CONVERGENCE RATE TO EQUILIBRIUM IN WASSERSTEIN DISTANCE FOR REFLECTED JUMP-DIFFUSIONS

ANDREY SARANTSEV

Abstract. Convergence rate to the stationary distribution for continuous-time Markov processes can be studied using Lyapunov functions. Recent work by the author provided explicit rates of convergence in special case of a reflected jump-diffusion on a half-line. These results are proved for total variation distance and its generalizations: measure distances defined by test functions regardless of their continuity. Here we prove similar results for Wasserstein distance, convergence in which is related to convergence for continuous test functions. In some cases, including the reflected Ornstein-Uhlenbeck process, we get faster exponential convergence rates for Wasserstein distance than for total variation distance.

1. Introduction

Consider a Markov process \( X = (X(t), t \geq 0) \) on a metric state space \( S \) with transition function \( P^t(x, \cdot) \). Assume it has a unique stationary distribution, or invariant measure \( \pi \): If \( X(0) \sim \pi \) then \( X(t) \sim \pi \) for all \( t \geq 0 \). We are interested in convergence

\[
P^t(x, \cdot) \to \pi(\cdot), \quad t \to \infty.
\]

In which norm does it hold? How fast is this convergence? One well-known tool is a Lyapunov function. Define the \( g \)-norm for signed measures on \( S \) as follows:

\[
\|\nu\|_g := \sup_{|f| \leq g} |(\nu, f)|, \quad (\nu, f) := \int_S f(x) \nu(dx).
\]

For \( g = 1 \), this becomes a total variation norm \( \|\cdot\|_{TV} \). Assume \( \mathcal{L} \) is the generator of this process. If there is a function \( V : S \to [1, \infty) \) such that for a compact set \( K \subseteq S \),

\[
\mathcal{L}V(x) \leq -kV(x), \quad x \in S \setminus K,
\]

called a Lyapunov function, then (under additional technical assumptions) the convergence in (1) is exponential:

\[
\|P^t(x, \cdot) - \pi(\cdot)\|_{TV} \leq C(x)e^{-\kappa t}
\]

for some constant \( C(x) \) depending on \( x \in S \), and \( \kappa > 0 \). This statement can also be proved for the \( V \)-norm \( \|\cdot\|_V \) from [2], for the same Lyapunov function \( V \), see [1, 5, 15, 16]. However, to find or estimate an exact constant \( \kappa \) is challenging: [17, 20]. But we can conclude that \( \kappa = k \) for \( S = [0, \infty) \), \( K = \{0\} \), and the process \( X \) is stochastically ordered, [14, 22]. These results are applied to reflected diffusions and jump-diffusions. A reflected diffusion can be informally described informally as follows: Inside the half-line, it behaves between jumps as a solution to one-dimensional stochastic differential equation with drift \( g \) and diffusion \( \sigma^2 \):

\[
dX(t) = g(X(t)) \, dt + \sigma(X(t)) \, dW(t),
\]

where \( W \) is a Brownian motion. When it hits \( 0 \), it is reflected back to the half-line. A reflected jump-diffusion behaves as a reflected diffusion but, in addition, it can jump with
a certain rate. The jump destination is also random, and there are finitely many jumps in finite time. We applied our results to queueing theory [2] and risk theory [7].

For the proof, we use the coupling method: Take two copies $X_1$ and $X_2$ of this process starting from points $X_2(0) \geq X_1(0) \geq 0$. By stochastic ordering, we can couple them (that is, create them on the same probability space) so that $0 \leq X_1(t) \leq X_2(t)$. Take $\tau$ to be the hitting moment of zero by $X_2$. Then by stochastic ordering $X_1(\tau) = 0$, and we can assume $X_1(t) = X_2(t)$ for $t > \tau$. This enables us to estimate the distance between $P^t(x_1, \cdot)$ and $P^t(x_2, \cdot)$. This coupling method is commonly used for convergence proofs; see articles [8, 11, 13] and the book [12]. In our case, when $S = \mathbb{R}_+$ and $K = \{0\}$, this method is particularly powerful.

