state space is infinite-dimensional, then in most "typical" situations the process is non-Harris and, therefore these assumptions are not fulfilled. For instance, if we go back to the above SDDE, then it is easy to check that this processes is not ψ-irreducible, the state space is not locally compact and, as was pointed in [11], all small sets of this process are degenerate (i.e., consists of no more than one point).

An alternative to the local Dobrushin condition was suggested by Bakry, Cattiaux and Guillin in [1]. They obtained estimates of convergence rate in the total variation metric, provided that the LDC holds, and a Markov process has a unique invariant measure, which satisfies a local Poincaré inequality on a large enough set.

Let us discuss another alternative to this set of assumptions, which was developed by Hairer, Mattingly, and Scheutzow [11] specifically for establishing exponential convergence rates of SDDEs, stochastic PDEs, and other infinite-dimensional processes in the Wasserstein metric. Exploiting a new notion of a

d-small set (a generalization of the notion of a small set), in conjunction with the LDC, and without any additional assumptions on the irreducibility of the process, the authors proved the existence of a spectral gap in a suitable norm, and, hence, the exponential convergence to stationarity.

We extend this result and consider the more general situation where a spectral gap may not exist. For discrete time Markov processes (Theorem 2.1) we prove that existence of a "good" d-small set and the LDC implies subgeometrical convergence in the Wasserstein metric. In the continuous time setting (Theorem 2.4) we obtain the same rate of convergence provided that there exists a "good" d-small set and the Douc–Fort–Guillin supermartingale condition holds. Thus, we also extend the results of [4, 5].

We apply our conditions to study the asymptotic behavior of strong solutions of SDDEs. We prove that Veretennikov–Khasminskii-type conditions are sufficient for subexponential ergodicity (Theorem 3.3). This extends the results of [4, 15, 25, 26].

The rest of the paper is organized as follows. Section 2 contains definitions and the main results. Applications to SDDEs and to an autoregressive model are presented in Section 3. The proofs of the main results are placed in Section 4.

2. Main results. Let $X = (X_n)_{n\in\mathbb{Z}_+}$ be a homogeneous Markov chain on a measurable space $(E, \mathcal{B}(E))$ with transition functions $P^n(x, A) := P_x(X_n \in A)$, where $x \in E$, $A \in \mathcal{B}(E)$, $n \in \mathbb{Z}_+$. As usual for $n = 1$ we will drop the upper index and write $P(x, A)$. For a measurable function $f : E \to [0, \infty)$, let $\mathcal{P}_f(E)$ be the set of probability measures on $(E, \mathcal{B}(E))$ which integrate f. We will write $\mathcal{P}(E)$ for the set of all probability measures on $(E, \mathcal{B}(E))$. If $\mu \in \mathcal{P}_f(E)$, denote $\mu(f) := \int_E f(x)\mu(dx)$. We define Markov semigroup operators as usual,

$$P\varphi(x) := \int_E \varphi(t)P(x, dt), \qquad P\mu(dx) := \int_E P(t, dx)\mu(dt).$$

Recall (see, e.g., [2]) that if d is a semimetric on E, then the *Wasserstein semidistance* W_d between probability measures $\mu, \nu \in \mathcal{P}(E)$ is given by

$$W_d(\mu, \nu) := \inf_{\lambda\in\mathcal{C}(\mu,\nu)} \int_{E\times E} d(x, y)\lambda(dx, dy),$$

where $\mathcal{C}(\mu, \nu)$ is the set of all probability measures on $(E \times E, \mathcal{B}(E \times E))$ with marginals μ and ν. If d is a proper metric, then W_d is a distance.

We consider also the *total variation metric* on the space $\mathcal{P}(E)$, which is defined by the following formula:

$$d_{\mathrm{TV}}(\mu, \nu) := 2 \sup_{A\in\mathcal{B}(E)} \bigl|\mu(A) - \nu(A)\bigr|, \qquad \mu, \nu \in \mathcal{P}(E).$$

Recall that if the space E is equipped with the discrete metric $d_0(x, y) := \mathrm{I}(x \neq y)$, $x, y \in E$, then the Wasserstein distance is just half of the total variation distance, that is, $W_{d_0}(\mu, \nu) = d_{\mathrm{TV}}(\mu, \nu)/2$, $\mu, \nu \in \mathcal{P}(E)$.

DEFINITION 2.1. A set $A \in \mathcal{B}(E)$ is called *small* for a Markov operator P if there exists $\varepsilon > 0$ such that for all $x, y \in A$,

$$\tfrac{1}{2} d_{\mathrm{TV}}\big(P(x,\cdot), P(y,\cdot)\big) \le 1 - \varepsilon.$$

For instance, any one-point set is small. However, as discussed above, a Markov process might have no small sets that consist of more than one point. To study such Markov processes Hairer, Mattingly and Scheutzow [11] introduce the following concept.

DEFINITION 2.2. A set $A \in \mathcal{B}(E)$ is called *d-small* for a Markov operator P if there exists $\varepsilon > 0$ such that for all $x, y \in A$,

$$W_d\big(P(x,\cdot), P(y,\cdot)\big) \le (1-\varepsilon)\, d(x,y).$$

Note that our definition of a d-small set is a bit different from the definition of [11]. Namely, the multiplier $d(x,y)$ appears on the right-hand side of the above inequality.

If $d(x,y) = \mathrm{I}(x \neq y)$, then the notions of a small set and a d-small set coincide. In the general case, the latter notion is much weaker than the former. In Section 3.1 we give an example of a Markov operator P that has a d-small state space and no nontrivial small sets.

Before we present our main result, let us recall that the total variation metric is contracting, that is, for any Markov semigroup $(P^t)_{t\ge 0}$ one has

$$d_{\mathrm{TV}}\big(P^t(x,\cdot), P^t(y,\cdot)\big) \le d_{\mathrm{TV}}\big(P^s(x,\cdot), P^s(y,\cdot)\big), \qquad x, y \in E$$

whenever $0 \le s \le t$. In general, the Wasserstein metric W_d may not be contracting. However, as discussed in detail in [11], it is natural to focus only on Wasserstein metrics that are contracting for the process X, since, in the general case, the Lyapunov drift condition is not sufficient even for a weak convergence toward the invariant measure. Note that the contractivity condition itself does not imply any convergence at all, either. It is the combination of the contractivity, the Lyapunov drift condition and the existence of a "good" d-small set, which yields the existence and uniqueness of the invariant measure and subgeometric convergence in the Wasserstein metric.

For a function $f : \mathbb{R}_+ \to (0;\infty)$ define

$$H_f(x) := \int_1^x \frac{1}{f(u)}\, du, \qquad x \ge 1.$$

Since H_f is increasing, the inverse function H_f^{-1} is well defined.

THEOREM 2.1. *Suppose there exist a measurable function $V : E \to [0;\infty)$ and a metric d on E such that the following conditions hold*:

(1) *V is a Lyapunov function*; *that is*, *there exist a concave differentiable function $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ increasing to infinity with $\varphi(0) = 0$ and a constant $K \geq 0$ such that*

$$PV \leq V - \varphi \circ V + K. \tag{2.1}$$

(2) *The space (E, d) is a complete separable metric space.*

(3) *The metric d is contracting and bounded by* 1; *that is*, *for any $x, y \in E$*,

$$W_d\big(P(x, \cdot), P(y, \cdot)\big) \leq d(x, y) \leq 1. \tag{2.2}$$

(4) *The level set $L := \{x, y \in E : V(x) + V(y) \leq R\}$ is d-small for some $R > \varphi^{-1}(2K)$*; *that is*, *there exists $\rho > 0$ such that*

$$W_d\big(P(x, \cdot), P(y, \cdot)\big) \leq (1 - \rho)\, d(x, y)$$

for any $x, y \in L$.

Then the process X has a unique stationary measure π and

$$\int_E \varphi\big(V(u)\big)\pi(du) \leq K.$$

Moreover, *for any $\varepsilon > 0$ there exist constants C_1 and C_2 such that for all $x \in E$*,

$$W_d\big(P^n(x, \cdot), \pi\big) \leq \frac{C_1(1 + V(x))}{\varphi(H_\varphi^{-1}(C_2 n))^{1-\varepsilon}}, \qquad n \in \mathbb{Z}_+. \tag{2.3}$$

REMARK 2.2. (i) If φ is a linear function, then the rate of convergence is exponential and this case is covered by [11], Theorem 4.8.

(ii) If $d(x, y) = \mathrm{I}(x \neq y)$, then the Wasserstein metric coincides with the total variation metric and this case is covered by [5], Proposition 2.5.

REMARK 2.3. Conditions (3) and (4) of the theorem are a bit more general than the corresponding conditions from [11], Theorem 4.8. Namely, we do not assume here that $W_d(P(x, \cdot), P(y, \cdot)) \leq (1 - \rho)\, d(x, y)$ for all $x, y \in E$ such that $d(x, y) < 1$. We suppose that this inequality is satisfied only for x, y belonging to the sublevel set.

Note that if φ grows to infinity not very rapidly (as x^γ for some $0 < \gamma < 1$ or slower), then the estimate of convergence rate given by (2.3) can be as close as possible to the estimate of convergence rate in the total variation distance obtained in [5], Proposition 2.5. Specific examples of convergence rates (polynomial, logarithmic, etc.) for different functions φ are given in [5], Section 2.3.

While the proof of the theorem is postponed to Section 4, we outline now the main steps.