In this short note, we apply this method to Wasserstein distance instead of $\| \cdot \|_V$-norm. Wasserstein distance is defined via optimal couplings of measures; one can think of it as “earth mover” distance: If we have two piles of sand with equal volume, how much work does it require to move the first pile in the place of the second pile? This distance is related to optimal transport problems, see the fundamental monograph [23], in particular Chapter 6. As noted there, convergence in Wasserstein distance of order $p \geq 1$ is equivalent to uniform boundedness of $p$th moments and weak convergence of measures. The latter, in turn, is equivalent to convergence for continuous bounded test functions. Thus Wasserstein distance is fundamentally different from total variation norm or other norms as in [2], which use all test functions $f : S \to \mathbb{R}$, not just continuous ones. Previous work on convergence in this distance is scant, see [3]. In this article, we find convergence rates and compare them with that for total variation and other similar distances. For some processes, including a reflected Ornstein-Uhlenbeck process, convergence is faster in Wasserstein distance than in total variation or the $\| \cdot \|_V$-norm.

We note that the interplay between Wasserstein distance and Lyapunov functions is remarkable. Concentration of measure Talagrand inequalities for a distribution $\mathbb{P}$ compare Wasserstein distance from $\mathbb{P}$ to a test measure $\mathbb{Q}$ with relative entropy of $\mathbb{Q}$ with respect to $\mathbb{P}$. These inequalities were found for common stochastic processes on $[0, T]$; see [18, 4, 19, 21] and references therein. These concentration inequalities are related to other functional inequalities for finite time horizon, which can be obtained using Lyapunov functions, [1]. But these articles do not address the long-term behavior of stochastic processes, even if these processes are convergent in the long run (like an Ornstein-Uhlenbeck process). In this work, we fill this gap and combine Lyapunov functions with Wasserstein distance framework to study rates of long-term convergence.

This short note is organized as follows. In Section 2, we define all notation and concepts. In Section 3, we state our main results and give examples. Section 4 is devoted to proofs.

2. Notation, Definitions, and Background

Define the Skorohod space $\mathbb{D}([s, t])$ of right-continuous function with left limits $[s, t] \to \mathbb{R}$, with the distance $\rho_{\mathbb{D}}$ satisfying the following estimate:

\begin{align}
\rho_{\mathbb{D}}(x, y) &\leq \sup_{s \leq u \leq t} |x(u) - y(u)|, \quad x, y \in D[s, t].
\end{align}

For two probability measures $\mathbb{P}$ and $\mathbb{Q}$ on the same metric space $(S, \rho)$, and for a $p \geq 1$, Wasserstein distance of order $p$ is defined as

\begin{align}
\mathcal{W}_p(\mathbb{P}, \mathbb{Q}) &= \inf \left[ \mathbb{E} \rho(X, Y)^p \right]^{1/p},
\end{align}
where the infimum is taken over all couplings \((X, Y)\) such that \(X \sim \mathbb{P}\) and \(Y \sim \mathbb{Q}\). A family of finite measures \((Q_x)_{x \geq 0}\) on \(\mathbb{R}_+\) is stochastically ordered if

\[
Q_x(\mathbb{R}_+) = \text{const}, \quad Q_x([z, \infty)) \leq Q_y([z, \infty)), \quad 0 \leq x \leq y, \quad z \geq 0.
\]

We operate on a filtered probability space \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})\). Consider a reflected jump-diffusion on \(\mathbb{R}_+\) with drift \(g : \mathbb{R}_+ \to \mathbb{R}\), diffusion \(\sigma : \mathbb{R}_+ \to \mathbb{R}_+\), and family of jump measures \((\nu_x)_{x \geq 0}\) on \(\mathbb{R}_+\). Take an \((\mathcal{F}_t)_{t \geq 0}\)-Brownian motion \(W = (W(t), t \geq 0)\).

We formally define a reflected diffusion without jumps, formalizing the informal description from the Introduction. This is an \(\mathbb{F}\)-continuous \((\mathcal{F}_t)_{t \geq 0}\)-adapted \(\mathbb{R}_+\)-valued process \(X = (X(t), t \geq 0)\) such that there exists an adapted continuous nondecreasing process \(\ell = (\ell(t), t \geq 0)\) with \(\ell(0) = 0\), which can increase only when \(X(t) = 0\), such that

\[
(6) \quad X(t) = X(0) + \int_0^t g(X(s)) \, ds + \int_0^t \sigma(X(s)) \, dW(s) + \ell(t).
\]

Add jumps to \((6)\) to get a reflected jump-diffusion by piecing out: At a point \(x \in \mathbb{R}_+\) it jumps with rate \(N(x) := \nu_x(\mathbb{R}_+)\), with jump destination distributed as \(N^{-1}(x) \nu_x(\cdot)\). This is done by piecing out: Assuming that \(N(x) = \Lambda\) is constant (as in Assumption 1 below), we fix jump times \(\tau_1 < \tau_2 < \ldots\) as Poisson point process with intensity \(\Lambda\): That is, \(\tau_{k+1} - \tau_k \sim \text{Exp}(\Lambda)\) i.i.d. with convention \(\tau_0 := 0\). Finally, we run this process as a reflected diffusion without jumps between \(\tau_k\) and \(\tau_{k+1}\). More details are in [22] Section 2 and in [10].