SKETCH OF THE PROOF OF THEOREM 2.1. To prove the theorem we develop the idea of constructing an auxiliary contracting semimetric [9–11]. Namely, let

l be a semimetric on the space E such that $d(x,y) \le l(x,y)$ for all $x, y \in E$. It is possible to prove (for some "good" l) that for any probability measures $\mu, \nu \in \mathcal{P}_{\varphi\circ V}(E)$

$$W_l(P\mu, P\nu) \le \big(1 - \chi(\mu,\nu)\big) W_l(\mu,\nu),$$

where χ is a positive function (this is done in Lemma 4.3). Hence

$$W_d\big(P^n\mu, P^n\nu\big) \le W_l\big(P^n\mu, P^n\nu\big) \le \prod_{i=0}^{n-1}\big(1 - \chi\big(P^i\mu, P^i\nu\big)\big) W_l(\mu,\nu).$$

Of course, since we want to obtain subgeometric estimates of $W_d(P^n\mu, P^n\nu)$, there is no hope that $\inf_{\mu,\nu\in\mathcal{P}_{\varphi\circ V}(E)} \chi(\mu,\nu)$ is positive (this lower bound was greater than zero in [9–11], where geometric estimates were obtained). Yet, a good (albeit nonuniform) estimate of $\chi(P^{i+1}\mu, P^{i+1}\nu)$ can be derived. However, this estimate depends not only on $W_l(P^i\mu, P^i\nu)$ but also on $\mu(P^i(\varphi\circ V))$ and $\nu(P^i(\varphi\circ V))$. The latter two expressions are unbounded if μ, ν are fixed, and i runs over positive integers. Fortunately, there are sufficiently many integers i such that these two expressions are "small" (Lemma 4.1). This allows us to overcome this obstacle (Lemma 4.4) and obtain subgeometric bounds on $W_d(P^n\mu, P^n\nu)$. The last step is to prove the existence and uniqueness of the stationary measure (Lemma 4.5). $\square$

Now we give a similar result for continuous time Markov processes. Let $X = (X_t)_{t\ge 0}$ be a time-homogeneous strong Markov process, and let $(P_t)_{t\ge0}$ be the associated Markov semigroup. Recall [6], Theorem 2, that if a Markov process has càdlàg paths, then the strong Markov property is implied by the Feller property.

THEOREM 2.4. *Suppose there exist a measurable function $V : E \to [0;\infty)$ and a metric d on E such that the following conditions hold*:

(1) *V is a Lyapunov function*; *that is*, *there exist a concave differentiable function $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ increasing to infinity with $\varphi(0) = 0$ and a constant $K \ge 0$ such that for all $t \ge 0$, $x \in E$*

$$\mathrm{E}_x V(X_t) \le V(x) - \mathrm{E}_x \int_0^t \varphi\big(V(X_u)\big)\,du + Kt. \tag{2.4}$$

(2) *The space (E, d) is a complete separable metric space.*

(3) *The metric d is bounded by* 1 *and contracting for all $t \ge t_0$, for some $t_0 \ge 0$; that is, for any $x, y \in E$*

$$W_d\big(P^t(x,\cdot), P^t(y,\cdot)\big) \le d(x,y) \le 1.$$

(4) *The level set $L := \{x, y \in E : V(x) + V(y) \le R\}$ is d-small for all $R > 0$ and all $t \ge t_0$, that is, there exists $\rho = \rho(R,t) > 0$ such that*

$$W_d\big(P^t(x,\cdot), P^t(y,\cdot)\big) \le (1-\rho)\, d(x,y)$$

for any $x, y \in L$.

Then the process X has a unique stationary measure π and $\pi(\varphi \circ V) \leq K$. Moreover, for any $\varepsilon > 0$ there exist constants C_1 and C_2 such that for all $x \in E$,

$$W_d\bigl(P^t(x, \cdot), \pi\bigr) \leq \frac{C_1(1 + V(x))}{\varphi(H_\varphi^{-1}(C_2 t))^{1-\varepsilon}}, \qquad t \geq 0. \tag{2.5}$$

REMARK 2.5. (i) The linear case $\varphi(x) = \lambda x$, $\lambda > 0$ is [11], Theorem 4.8.

(ii) The case where the metric d is discrete, that is, $d(x, y) = \mathrm{I}(x \neq y)$, is [4], Theorem 3.2.

REMARK 2.6. (i) Condition (1) of Theorem 2.4 is equivalent to the Douc–Fort–Guillin supermartingale condition [4], equation (3.2); that is, inequality (2.4) holds if and only if the process $Z := (Z_t)_{t \geq 0}$,

$$Z_t := V(X_t) + \int_0^t \varphi\bigl(V(X_u)\bigr)\, du - Kt, \qquad t \geq 0$$

is a supermartingale with respect to the natural filtration of the process X.

(ii) Let L be the extended generator (see, e.g., [22], Definition 7.1.8) of the Markov process X. If the function V belongs to the domain of L and

$$LV \leq -\varphi \circ V + K,$$

where $K > 0$ and $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ is a concave differentiable function increasing to infinity with $\varphi(0) = 0$, then condition (1) of Theorem 2.4 holds.

The proof of this theorem is given in Section 4. Let us describe here the main idea.

SKETCH OF THE PROOF OF THEOREM 2.4. Combining the technique from [4, 7, 18], we find a function $W : E \to [0; \infty)$ such that

$$P^{t_0} W(x) \leq W(x) - \varphi\bigl(K_1 W(x)\bigr) + K_2, \qquad x \in E$$

for some positive K_1, K_2. Therefore, by Theorem 2.1, the skeleton chain $(X_{nt_0})_{n \in \mathbb{Z}_+}$ has a unique invariant measure. It is possible to prove that this measure is also invariant for the Markov process X, and inequality (2.5) holds. □

Thus Theorems 2.1 and 2.4 suggest a new method for proving results concerning subgeometrical convergence. Namely, one needs to find a suitable contracting metric d and a suitable Lyapunov function V with d-small sublevel sets, such that the conditions of the theorems hold. It extends the ability of the existing methods by allowing to choose the metric d (which might be different from the discrete metric).

3. Examples and applications. Let us give some applications of the results of the previous section. The focus here is on stochastic delay equations; however, it is possible to apply the results of this kind to study convergence in the Wasserstein metric for other classes of Markov processes; see, for example, [11], Section 5.3, for estimates of convergence rates of stochastic partial differential equations.

We first recall some terminology from [16]. A Markov chain $X = (X_n)_{n\in\mathbb{Z}_+}$ is said to be ψ*-irreducible* if there exists a nontrivial measure ψ on $\mathcal{B}(E)$ such that for any $x \in E$ and any set $A \in \mathcal{B}(E)$ with $\psi(A) > 0$, one has $\mathrm{P}_x(T_A < \infty) > 0$, where T_A is the first return time to the set A, that is, $T_A := \inf\{n \geq 1 : X_n \in A\}$.

A set $H \in \mathcal{B}(E)$ is called *absorbing* if $P(x, H) = 1$ for all $x \in H$, and *Harris* if there exists a measure ψ on $\mathcal{B}(E)$ with $\psi(H) > 0$ such that for any $x \in H$ and any set $A \in \mathcal{B}(E)$ with $\psi(A) > 0$ one has $\mathrm{P}_x(T_A < \infty) = 1$.

An invariant measure π is called *singular* if for any $x \in E$ there exists an absorbing set S_x such that $x \in S_x$ and $\pi(S_x) = 0$. In other words, the Markov chain, whatever the starting point is, will remain in the set of π-measure 0.

3.1. *Autoregressive model.* Consider the following peculiar AR(1) process, which belongs to case (b).

EXAMPLE 3.1. Let $X = (X_n)_{n\in\mathbb{Z}_+}$ be an autoregressive process satisfying the following equation:

$$X_{n+1} = \tfrac{1}{10} X_n + \varepsilon_{n+1}, \qquad n \in \mathbb{Z}_+,$$

where $\varepsilon_1, \varepsilon_2, \dots$ are i.i.d. random variables uniformly distributed on the set $\{0, \frac{1}{10}, \dots, \frac{9}{10}\}$ and $X_0 \in [0; 1)$. In other words, to get X_{n+1} from X_n one needs to take the decimal notation of X_n (which starts with 0 followed by the decimal point) and insert a random digit immediately after the decimal point. Other digits in the decimal notation of X_n are shifted right by one position.

Clearly, X is a Markov process with state space $(E, \mathcal{E}) = ([0; 1), \mathcal{B}([0; 1)))$. Let d be the Euclidean metric on this space [i.e., $d(x, y) = |x - y|$, $x, y \in E$]. One can easily prove that the process X has a unique invariant measure π, which is uniformly distributed on the interval $[0; 1)$. Moreover, the sequence $\{X_n\}$ weakly converges to π as $n \to \infty$.

This autoregression has a number of very interesting and unusual features. First, it has a reconstruction property. Namely, if we have just one observation of X_n, where the integer n can be arbitrarily large, then it is possible to find an initial value X_0 with probability 1 by the following simple formula: $X_0 = \{10^n X_n\}$, where $\{b\}$ denotes the fractional part of a real b. In other words, one just needs to shift right the decimal point by n positions and drop all the digits which will be on the left of the decimal point.

Therefore for $x, y \in E$, $x \neq y$, the probability measures $P(x, \cdot)$ and $P(y, \cdot)$ are singular. Hence the process X has no nontrivial small sets. On the other hand, the whole state space E is d-small. Indeed, it is easily seen that $W_d(P(x,\cdot), P(y,\cdot)) \le |x-y|/10$, for any $x, y \in E$.

Observe also that the process X is not ψ-irreducible, and, furthermore, it has uncountably many pairwise disjoint absorbing sets. Indeed, it is sufficient to note that for any $x \in E$ the set $S_x := \{y \in E \mid \exists m, n \in \mathbb{Z}_+ : \{10^m y\} = \{10^n x\}\}$ is absorbing, countable and for $x, y \in E$ either $S_x = S_y$ or $S_x \cap S_y = \varnothing$. By the same argument, the chain X has no Harris sets. Since $\pi(S_x) = 0$, we see that the measure π is singular.

Finally, let us point out that for any $x \in E$, the sequence $P^n(x,\cdot)$ does not converge to π in total variation [moreover, $d_{\mathrm{TV}}(P^n(x,\cdot),\pi) = 2$ for any positive integer n]. On the other hand, $P^n(x,\cdot)$ converges exponentially to π in the Wasserstein metric [moreover, $W_d(P^n(x,\cdot),\pi) \le 10^{-n}$ for any positive integer n].

3.2. *Stochastic delay equations.* In this subsection we present our results on convergence of SDDEs in the Wasserstein metric.

Fix $r > 0$, positive integers n, m, and let $\mathcal{C} = \mathcal{C}([-r; 0], \mathbb{R}^n)$ be the space of continuous functions from $[-r; 0]$ to $\mathbb{R}^n$ equipped with the supremum norm $\|\cdot\|$. Following [11], introduce the following family of metrics on the space $\mathcal{C}$:

$$d_\beta(x, y) = 1 \wedge \|x - y\|/\beta, \qquad \beta > 0.$$

Consider the stochastic differential delay equation

$$\begin{cases} dX(t) = f(X_t)\,dt + g(X_t)\,dW(t), & t \ge 0, \\ X_0 = x, \end{cases} \tag{3.1}$$

where $f : \mathcal{C} \to \mathbb{R}^n$, $g : \mathcal{C} \to \mathbb{R}^{n\times m}$, W is an m-dimensional Brownian motion, $x \in \mathcal{C}$ is the initial condition and as usual we use the notation $X_t(s) := X(t+s)$, $-r \le s \le 0$. It is clear that the process $X = (X_t)_{t\ge 0}$ defined on the state space $(\mathcal{C}, \mathcal{B}(\mathcal{C}))$ is Markov.

Throughout this section we assume that the drift and the diffusion satisfy the following conditions:

- the drift satisfies a one-sided Lipschitz condition, and the diffusion is Lipschitz; that is, there exists $K > 0$ such that for any $x, y \in \mathcal{C}$

$$2\langle f(x) - f(y), x(0) - y(0)\rangle^+ + |||g(x) - g(y)|||^2 \le K\|x - y\|^2; \tag{3.2}$$

- the diffusion is nondegenerate; that is, for any $x \in \mathcal{C}$ the matrix $g(x)$ admits a right inverse $g^{-1}(x)$ and

$$\sup_{x\in\mathcal{C}} |||g^{-1}(x)||| < \infty; \tag{3.3}$$

- (3.4) $\qquad f$ is continuous and bounded on bounded subsets of $\mathcal{C}$.

Here $\langle \cdot, \cdot \rangle$ is the standard scalar product in $\mathbb{R}^n$; for a real b we write $b^+ := \max(b, 0)$, and $|||M|||$ denotes the Frobenius norm of a matrix M, that is, $|||M|||^2 = \sum M_{ij}^2$. As in [26] we also define

$$\lambda_+ = \sup_{\substack{x \in \mathcal{C} \\ x(0) \neq 0}} \left\langle g(x) g^T(x) \frac{x(0)}{|x(0)|}, \frac{x(0)}{|x(0)|} \right\rangle, \qquad \Lambda = \sup_{x \in \mathcal{C}} \frac{\operatorname{Tr} g(x) g^T(x)}{n}.$$

Conditions (3.2) and (3.4) imply [27] the existence and uniqueness of the strong solution of SDDE (3.1).

Now we give a general theorem, which describes convergence rates in the Wasserstein metric W_{d_β}. Theorem 3.2(i) is a generalization of [11], Assumption 5.1.

THEOREM 3.2. *Suppose conditions* (3.2)–(3.4) *hold, and there exists a Lyapunov function* $V : \mathcal{C} \to \mathbb{R}_+$ *that satisfies inequality* (2.4). *If either*

(i) $\lim_{\|x\| \to \infty} V(x) = \infty$

or

(ii) $V(x) = U(x(0))$, *for some function* $U : \mathbb{R}^n \to \mathbb{R}_+$, $\lim_{|v| \to \infty} U(v) = \infty$, *the diffusion coefficient is uniformly bounded, and the drift coefficient can be decomposed into two terms,*

$$f(x) = f_1(x) + f_2\big(x(0)\big), \qquad x \in \mathcal{C}, \tag{3.5}$$

where the function f_1 *is bounded*;

then SDDE (3.1) *has a unique invariant measure* π. *Furthermore, for any* $\beta > 0$, *the rate of convergence of* $\mathrm{Law}(X_t)$ *to* π *in the Wasserstein metric* W_{d_β} *is given by* (2.5).

PROOF. Fix $\beta > 0$. Let us check that the process X and the function V satisfy the conditions of Theorem 2.4. It follows from [11], Proposition 5.4, and [24], Lemma 3.7.2, that the process X is Feller. Since X has continuous paths, we see that X is strongly Markovian. The first condition of the theorem is satisfied by assumption. The second condition also holds. In case (i) it follows directly from [11], Section 5.2, that there exists a $\gamma \in (0; \beta)$ such that the third and the fourth conditions are met. In case (ii), arguing as in [11], Proposition 5.3 and Lemma 3.8, one can show that the set $\{x \in \mathcal{C} : |x(0)| \le R\}$, $R \ge 0$ is d_γ-small for some $\gamma \in (0; \beta)$, and the metric d_γ is contracting. Thus, in both cases the conditions of Theorem 2.4 are satisfied.

Apply Theorem 2.4 to the process X. It follows from this theorem that SDDE (3.1) has a unique invariant measure π, and the rate of convergence of $\mathrm{Law}(X_t)$ to π in the metric W_{d_γ} is provided in (2.5). To complete the proof, it remains to note that for any measures $\mu_1, \mu_2 \in \mathcal{P}(E)$ one has $W_{d_\beta}(\mu_1, \mu_2) \le W_{d_\gamma}(\mu_1, \mu_2)$. $\square$

Ergodic properties of stochastic differential equations (SDE) were studied by Veretennikov [25, 26], Malyshkin [15], Klokov [14], Douc, Fort and Guillin [4] and many others. It is known that the Veretennikov–Khasminskii condition on the drift combined with a certain nondegeneracy condition on the diffusion is sufficient for the existence and uniqueness of the invariant measure for the strong solution of an SDE. Moreover, these conditions yield exponential, subexponential or polynomial (depending on the value of the constant α, see below) convergence toward the invariant measure in the total variation metric [4, 19]. The following theorem extends these results to SDDE.

THEOREM 3.3. *Suppose conditions* (3.2)–(3.4) *hold*, $\Lambda < \infty$ *and the function* f_1 *in decomposition* (3.5) *is bounded.*

(i) *Assume additionally that for some constants* $\alpha \in (0, 1]$, $M > 0$, $\varkappa > 0$, *the generalized Veretennikov–Khasminskii condition holds, that is,*

$$\langle f(x), x(0)\rangle \le -\varkappa |x(0)|^{\alpha}, \qquad x \in \mathcal{C}, |x(0)| \ge M. \tag{3.6}$$

Then SDDE (3.1) *has a unique invariant measure* π, *and* $\operatorname{Law}(X_t)$ *converges to* π *in the Wasserstein metric* W_{d_β} *subexponentially* (*if* $0 < \alpha < 1$) *or exponentially* (*if* $\alpha = 1$); *that is, for any* $\beta > 0$ *there exists positive constants* C_1 *and* C_2 *such that*

$$W_{d_\beta}\bigl(P^t(x,\cdot),\pi\bigr) \le C_1 \exp\{C_1\|x\|^{\alpha} - C_2 t^{\alpha/(2-\alpha)}\}, \qquad x \in \mathcal{C}, t > 0. \tag{3.7}$$

(ii) *If* (3.6) *holds with* $\alpha = 0$ *and* $\varkappa > n\Lambda/2$, *then SDDE* (3.1) *has a unique invariant measure* π, *but* $\operatorname{Law}(X_t)$ *converges to* π *in the Wasserstein metric* W_{d_β} *only polynomially*; *that is, for any* $\beta > 0$, $\varepsilon > 0$ *there exist* $C > 0$ *such that*

$$W_{d_\beta}\bigl(P^t(x,\cdot),\pi\bigr) \le C\bigl(1 + \|x\|^{2+2\varkappa_0}\bigr)t^{-\varkappa_0+\varepsilon}, \qquad x \in \mathcal{C}, t > 0,$$

where $\varkappa_0 = (\varkappa - n\Lambda/2)\lambda_+^{-1}$.

PROOF. The proof is based on the application of Theorem 3.2(ii) with a suitable Lyapunov function V. (i) Following [14], Section 3 (see also [4], Proposition 5.2), let $U : \mathbb{R}^n \to [0; \infty)$ be a twice continuously differentiable function such that $U(v) = \exp\{k|v|^{\alpha}\}$ for $|v| \ge M_0$. The parameters $M_0 \ge M$ and $k \ge 0$ will be chosen later. Take $V(x) = U(x(0))$. By Ito's Lemma, for any $x \in \mathcal{C}$ and $t > 0$ one has

$$\begin{aligned}
\mathrm{E}_x V(X_t) &\le V(x) + \alpha k \mathrm{E}_x \int_0^t \mathrm{I}\bigl(|X(s)| \ge M_0\bigr)V(X_s)|X(s)|^{\alpha-2}\langle X(s), f(X_s)\rangle\, ds \\
&\qquad + \frac{1}{2}\alpha k \mathrm{E}_x \\
&\qquad \times \int_0^t \mathrm{I}\bigl(|X(s)| \ge M_0\bigr)V(X_s)|X(s)|^{\alpha-2}\bigl(\lambda_+\alpha k|X(s)|^{\alpha} + C_1\bigr)\, ds \\
&\qquad + C_2 t \\
&\le V(x) - C_3\alpha k \mathrm{E}_x \int_0^t \mathrm{I}\bigl(|X(s)| \ge M_0\bigr)V(X_s)|X(s)|^{2\alpha-2}\, ds + C_2 t,
\end{aligned}$$

where $C_1 = \lambda_+(\alpha - 2) + n\Lambda$, $C_2 > 0$, $C_3 = \varkappa - \frac{1}{2}\lambda_+\alpha k - \frac{1}{2}C_1 M_0^{-\alpha}$ and in the second inequality we made use of (3.6).