**Assumption 1.** The functions \(g\) and \(\sigma\) are continuous, this family of measures is weakly continuous: \(\nu_y \to \nu_x\) as \(y \to x\) weakly, and \(N(x) = \Lambda\) is a constant.

**Remark 1.** It suffices to assume \(\sup_{x \geq 0} N(x) = \overline{N} < \infty\) instead of constancy from Assumption 1. Indeed, we can replace measures \(\nu_x\) with \(\nu'_x := \nu_x + (\overline{N} - N(x)) \delta_x\). Indeed, jumping from \(x\) to \(x\) means not jumping at all, and new measures \(\nu'_x\) satisfy \(\nu'_x(\mathbb{R}_+) = \overline{N} = \text{const}\).

It was shown in [22] that under Assumption 1, there exists in the weak sense a unique in law version of this process for any initial condition \(X(0)\) (which can be random). This process is Feller continuous strong Markov, with generator \(\mathcal{L}f\) for \(f \in C^2(\mathbb{R}_+)\) with \(f'(0) = 0\):

\[
\mathcal{L}f(x) = g(x)f'(x) + \frac{\sigma^2(x)}{2}f''(x) + \int_0^\infty [f(y) - f(x)] \nu_x(dy),
\]

If the initial distribution is \(X(0) \sim \rho\), then \(X(t) \sim \rho P^t\) with \(\rho P^t\) for \(t \geq 0\) defined as

\[
\rho P^t(B) := (\rho, P^t(\cdot, B)) = \int_0^\infty \rho(dx) P^t(x, B), \quad B \subseteq [0, \infty).
\]

For example, if \(x \geq 0\), then \(\delta_x P^t(B) = P^t(x, B)\). A probability distribution \(\pi\) on \(\mathbb{R}_+\) is called a stationary distribution for this Markov process if \(\pi = \pi P^t\) for \(t \geq 0\); or, equivalently, if \(X(0) \sim \pi\) implies \(X(t) \sim \pi\) for all \(t \geq 0\). This Markov process is called stochastically ordered if for every \(t \geq 0\), the family of measures \((P^t(x, \cdot))_{x \geq 0}\) is stochastically ordered. For \(0 \leq s \leq t\) and \(x \geq 0\), define \(P^{[s,t]}(x, \cdot)\) to be the distribution in the Skorohod space \(D[s, t]\) of the process \(X = (X(u), u \in [s, t])\) starting from \(X(0) = x\).
3. Main Results

For a $\lambda > 0$, define $k(\lambda) := -\sup_{x > 0} c(x, \lambda)$, where

$$c(x, \lambda) := \lambda g(x) + \frac{\lambda^2}{2}\sigma^2(x) + \int_0^\infty \left( e^{\lambda(y-x)} - 1 \right) \nu_x(dy), \quad x \in \mathbb{R}_+.$$  \hfill (7)

Assumption 2. The family of jump measures $(\nu_x)_{x \geq 0}$ is stochastically ordered.

Assumption 3. The integral in (7) is well-defined for all $x$, $\lambda$, and $k(\lambda) > 0$ for some $\lambda > 0$.

It was shown in [22] that Assumption 3 holds under the following conditions:

$$m(x) := g(x) + \int_0^\infty (y - x) \nu_x(dy) \leq -K_1 < 0, \quad \sigma(x) \leq K_2 < \infty,$$  \hfill (8)

for constants $K_1, K_2 > 0$. This can be viewed as “effective drift” at point $x \in \mathbb{R}_+$: The sum of “true drift” $g(x)$ and “implied drift” from jumps.