Let $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ be a concave differentiable function with $\varphi(0) = 0$ and $\varphi(t) = t(\ln t)^{(2\alpha-2)/\alpha}$ for $t \ge e^2$. Take $k = \frac{\varkappa}{2\lambda_+\alpha}$, and $M_0 = (\frac{C_1}{\varkappa})^{1/\alpha} \vee (\frac{2}{k})^{1/\alpha} \vee M$. Then $U(M_0) \ge e^2$ and

$$
\begin{aligned}
\mathrm{E}_x V(X_t) &\le V(x) - C_4 \mathrm{E}_x \int_0^t \mathrm{I}(|X(s)| \ge M_0) V(X_s) |X(s)|^{2\alpha-2}\, ds + C_2 t \\
&= V(x) - C_5 \mathrm{E}_x \int_0^t \mathrm{I}(|X(s)| \ge M_0) \varphi(V(X_s))\, ds + C_2 t \\
&\le V(x) - C_5 \mathrm{E}_x \int_0^t \varphi(V(X_s))\, ds + C_6 t,
\end{aligned}
$$

where $C_4 := \alpha k \varkappa/4$, $C_5 := C_4 k^{2/\alpha-2}$ and $C_6 > 0$. Thus the function V satisfies inequality (2.4). Theorem 3.2(ii) now yields the existence and the uniqueness of the invariant measure π and implies estimate (3.7).

(ii) Now let $U(v) = |v|^k$, where $k > 2$. We take $V(x) = U(x(0))$ and proceed as follows:

$$
\begin{aligned}
\mathrm{E}_x V(X_t) &\le V(x) + \frac{1}{2} k \mathrm{E}_x \int_0^t |X(s)|^{k-2} \bigl(2\langle X(s), f(X_s)\rangle + (k-2)\lambda_+ + n\Lambda\bigr)\, ds \\
&\le V(x) - k C_1 \mathrm{E}_x \int_0^t \mathrm{I}(|X(s)| \ge M) |X(s)|^{k-2}\, ds + C_2 t,
\end{aligned}
$$

where $C_1 = \varkappa - \frac{k-2}{2}\lambda_+ - \frac{n\Lambda}{2}$, $C_2 > 0$. Set

$$
k = 2 + \frac{2\varkappa - n\Lambda}{\lambda_+} - \varepsilon,
$$

where $\varepsilon > 0$. By choosing $\varepsilon > 0$ small enough we can ensure that $k > 2$. Take $\varphi(u) = u^{(k-2)/k}$. Then

$$
\mathrm{E}_x V(X_t) \le V(x) - C_3 \mathrm{E}_x \int_0^t \varphi(V(X_s))\, ds + C_4 t
$$

for some $C_3, C_4 > 0$. Thus the function V satisfies condition (2.4), and the statement of the theorem follows now from Theorem 3.2(ii). $\square$

EXAMPLE 3.4. Consider the following peculiar SDDE:

$$
dX(t) = f(X(t))\, dt + g(X(t-1))\, dW(t),
$$

where $n = m = 1$, the functions f and g satisfies (3.2)–(3.4), f also satisfies (3.6), and g is a strictly increasing bounded positive continuous function. The strong solution of this SDDE also belongs to case (b). This SDDE has the reconstruction property [23]; that is, if we know X_t for any $t > 0$, then we can reconstruct the initial condition X_0 with probability one. Hence, the measures $P^t(x,\cdot)$ and $P^t(y,\cdot)$ are always singular for any $t>0$ and $x\neq y$. It follows from Theorem 3.3 that this SDDE has a unique invariant measure π. However, the reconstruction property implies that $d_{\mathrm{TV}}(P^t(x,\cdot),\pi)$ does not converge to 0 as $t\to\infty$, and the measure π is singular. On the other hand, if we replace the total variation metric d_{TV} by the Wasserstein metric W_{d_β} (these two metrics can be arbitrarily close to each other for sufficiently small β), then we see that $W_{d_\beta}(P^t(x,\cdot),\pi)$ converges to 0 subexponentially.

4. Proofs of the main results. To prove Theorems 2.1 and 2.4 we introduce some notation. Consider a semimetric $l(x,y):=d(x,y)^{1/p}(1+\beta\varphi(V(x)+V(y)))^{1/q}$, where $\beta>0$, $p,q>1$ and $1/p+1/q=1$. These parameters will be chosen later. We start with two auxiliary lemmas.

LEMMA 4.1. *Assume that a function $V:E\to[0;\infty)$ satisfies condition* (1) *of Theorem* 2.1. *Then for any $n\in\mathbb{Z}_+$*

$$\sum_{i=0}^{n-1} P^i(\varphi\circ V)\le nK+V. \tag{4.1}$$

Furthermore, if a measure π is invariant for the process X, then $\pi\in\mathcal{P}_{\varphi\circ V}(E)$ and $\pi(\varphi\circ V)\le K$.

PROOF. Let us rewrite (2.1) in the following form: $\varphi\circ V-K\le V-PV$. Applying the operator P^i, $i\in\mathbb{Z}_+$ to the both sides of this expression and summing the result over all $0\le i<n$, we get

$$\sum_{i=0}^{n-1} P^i(\varphi\circ V)-nK\le V-P^nV,$$

which proves (4.1).

To prove the second part of the lemma we combine the first part of the lemma with a cut-off argument; see, for example, [8], Proposition 4.24. Fix $L>0$. Then, for any nonnegative integer i, we have

$$\begin{aligned}\int_E((\varphi\circ V)(x)\wedge L)\pi(dx)&=\int_E P^i\big((\varphi\circ V)\wedge L\big)(x)\pi(dx)\\&\le\int_E\big(P^i(\varphi\circ V)(x)\wedge L\big)\pi(dx).\end{aligned}$$

Summing the both sides of the above inequality over all $0\le i<n$, we derive

$$\int_E((\varphi\circ V)(x)\wedge L)\pi(dx)\le\int_E\left(\left(\frac{1}{n}\sum_{i=0}^{n-1}P^i(\varphi\circ V)(x)\right)\wedge L\right)\pi(dx).$$

This, combined with (4.1), yields

$$\int_E ((\varphi \circ V)(x) \wedge L)\pi(dx) \le K + \int_E \left(\frac{V(x)}{n} \wedge L\right)\pi(dx).$$

Lebesgue's dominated convergence theorem implies that the integral on the right-hand side of the above inequality tends to 0 as $n \to \infty$. Thus

$$\int_E ((\varphi \circ V)(x) \wedge L)\pi(dx) \le K$$

and the second part of the lemma follows from Fatou's lemma. $\square$

The following Lemma 4.2 is due to Petrov.

LEMMA 4.2 ([21]). *Let $a_0, a_1, \dots$ be a sequence of positive numbers, and assume that for all $n \in \mathbb{Z}_+$ one has*

$$a_{n+1} \le a_n\bigl(1 - \psi(a_n)\bigr), \qquad 0 \le a_0 \le 1,$$

where $\psi : [0; \infty) \to [0; 1]$ is a continuous increasing function with $\psi(0) = 0$ and $\psi(x) > 0$ for $x > 0$. Then

$$a_n \le g^{-1}(n) \tag{4.2}$$

for all $n \in \mathbb{Z}_+$, where

$$g(x) := \int_x^1 \frac{dt}{t\psi(t)}, \qquad 0 < x \le 1.$$

PROOF. We see that the function g^{-1} is well defined. This follows from the fact that the function g is nonnegative, unbounded and strictly decreasing. Since ψ is positive, we have $a_{n+1} \le a_n$. By the mean value theorem, there exists $s \in [a_{n+1}; a_n]$ such that

$$g(a_{n+1}) - g(a_n) = g'(s)(a_{n+1} - a_n) = -\frac{a_{n+1} - a_n}{s\psi(s)} \ge \frac{a_n\psi(a_n)}{s\psi(s)} \ge 1.$$

Hence $g(a_n) \ge n$ and $a_n \le g^{-1}(n)$. $\square$

The next key lemma gives the estimate of the contraction rate in one step.

LEMMA 4.3. *Assume that the conditions of Theorem 2.1 hold. Then there exist $\beta = \beta(p, q)$ and positive $c_1(p, q), c_2(p, q), c_3(p, q)$ such that for any $\mu, \nu \in \mathcal{P}_{\varphi\circ V}(E)$ one has*

$$W_l(P\mu, P\nu) \le \bigl(1 - c_1 \wedge c_2\varphi'\bigl(\varphi^{-1}\bigl(cW_l(\mu, \nu)^{-p}\bigr)\bigr)\bigr)W_l(\mu, \nu),$$

where $c := c_3(\mu(\varphi \circ V) + \nu(\varphi \circ V))^p$ and the semimetric l was introduced at the beginning of this section.

Here, as usual, $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$ for real a, b. To simplify the formulas, we will drop a pair of parentheses and write $1 - a \wedge b$ for $1 - (a \wedge b)$.

PROOF OF LEMMA 4.3. We start as in the proof of [11], Theorem 4.8, by observing that since W_l is convex, the Jensen inequality implies

$$W_l(P\mu, P\nu) \le \int_{E\times E} W_l\bigl(P(x,\cdot), P(y,\cdot)\bigr)\alpha(dx, dy) \tag{4.3}$$

for any $\mu, \nu \in \mathcal{P}_{\varphi\circ V}(E)$ and any $\alpha \in \mathcal{C}(\mu, \nu)$. Applying the Cauchy–Schwarz inequality and the Jensen inequality for concave functions, we find that

$$\begin{aligned} &W_l\bigl(P(x,\cdot), P(y,\cdot)\bigr)\\ &\quad= \inf_\lambda \int_{E\times E} l(u,v)\lambda(du, dv)\\ &\quad\le \inf_\lambda \biggl(\int_{E\times E} d(u,v)\lambda(du,dv)\biggr)^{1/p}\\ &\qquad\times\biggl(1+\beta\int_{E\times E}\varphi\bigl(V(u)+V(v)\bigr)\lambda(du,dv)\biggr)^{1/q}\\ &\quad\le W_d\bigl(P(x,\cdot),P(y,\cdot)\bigr)^{1/p}\bigl(1+\beta\varphi\bigl(PV(x)+PV(y)\bigr)\bigr)^{1/q}, \end{aligned}\tag{4.4}$$

where the infimum is taken over all measures $\lambda \in \mathcal{C}(P(x,\cdot), P(y,\cdot))$.

To estimate the right-hand side of the last inequality we consider three different cases. Note once again that contrary to the proof of [11], Theorem 4.8, it is impossible here to obtain a nontrivial upper uniform bound for $W_l(P(x,\cdot), P(y,\cdot))/l(x,y)$.

Fix a large $M > R$.

Case 1. $V(x) + V(y) \le R$. In this case we proceed similar to [9, 11]. Using (4.4) and conditions (1) and (4) of the theorem, we obtain

$$W_l\bigl(P(x,\cdot), P(y,\cdot)\bigr) \le (1-\rho)^{1/p} d(x,y)^{1/p}\bigl(1+\beta\varphi(2K+R)\bigr)^{1/q}.$$

Setting

$$\beta = \frac{(1+\rho/(2-2\rho))^{q-1}-1}{\varphi(2K+R)}$$

we get

$$W_l\bigl(P(x,\cdot), P(y,\cdot)\bigr) \le (1-\rho/2)^{1/p} d(x,y)^{1/p} \le (1-\rho/2p)l(x,y).$$

Case 2. $R < V(x) + V(y) \le M$. In this case we make use of (2.1) and the concavity of φ to derive

$$\begin{aligned} &\varphi\bigl(PV(x)+PV(y)\bigr)\\ &\quad\le \varphi\bigl(V(x)+V(y)-\varphi\bigl(V(x)\bigr)-\varphi\bigl(V(y)\bigr)+2K\bigr)\\ &\quad\le \varphi\bigl(V(x)+V(y)-\varphi\bigl(V(x)+V(y)\bigr)+2K\bigr). \end{aligned}\tag{4.5}$$

Clearly, if $u \in (R; M]$, then again by the concavity of φ we have

$$\begin{aligned}\varphi\big(u - \varphi(u) + 2K\big) &\le \varphi(u)\left(1 - \big(\varphi(u) - 2K\big)\frac{\varphi'(u)}{\varphi(u)}\right)\\ &\le \varphi(u)\big(1 - \theta\varphi'(M)\big),\end{aligned}$$

where $\theta := 1 - 2K/\varphi(R)$. This inequality, combined with (4.4), (4.5) and contraction property (2.2), yields

$$\begin{aligned}&W_l\big(P(x,\cdot), P(y,\cdot)\big)\\ &\quad\le d(x,y)^{1/p}\big(1 + \beta\varphi\big(PV(x) + PV(y)\big)\big)^{1/q}\\ &\quad\le d(x,y)^{1/p}\big(1 + \beta\varphi\big(V(x) + V(y)\big)\big(1 - \theta\varphi'(M)\big)\big)^{1/q}\\ &\quad\le l(x,y)\left(1 - \frac{\theta\beta\varphi(R)}{1 + \beta\varphi(R)}\varphi'(M)\right)^{1/q}\\ &\quad\le l(x,y)\left(1 - \frac{\theta\beta\varphi(R)}{q(1 + \beta\varphi(R))}\varphi'(M)\right).\end{aligned}$$

Case 3. $V(x) + V(y) > M$. This is the easiest situation because in this case we would like to derive a very weak estimate of $W_l(P(x,\cdot), P(y,\cdot))$. Combining (2.2), (4.4) and (4.5), we get

$$\begin{aligned}&W_l\big(P(x,\cdot), P(y,\cdot)\big)\\ &\quad\le d(x,y)^{1/p}\big(1 + \beta\varphi\big(V(x) + V(y) - \varphi\big(V(x) + V(y)\big) + 2K\big)\big)^{1/q}\\ &\quad\le d(x,y)^{1/p}\big(1 + \beta\varphi\big(V(x) + V(y)\big)\big)^{1/q}\\ &\quad= l(x,y).\end{aligned}$$

Now we return to the main line of the proof. Introduce

$$c_1 = c_1(p,q,R,K) := \frac{\theta\beta\varphi(R)}{q(1 + \beta\varphi(R))}, \qquad c_2 = \rho/2p.$$

Note that the values of c_1 and c_2 depend neither on the choice of M nor on measures μ and ν. We see from (4.3) and the above estimates of $W_l(P(x,\cdot), P(y,\cdot))$ that for all $M > R$ one has

$$\begin{aligned}&W_l(P\mu, P\nu)\\ &\quad\le \big(1 - c_2 \wedge c_1\varphi'(M)\big)\int_{E\times E} l(x,y)\alpha(dx,dy)\\ &\qquad + \big(c_2 \wedge c_1\varphi'(M)\big)\int_{\{V(x)+V(y)>M\}} l(x,y)\alpha(dx,dy).\end{aligned} \tag{4.6}$$

The second integral on the right-hand side of (4.6) is estimated using Chebyshev inequality. Namely,

$$\begin{aligned}\int_{\{V(x)+V(y)>M\}} & l(x,y)\alpha(dx,dy)\\ &\le \int_{\{V(x)+V(y)>M\}} \bigl(1+\beta\varphi\bigl(V(x)+V(y)\bigr)\bigr)^{1/q}\alpha(dx,dy)\\ &\le C\int_{\{V(x)+V(y)>M\}} \varphi\bigl(V(x)+V(y)\bigr)^{1/q}\alpha(dx,dy)\\ &\le C\varphi(M)^{-1/p}\int_{E\times E}\varphi\bigl(V(x)+V(y)\bigr)\alpha(dx,dy)\\ &\le C\varphi(M)^{-1/p}\bigl(\mu(\varphi\circ V)+\nu(\varphi\circ V)\bigr),\end{aligned}$$

where $C = 1/K + \beta + 1$, and in the second inequality we used the bound $\varphi(M) > K$. Note that $\mu(\varphi\circ V)$ as well as $\nu(\varphi\circ V)$ are finite because it was assumed that $\mu,\nu\in\mathcal{P}_{\varphi\circ V}(E)$.

Recall that α is an arbitrary element of $\mathcal{C}(\mu,\nu)$. Hence we can take the infimum over all $\alpha\in\mathcal{C}(\mu,\nu)$ in (4.6) and use the above inequality to derive

$$\begin{aligned}W_l(P\mu,P\nu) &\le \bigl(1-c_2\wedge c_1\varphi'(M)\bigr)W_l(\mu,\nu)\\ &\quad + C\bigl(c_2\wedge c_1\varphi'(M)\bigr)\bigl(\mu(\varphi\circ V)+\nu(\varphi\circ V)\bigr)\varphi(M)^{-1/p}.\end{aligned}\tag{4.7}$$

Now we can choose M in such a way, that the right-hand side of the above expression is always smaller than $W_l(\mu,\nu)$. Namely, it is sufficient to require that

$$C\bigl(\mu(\varphi\circ V)+\nu(\varphi\circ V)\bigr)\varphi(M)^{-1/p}\le W_l(\mu,\nu)/2.$$

This inequality holds for

$$M = \varphi^{-1}\bigl(c_3\bigl(\mu(\varphi\circ V)+\nu(\varphi\circ V)\bigr)^p W_l(\mu,\nu)^{-p}\bigr),$$

where $c_3 = c_3(p,q,R,K) = 2^p(1/K+\beta+1)^p$. The substitution of the last expression into (4.7) proves the lemma. $\square$

LEMMA 4.4. *Assume that the conditions of Theorem 2.1 are satisfied. Let $\mu,\nu\in\mathcal{P}_{\varphi\circ V}(E)$ and let $(n_k)_{k\in\mathbb{Z}_+}$ be an increasing sequence of positive integers such that for all $k\in\mathbb{Z}_+$*

$$P^{n_k}\mu(\varphi\circ V)+P^{n_k}\nu(\varphi\circ V)\le C(\mu,\nu),$$

where $C(\mu,\nu)\ge 1$. Then there exist positive C_1, C_2 that do not depend on μ,ν such that for all $k\in\mathbb{Z}_+$,

$$W_l\bigl(P^{n_k}\mu,P^{n_k}\nu\bigr)\le C_1C(\mu,\nu)\frac{1}{\varphi(H_\varphi^{-1}(C_2k))^{1/p}}.\tag{4.8}$$

PROOF. We begin by observing that for any measures $\zeta_1, \zeta_2 \in \mathcal{P}_{\varphi\circ V}(E)$ one has

$$\begin{aligned}
W_l(\zeta_1, \zeta_2) &\le \int_{E\times E} \bigl(1+\beta\varphi\bigl(V(x)+V(y)\bigr)\bigr)^{1/q}\zeta_1(dx)\zeta_2(dy)\\
&\le \Bigl(1+\beta\int_{E\times E}\varphi\bigl(V(x)+V(y)\bigr)\zeta_1(dx)\zeta_2(dy)\Bigr)^{1/q}\\
&\le \bigl(1+\beta\zeta_1(\varphi\circ V)+\beta\zeta_2(\varphi\circ V)\bigr)^{1/q},
\end{aligned}$$

where we used the concavity of the function φ and the bound $d \le 1$. Hence,

$$\begin{aligned}
W_l\bigl(P^{n_0}\mu, P^{n_0}\nu\bigr) &\le \bigl(1+\beta P^{n_0}\mu(\varphi\circ V)+\beta P^{n_0}\nu(\varphi\circ V)\bigr)^{1/q}\\
&\le \bigl(1+\beta C(\mu,\nu)\bigr)^{1/q} \le (1+\beta)C(\mu,\nu).
\end{aligned}\tag{4.9}$$

Introduce $c_0 := 1+\beta$ and denote

$$a_n := \frac{W_l(P^n\mu, P^n\nu)}{c_0 C(\mu,\nu)}, \qquad n \in \mathbb{Z}_+.$$

It follows from Lemma 4.3 that $0 \le a_{n+1} \le a_n$ for all $n \in \mathbb{Z}_+$. Besides, by definition and (4.9) we have $a_{n_0} \le 1$. The function φ' is decreasing, therefore using Lemma 4.3, we derive

$$\begin{aligned}
a_{n_{k+1}} &\le a_{n_k+1}\\
&\le \bigl(1-c_1\\
&\qquad \wedge c_2\varphi'\bigl(\varphi^{-1}\bigl(c_3c_0^{-p}\bigl(P^{n_k}\mu(\varphi\circ V)+P^{n_k}\nu(\varphi\circ V)\bigr)^pC(\mu,\nu)^{-p}a_{n_k}^{-p}\bigr)\bigr)\bigr)a_{n_k}\\
&\le \bigl(1-c_1\wedge c_2\varphi'\bigl(\varphi^{-1}\bigl(c_4a_{n_k}^{-p}\bigr)\bigr)\bigr)a_{n_k},
\end{aligned}$$

where $c_4 = c_0^{-p}c_3$. Since $a_{n_0} \le 1$, it is possible to apply Lemma 4.2 to the sequence $(a_{n_k})_{k\in\mathbb{Z}_+}$. It follows from (4.2) that $a_{n_k} \le g^{-1}(k)$, where

$$\begin{aligned}
g(x) &= \int_x^1 \frac{dt}{c_1t\wedge c_2t\varphi'(\varphi^{-1}(c_4t^{-p}))} = c_5\int_x^{c_6}\frac{dt}{t\varphi'(\varphi^{-1}(c_4t^{-p}))}+c_7\\
&= c_8\int_{c_9}^{\varphi^{-1}(c_4x^{-p})}\frac{du}{\varphi(u)}+c_7 = c_8H_\varphi\bigl(\varphi^{-1}\bigl(c_4x^{-p}\bigr)\bigr)+c_{10}
\end{aligned}$$

and $c_5, c_6, \dots$ are some positive constants. Note that to obtain the third identity, we made the change of variables $u = \varphi^{-1}(c_4t^{-p})$. Thus we finally get $a_{n_k} \le c_{11}\varphi(H_\varphi^{-1}(c_{12}k))^{-1/p}$ and hence

$$W_l\bigl(P^{n_k}\mu, P^{n_k}\nu\bigr) \le c_{13}C(\mu,\nu)\varphi\bigl(H_\varphi^{-1}(c_{12}k)\bigr)^{-1/p}.$$

This completes the proof of the lemma. $\square$

LEMMA 4.5. *Under the conditions of Theorem 2.1, the process X has a unique stationary measure π.*

As was pointed out by the referee, if we additionally assumed that the sublevel sets of V are compact, and the process X is Feller, then the proof of the lemma would be trivial. Indeed, in this case the statement of the lemma would follow directly from the Krylov–Bogoliubov theorem; see [9], page 20. However, we do not make this assumption because we would like to apply Theorem 2.1 to Markov processes with a nonlocally compact state space and in particular, to strong solutions of stochastic delay equations defined on $\mathcal{C}([-r;0],\mathbb{R}^n)$; see Section 3.2.

PROOF OF LEMMA 4.5. First let us prove the existence of a stationary measure. Fix $x \in E$. Let us verify that the sequence of measures $(P^n\delta_x)_{n\in\mathbb{Z}_+}$ has a Cauchy subsequence. For $n < m \in \mathbb{Z}_+$, define

$$A(n,m) := \#\{i \in [n;m) : P^i(\varphi\circ V)(x) \le 4K + 4V(x) + 1\},$$

$$B(n,m) := \#\{i \in [0;n) : (P^i(\varphi\circ V)(x) \vee P^{m-n+i}(\varphi\circ V)(x)) \le 4K + 4V(x) + 1\}.$$

Here the symbol # denotes the cardinality of a finite set. It follows from the above definitions that for $n < m$,

$$B(n,m) \ge A(0,n) + A(m-n,n) - n. \tag{4.10}$$

Introduce the following sequence. Let $r_{-1} = -1$ and for $k \in \mathbb{Z}_+$,

$$r_k := \inf\{s > r_{k-1} : (P^s(\varphi\circ V)(x) \vee P^{m-n+s}(\varphi\circ V)(x)) \le 4K + 4V(x)\}.$$

We see that $r_{B(n,m)-1} < n$. We apply Lemma 4.4 to the sequence $(r_k)_{k\in\mathbb{Z}_+}$, the measures δ_x and $P^{m-n}\delta_x$ and take $C(\delta_x, P^{m-n}\delta_x) = 4K + 4V(x) + 1$. Then, by (4.8),

$$\begin{aligned} W_l(P^n\delta_x, P^m\delta_x) &\le W_l(P^{r_{B(n,m)-1}}\delta_x, P^{r_{B(n,m)-1}}(P^{m-n}\delta_x)) \\ &\le C_1(4K + 4V(x) + 1)\varphi(H_\varphi^{-1}(C_2B(n,m) - C_2))^{-1/p}, \end{aligned} \tag{4.11}$$

where we used Lemma 4.3 to obtain the first inequality. Recall that the constants C_1, C_2 are independent of n, m.

It follows from (4.1) that for any fixed n there exists an arbitrarily large m such that $A(mn,(m+1)n) \ge 3n/4$. Since $A(0,n) \ge 3n/4$, inequality (4.10) implies that for any fixed n there exists an arbitrarily large m such that $B(n,m) \ge n/2$. It is clear that for all such m, one has

$$\begin{aligned} &W_l(P^n\delta_x, P^m\delta_x) \\ &\quad\le C_1(4K + 4V(x) + 1)\varphi(H_\varphi^{-1}(C_2n/2 - C_2))^{-1/p} =: \Psi(n). \end{aligned}$$

It is evident that $\Psi(n) \to 0$, as $n \to \infty$.

Now we can construct the desired Cauchy subsequence. We set $n_0 = 0$, and for $k \in \mathbb{Z}_+$,

$$n_{k+1} := \inf\{m > n_k : B(n_k, m) \geq n_k/2 \text{ and } \Psi(m) \leq e^{-(k+1)}\}.$$

By the above arguments, we see that the sequence $(n_k)_{k\in\mathbb{Z}_+}$ is well defined, $B(n_k, n_{k+1}) \geq n_k/2$, and $\Psi(n_k) \leq e^{-k}$. Now we claim that the sequence $(P^{n_k}\delta_x)_{k\in\mathbb{Z}_+}$ is a Cauchy sequence in the space $(\mathcal{P}(E), W_d)$. Indeed, using (4.11) and the definition of n_k we derive

$$\begin{aligned}
W_d(P^{n_k}\delta_x, P^{n_{k+m}}\delta_x) &\leq \sum_{i=k}^{k+m-1} W_d(P^{n_i}\delta_x, P^{n_{i+1}}\delta_x) \\
&\leq \sum_{i=k}^{k+m-1} W_l(P^{n_i}\delta_x, P^{n_{i+1}}\delta_x) \\
&\leq \sum_{i=k}^{k+m-1} \Psi(n_i) \leq \sum_{i=k}^{k+m-1} e^{-i} \leq 2e^{-k}
\end{aligned}$$

for all integers k, m. Since the space $(\mathcal{P}(E), W_d)$ is complete (see, e.g., [2], Theorem 1.1.3), we see that there exists a measure $\pi \in \mathcal{P}(E)$ such that $W_d(P^{n_k}\delta_x, \pi) \to 0$.

Let us verify that the measure π is stationary, that is, let us check that $P\pi = \pi$. Note that the metric W_d is contractive. Indeed, for any $\mu, \nu \in \mathcal{P}(E)$, we have

$$\begin{aligned}
W_d(P\mu, P\nu) &\leq \inf_{\lambda\in\mathcal{C}(\mu,\nu)} \int_{E\times E} W_d(P(x,\cdot), P(y,\cdot))\lambda(dx, dy) \\
&\leq \inf_{\lambda\in\mathcal{C}(\mu,\nu)} \int_{E\times E} d(x, y)\lambda(dx, dy) \\
&= W_d(\mu, \nu),
\end{aligned}$$

where we used the Jensen inequality and condition (2.2).

Therefore, for any $k \in \mathbb{Z}_+$, we obtain

$$\begin{aligned}
W_d(P\pi, \pi) &\leq W_d(P\pi, P^{n_k+1}\delta_x) \\
&\quad + W_d(P^{n_k}\delta_x, P^{n_k+1}\delta_x) + W_d(P^{n_k}\delta_x, \pi) \\
&\leq 2W_d(\pi, P^{n_k}\delta_x) + W_l(P^{n_k}\delta_x, P^{n_k+1}\delta_x).
\end{aligned} \tag{4.12}$$

The first term on the right-hand side of the last expression tends to 0, as $k \to \infty$. To estimate the second term, we observe that if n is a positive integer, then $A(0, n) \geq$

$3n/4$ and $A(1, n+1) \geq 3n/4 - 1$. Therefore, inequality (4.10) implies $B(n, n+1) > n/2 - 1$. This, combined with (4.11), yields

$$W_l(P^{n_k}\delta_x, P^{n_k+1}\delta_x) \leq C_1(4K + 4V(x) + 1)\frac{1}{\varphi(H_\varphi^{-1}(C_2 n_k/2 - 2C_2))^{1/p}}.$$

Hence $W_l(P^{n_k}\delta_x, P^{n_k+1}\delta_x) \to 0$ as $k \to \infty$, and we conclude from (4.12) that $W_d(P\pi, \pi) = 0$, which implies the stationarity of the measure π.

To complete the proof of the lemma it remains to prove the uniqueness of stationary measure. Suppose that, on the contrary, the process X has two stationary measures π_1 and π_2 and $\pi_1 \neq \pi_2$. By Lemma 4.1, $\pi_1, \pi_2 \in \mathcal{P}_{\varphi \circ V}(E)$ and hence $0 < W_l(\pi_1, \pi_2) < \infty$. We make use of stationarity of the measures and Lemma 4.3 to obtain

$$W_l(\pi_1, \pi_2) = W_l(P\pi_1, P\pi_2) < W_l(\pi_1, \pi_2).$$

This contradiction proves the lemma. $\square$

PROOF OF THEOREM 2.1. It follows from Lemmas 4.1 and 4.5, that the process X has a unique stationary measure $\pi \in \mathcal{P}_{\varphi\circ V}(E)$ and $\pi(\varphi \circ V) \leq K$. Fix $x \in E$ and consider the following sequence. Let $n_0 = 0$ and

$$n_{k+1} := \inf\{m > n_k : P^m(\varphi \circ V) \leq 2K + 2V(x) + 1\}, \qquad k \in \mathbb{Z}_+.$$

We make use of stationarity of π, the bound $\pi(\varphi \circ V) \leq K$ and the definition of n_k to derive

$$P^{n_k}\delta_x(\varphi \circ V) + P^{n_k}\pi(\varphi \circ V) = P^{n_k}\delta_x(\varphi \circ V) + \pi(\varphi \circ V) \leq 3K + 2V(x) + 1.$$

Let us apply Lemma 4.4 to the measures δ_x, π, to the sequence $(n_k)_{k\in\mathbb{Z}_+}$ and take $C(\delta_x, \pi) = 3K + 2V(x) + 1$. Clearly, $C(\delta_x, \pi) > 1$. It follows from (4.8) that

$$W_l(P^{n_k}\delta_x, \pi) \leq C_1(3K + 2V(x) + 1)\frac{1}{\varphi(H_\varphi^{-1}(C_2 k))^{1/p}}.$$

On the other hand, it follows from (4.1) that $n_k \leq 2k$. To complete the proof, it remains to take $1/p = 1 - \varepsilon$ and note that

$$\begin{aligned} W_d(P^{2k}\delta_x, \pi) &\leq W_l(P^{2k}\delta_x, \pi) = W_l(P^{2k}\delta_x, P^{2k-n_k}\pi) \leq W_l(P^{n_k}\delta_x, \pi) \\ &\leq C_1(3K + 2V(x) + 1)\frac{1}{\varphi(H_\varphi^{-1}(C_2 k))^{1/p}}. \end{aligned}$$

$\square$

To switch from discrete time to continuous time and prove Theorem 2.4, we combine different methods from [4, 7, 18]. First of all for a set $C \in \mathcal{B}(E)$, introduce the hitting time delayed by $\delta > 0$

$$\tau_C(\delta) := \inf\{t \geq \delta : X_t \in C\}$$

and the hitting and return times of the skeleton chain

$$\sigma_{m,C} := \inf\{n \in \mathbb{Z}_+ : X_{mn} \in C\};$$

$$T_{m,C} := \inf\{n \in \mathbb{Z}_+, n \geq 1 : X_{mn} \in C\},$$

where $m > 0$. Denote for brevity $C_R := \{x \in E : V(x) \leq R\}$.

LEMMA 4.6. *If $R > 0$ and $\varphi(R) > K$, then under the conditions of Theorem 2.4*

$$\mathrm{E}_x \tau_{\{V(x) \leq R\}}(\delta) \leq \frac{\delta\varphi(R) + V(x)}{\varphi(R) - K}$$

for all $x \in E$ and $\delta > 0$.

PROOF. Fix $L > \delta$. Observe that if $\delta \leq u < \tau_{C_R}(\delta)$, then by definition $V(X_u) \geq R$. Combining this with (2.4) we obtain

$$\begin{aligned}\mathrm{E}_x\big(\tau_{C_R}(\delta) \wedge L\big) &= \delta + \mathrm{E}_x \int_\delta^{\tau_{C_R}(\delta) \wedge L} du \\ &\leq \delta + \frac{1}{\varphi(R)} \mathrm{E}_x \int_\delta^{\tau_{C_R}(\delta) \wedge L} \varphi\big(V(X_u)\big)\, du \\ &\leq \delta + \frac{V(x) + K\mathrm{E}_x(\tau_{C_R}(\delta) \wedge L)}{\varphi(R)}.\end{aligned}$$

Therefore

$$\mathrm{E}_x(\tau_{C_R(\delta)} \wedge L) \leq \frac{\delta\varphi(R) + V(x)}{\varphi(R) - K}.$$

The desired inequality follows now from the Fatou lemma. $\square$

LEMMA 4.7. *Let $m > 0$. If $R > Km$ and $\varphi(R - Km) > K$, then under the conditions of Theorem 2.4,*

$$\mathrm{E}_x T_{m,C_R} \leq c_1 V(x) + c_2, \qquad x \in E,$$

where $c_1 = c_1(m, R, K)$ and $c_2 = c_2(m, R, K)$ are positive functions that do not depend on x.

PROOF. The proof of the lemma uses the ideas from the proof of [7], Proposition 22(ii). However, note that we cannot apply this proposition directly because in contrast to Fort and Roberts, we assumed neither that the set $\{V(x) \leq R\}$ is petite nor that the process X is Harris-recurrent with invariant measure.

Introduce $R' < R - Km$ such that $\varphi(R') > K$. The existence of such R' follows from the conditions of the lemma. Consider the following sequence of stopping times:

$$\tau^0 := 0, \qquad \tau^1 := \tau_{C_{R'}}(m), \qquad \tau^n := \inf\{t \geq \tau^{n-1} + m : X_t \in C_{R'}\}$$

and let $M := \sup_{x \in C_{R'}} \mathrm{E}_x \tau_{C_{R'}}(m)$. By Lemma 4.6,

$$M \le \frac{m\varphi(R') + R'}{\varphi(R') - K}.$$

For $n \in \mathbb{Z}_+$, $n \ge 1$ define $Z_n := \mathrm{I}\{X_{\lceil \tau^n/m \rceil m} \in C_R\}$, where $\lceil b \rceil$ denotes the upper integer part of a real b. By definition, $Z_n \in \mathcal{F}_{\tau^{n+1}}$, where we denote $\mathcal{F}_t := \sigma\{X_s, 0 \le s \le t\}$. We combine the strong Markov property, the Chebyshev inequality and (2.4) to obtain

$$\begin{aligned} \mathrm{P}(Z_n = 1 \mid \mathcal{F}_{\tau^n}) &= 1 - \mathrm{E}_{X_{\tau^n}} \mathrm{I}\{X_{\lceil \tau^n/m \rceil m - \tau^n} \notin C_R\} \\ &\ge 1 - \frac{V(X_{\tau^n}) + Km}{R} \\ &\ge \frac{R - R' - Km}{R} =: \gamma. \end{aligned} \tag{4.13}$$

It follows from the choice of R' that $\gamma > 0$.

Introduce $\eta := \inf\{n \in \mathbb{Z}_+, n \ge 1 : Z_n = 1\}$. Using the strong Markov property, (4.13) and following the same lines as in the proof of [18], Lemma 3.1, we get for $n \ge 1$ and $x \in C_{R'}$,

$$\begin{aligned} \mathrm{E}_x \tau^n \mathrm{I}(\eta \ge n) &\le \mathrm{E}_x \tau^{n-1} \mathrm{I}(\eta \ge n-1) \mathrm{E}\big(\mathrm{I}(Z_{n-1} = 0) \mid \mathcal{F}_{\tau^{n-1}}\big) \\ &\quad + \mathrm{E}_x \mathrm{I}(\eta \ge n-1) \mathrm{E}\big(\tau^n - \tau^{n-1} \mid \mathcal{F}_{\tau^{n-1}}\big) \\ &\le (1-\gamma) \mathrm{E}_x \tau^{n-1} \mathrm{I}(\eta \ge n-1) + (1-\gamma)^{n-1} M. \end{aligned}$$

Since $\mathrm{E}_x \tau^0 \mathrm{I}(\eta \ge 0)$ is obviously zero, by induction we establish the following estimate:

$$\mathrm{E}_x \tau^n \mathrm{I}(\eta \ge n) \le n M (1-\gamma)^{n-1}, \qquad x \in C_{R'}.$$

Thus we have

$$\mathrm{E}_x \tau^\eta \le \sum_{n=1}^{\infty} \mathrm{E}_x \tau^n \mathrm{I}(\eta \ge n) \le \frac{M}{\gamma^2}, \qquad x \in C_{R'}.$$

We combine this with Lemma 4.6 to finally obtain

$$\begin{aligned} m E_x T_{m,C_R} &\le \mathrm{E}_x \tau^1 + \mathrm{E}_x \mathrm{E}_{X_{\tau^1}} \tau^\eta + m \\ &\le \frac{m\varphi(R') + V(x)}{\varphi(R) - K} + \frac{m\varphi(R') + R'}{\gamma^2(\varphi(R') - K)} + m \\ &\le c_1 V(x) + c_2 \end{aligned}$$

for all $x \in E$. This completes the proof of the statement. $\square$

PROOF OF THEOREM 2.4. First let us prove that there exist a Lyapunov function $W : E \to [0, \infty)$ and positive constants K_1, K_2 such that

$$P^{t_0} W(x) \le W(x) - \varphi\bigl(K_1 W(x)\bigr) + K_2, \qquad x \in E. \tag{4.14}$$

Choose a sufficiently large R (such that the conditions of Lemma 4.7 hold with $m = t_0$), and let

$$W(x) := \mathrm{E}_x \sum_{k=0}^{\sigma_{t_0,C_R}} \varphi\bigl(V(X_{kt_0})\bigr).$$

It follows from [16], Theorem 11.3.5(i) that for $x \in E$

$$P^{t_0} W(x) = W(x) - \varphi\bigl(V(x)\bigr) + \mathrm{I}(x \in C_R)\mathrm{E}_x \sum_{k=1}^{T_{t_0,C_R}} \varphi\bigl(V(X_{kt_0})\bigr). \tag{4.15}$$

Using an argument similar to that in the proof of [4], Proposition 4.8(i), we obtain for any $L > 0$ and $x \in E$,

$$\begin{aligned}
&\mathrm{E}_x \sum_{k=1}^{T_{t_0,C_R}\wedge L} \varphi\bigl(1 + V(X_{kt_0})\bigr) - \mathrm{E}_x \int_0^{T_{t_0,C_R}\wedge L} \varphi\bigl(1 + V(X_{st_0})\bigr)\,ds \\
&\quad \le \frac{1}{2}\varphi'(1) K t_0 \mathrm{E}_x (T_{t_0,C_R} \wedge L).
\end{aligned}$$

Furthermore, using condition (2.4) and the concavity of the function φ, we get for any $x \in E$,

$$\begin{aligned}
&\mathrm{E}_x \int_0^{T_{t_0,C_R}\wedge L} \varphi\bigl(1 + V(X_{st_0})\bigr)\,ds \\
&\quad = \frac{1}{t_0}\mathrm{E}_x \int_0^{t_0 T_{t_0,C_R}\wedge t_0 L} \varphi\bigl(1 + V(X_u)\bigr)\,du \\
&\quad \le \varphi(1)\mathrm{E}_x (T_{t_0,C_R} \wedge L) + \frac{1}{t_0}\mathrm{E}_x \int_0^{t_0 T_{t_0,C_R}\wedge t_0 L} \varphi\bigl(V(X_u)\bigr)\,du \\
&\quad \le V(x)/t_0 + \bigl(\varphi(1) + K\bigr)\mathrm{E}_x (T_{t_0,C_R} \wedge L).
\end{aligned}$$

Combining this with the previous inequality and using Lemma 4.7 and Fatou's lemma, we derive for any $x \in E$,

$$\begin{aligned}
&\mathrm{E}_x \sum_{k=1}^{T_{t_0,C_R}} \varphi\bigl(V(X_{kt_0})\bigr) \\
&\quad \le V(x)/t_0 + \bigl(\varphi(1) + K + \varphi'(1) K t_0\bigr)\mathrm{E}_x T_{t_0,C_R} \\
&\quad \le V(x)/t_0 + c_3\bigl(c_1 V(x) + c_2\bigr) \\
&\quad \le c_4 V(x) + c_5,
\end{aligned} \tag{4.16}$$

where c_1 and c_2 are defined in Lemma 4.7, $c_3 := \varphi(1) + K + \varphi'(1)Kt_0$, $c_4 := 1/t_0 + c_1c_3$, $c_5 = c_2c_3$. Therefore, by the concavity of φ,

$$W(x) \le \varphi\bigl(V(x)\bigr) + c_4V(x) + c_5 \le V(x)\bigl(\varphi'(1) + c_4\bigr) + \varphi(1) + c_5.$$

This bound, together with (4.15) and (4.16), yields

$$P^{t_0}W(x) \le W(x) - \varphi\bigl(c_6W(x)\bigr) + c_4R + c_5 + c_7$$

for some positive c_6, c_7. Hence the function W satisfies (4.14).

Now the statement of Theorem 2.4 follows from the corresponding statement for discrete time chains. Indeed, the application of Theorem 2.1 to the skeleton chain $(X_{nt_0})_{n\in\mathbb{Z}_+}$ yields the existence of a measure π such that $P^{t_0}\pi = \pi$. Note that for any $0 < s < t_0$ the measure $\pi_s := P^s\pi$ is also invariant for this skeleton chain. Indeed, $P^{t_0}\pi_s = P^{t_0+s}\pi = P^sP^{t_0}\pi = \pi_s$. On the other hand, Theorem 2.1 yields uniqueness of the invariant measure. Thus, $P^s\pi = \pi$ and the measure π is invariant for the process X. Arguing as in the proof of Lemma 4.1, we see that $\pi(\varphi \circ V) \le K$.

It follows from Theorem 2.1 that for any $\varepsilon > 0$ there exist constants C_1, C_2 such that for all $x \in E$, $n \in \mathbb{Z}_+$,

$$W_d\bigl(P^{nt_0}(x,\cdot),\pi\bigr) \le \frac{C_1(1+V(x))}{\varphi(H_\varphi^{-1}(C_2n))^{1-\varepsilon}}.$$

We combine this with condition (4) of the theorem to conclude that for any $t > t_0$,

$$\begin{aligned} W_d\bigl(P^t(x,\cdot),\pi\bigr) &= W_d\bigl(P^t(x,\cdot), P^{t_0+t-\lfloor t/t_0\rfloor t_0}\pi\bigr) \\ &\le W_d\bigl(P^{(\lfloor t/t_0\rfloor-1)t_0}(x,\cdot),\pi\bigr) \\ &\le \frac{C_1(1+V(x))}{\varphi(H_\varphi^{-1}(C_3t))^{1-\varepsilon}} \end{aligned}$$

for some $C_3 > 0$. Here $\lfloor b\rfloor$ denotes the lower integer part of a real b. This completes the proof of Theorem 2.4. $\square$

Acknowledgments. The author is grateful to Professor A. V. Bulinski and Professor A. Yu. Veretennikov for their help and constant attention to this work. The author also would like to thank Professor M. Hairer and F. V. Petrov for useful discussions and the referee for his valuable comments and suggestions which helped to improve the quality of the paper.

REFERENCES

[1] BAKRY, D., CATTIAUX, P. and GUILLIN, A. (2008). Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. *J. Funct. Anal.* **254** 727–759. MR2381160

[2] BOGACHEV, V. I. and KOLESNIKOV, A. V. (2012). The Monge–Kantorovich problem: Achievements, connections, and perspectives. *Russian Math*. *Surveys* **67** 785–890. MR3058744

[3] DOBRUSHIN, R. (1956). Central limit theorem for nonstationary Markov chains. I. *Theory Probab*. *Appl*. **1** 65–80.

[4] DOUC, R., FORT, G. and GUILLIN, A. (2009). Subgeometric rates of convergence of f-ergodic strong Markov processes. *Stochastic Process*. *Appl*. **119** 897–923. MR2499863

[5] DOUC, R., FORT, G., MOULINES, E. and SOULIER, P. (2004). Practical drift conditions for subgeometric rates of convergence. *Ann*. *Appl*. *Probab*. **14** 1353–1377. MR2071426

[6] DYNKIN, E. B. and YUSHKEVICH, A. A. (1956). Strong Markov processes. *Theory Probab*. *Appl*. **1** 134–139.

[7] FORT, G. and ROBERTS, G. O. (2005). Subgeometric ergodicity of strong Markov processes. *Ann*. *Appl*. *Probab*. **15** 1565–1589. MR2134115

[8] HAIRER, M. (2006). Ergodic properties of Markov processes. Lecture notes, Univ. Warwick. Available at http://www.hairer.org/notes/Markov.pdf.

[9] HAIRER, M. (2010). Convergence of Markov processes. Lecture notes, Univ. Warwick. Available at http://www.hairer.org/notes/Convergence.pdf.

[10] HAIRER, M. and MATTINGLY, J. C. (2011). Yet another look at Harris' ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis*, *Random Fields and Applications VI*. *Progress in Probability* **63** 109–117. Birkhäuser, Basel. MR2857021

[11] HAIRER, M., MATTINGLY, J. C. and SCHEUTZOW, M. (2011). Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations. *Probab*. *Theory Related Fields* **149** 223–259. MR2773030

[12] HERNÁNDEZ-LERMA, O. and LASSERRE, J. B. (2000). On the classification of Markov chains via occupation measures. *Appl*. *Math*. (*Warsaw*) **27** 489–498. MR1816473

[13] JARNER, S. F. and ROBERTS, G. O. (2002). Polynomial convergence rates of Markov chains. *Ann*. *Appl*. *Probab*. **12** 224–247. MR1890063

[14] KLOKOV, S. A. and VERETENNIKOV, A. Y. (2004). On the sub-exponential mixing rate for a class of Markov diffusions. *J*. *Math*. *Sci*. (*N*. *Y*.) **123** 3816–3823. MR2093828

[15] MALYSHKIN, M. N. (2001). Subexponential estimates of the rate of convergence to the invariant measure for stochastic differential equations. *Theory Probab*. *Appl*. **45** 466–479.

[16] MEYN, S. and TWEEDIE, R. L. (2009). *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2509253

[17] NUMMELIN, E. and TUOMINEN, P. (1982). Geometric ergodicity of Harris recurrent Markov chains with applications to renewal theory. *Stochastic Process*. *Appl*. **12** 187–202. MR0651903

[18] NUMMELIN, E. and TUOMINEN, P. (1983). The rate of convergence in Orey's theorem for Harris recurrent Markov chains with applications to renewal theory. *Stochastic Process*. *Appl*. **15** 295–311. MR0711187

[19] PARDOUX, E. and VERETENNIKOV, A. Y. (2001). On the Poisson equation and diffusion approximation. I. *Ann*. *Probab*. **29** 1061–1085. MR1872736

[20] POPOV, N. (1977). Conditions for geometric ergodicity of countable Markov chains. *Soviet Math*. *Dokl*. **18** 676–679.

[21] PETROV, F. V. (2012). Personal communication.

[22] REVUZ, D. and YOR, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd ed. Springer, Berlin. MR1725357

[23] SCHEUTZOW, M. (2005). Exponential growth rates for stochastic delay differential equations. *Stoch*. *Dyn*. **5** 163–174. MR2147280

[24] SHIRYAEV, A. N. (1996). *Probability*, 2nd ed. *Graduate Texts in Mathematics* **95**. Springer, New York. MR1368405

[25] VERETENNIKOV, A. Y. (1997). On polynomial mixing bounds for stochastic differential equations. *Stochastic Process*. *Appl*. **70** 115–127. MR1472961
[26] VERETENNIKOV, A. Y. (2000). On polynomial mixing and convergence rate for stochastic difference and differential equations. *Theory Probab. Appl*. **44** 361–374.
[27] VON RENESSE, M.-K. and SCHEUTZOW, M. (2010). Existence and uniqueness of solutions of stochastic functional differential equations. *Random Oper. Stoch*. *Equ*. **18** 267–284. MR2718125

FACULTY OF MATHEMATICS AND MECHANICS
DEPARTMENT OF PROBABILITY THEORY
LOMONOSOV MOSCOW STATE UNIVERSITY
MOSCOW, 119991
RUSSIA
AND
FACULTY OF INDUSTRIAL ENGINEERING AND MANAGEMENT
TECHNION—ISRAEL INSTITUTE OF TECHNOLOGY
HAIFA 32000
ISRAEL
E-MAIL: oleg.butkovskiy@gmail.com