Example 1. A simple example is a reflected spectrally positive Lévy process $X$, where $g$ and $\sigma$ are constant, and the jump measures $\nu_x$ are defined as follows, for some finite measure $\mu$ on $\mathbb{R}_+$: $\nu_x([0, x]) = 0$, and $\nu_x([x + z, \infty)) = \mu([z, \infty))$ for all $x, z \geq 0$. This process behaves as a Lévy process which behaves as a Brownian motion with drift $g$ and diffusion $\sigma^2$; jump times behave as a Poisson process with intensity $\mu(\mathbb{R}_+)$ and jump displacement (to the right) is distributed as the normalized measure $\mu$. As long as it hits zero, it reflects back to the positive half-line. Then the condition (8) becomes

$$g + \int_0^\infty z \mu(dz) < 0.$$  \hfill (9)

Example 2. Try $g = -1$ and $\sigma = 1$, $\mu \sim \text{Exp}(2)$. Then condition (9) holds, and

$$-k(\lambda) = c(x, \lambda) = -\lambda + \frac{\lambda^2}{2} + \int_0^\infty (e^{\lambda z} - 1) \mu(dz) = -\lambda + \frac{\lambda^2}{2} + \frac{2}{2 - \lambda} - 1.$$  

Thus $k$ assumes maximum $k(\lambda_0) = 0.0785$ for $\lambda_0 = 0.304$.

It was shown (see [22] and references therein) that under Assumptions 1, 2, 3 there exists a unique stationary distribution $\pi$ and $(\pi, V) < \infty$ for $V(x) = e^{\lambda x}$; and the $V$-distance between $P^t(x, \cdot)$ and $\pi(\cdot)$ converges to 0 at least as $e^{-k(\lambda)t}$. We show a similar result for the Wasserstein distance.

Theorem 1. Under Assumptions 1, 2, 3, for every $p \geq 1$, for $C := e\lambda / \lambda$, we have

(a) For all $x_1, x_2 \geq 0$ and $t \geq 0$, we have:

$$\mathcal{W}_p(P^t(x_1, \cdot), P^t(x_2, \cdot)) \leq C \cdot \exp \left[ p^{-1} \lambda(x_1 \lor x_2) \right] \exp \left[ -p^{-1} k(\lambda) pt \right].$$

(b) The same estimate holds for path distributions: For $x_1, x_2 \geq 0$ and $T > t > 0$,

$$\mathcal{W}_p \left( P^{[t, T]}(x_1, \cdot), P^{[t, T]}(x_2, \cdot) \right) \leq C \cdot \exp \left[ p^{-1} \lambda(x_1 \lor x_2) \right] \exp \left[ -p^{-1} k(\lambda) t \right].$$

(c) For probability measures $p_1$ and $p_2$ on $\mathbb{R}_+$ and for $t \geq 0$:

$$\mathcal{W}_p(p_1 P^t, p_2 P^t) \leq C \cdot [(p_1, V) + (p_2, V)]^{1/p} \exp \left[ -p^{-1} k(\lambda) t \right], \quad V(x) := e^{\lambda x}.$$

(d) For every $x \geq 0$ and $t \geq 0$, we get:

$$\mathcal{W}_p(P^t(x, \cdot), \pi) \leq C \cdot [(\pi, V) + V(x)]^{1/p} \exp \left[ -p^{-1} k(\lambda) t \right].$$
Assumption 4. The function \( \sigma \) is constant. For every \( z \in \mathbb{R} \), the function \( x \mapsto \nu_z((x+z, \infty)) \) is nonincreasing. There exists a constant \( G \) such that \( x \mapsto g(x) - Gx \) is nonincreasing.

Theorem 2. Under Assumptions 1, 2, 3, 4, if \( k(\lambda) > pG, p \geq 1 \), define \( K := p^{-1}k(\lambda) - G \).

(a) For all \( x_1, x_2 \geq 0, t \geq 0 \), we have:
\[
\mathcal{W}_p(P^t(x_1, \cdot), P^t(x_2, \cdot)) \leq \exp(p^{-1} \lambda(x_1 \vee x_2)) |x_1 - x_2| \exp(-Kt).
\]

(b) If \( G \leq 0 \), the same estimate holds for path distributions:
\[
\mathcal{W}_p(P^{[s,T]}(x_1, \cdot), P^{[s,T]}(x_2, \cdot)) \leq \exp(p^{-1} \lambda(x_1 \vee x_2)) |x_1 - x_2| \exp(-Kt).
\]

(c) For probability measures \( \rho_1, \rho_2 \) on \( \mathbb{R}_+ \), there exists a constant \( C_0 > 0 \) such that
\[
\mathcal{W}_p(\rho_1 P^t, \rho_2 P^t) \leq C_0 \exp(-Kt), \quad V(x) := e^{\lambda x}, \quad t \geq 0.
\]

(d) For every \( t, x \geq 0 \), we get:
\[
\mathcal{W}_p(P^t(x, \cdot), \pi) \leq [(\pi, V) + V(x)]^{1/p} \exp(-Kt).
\]

As follows from [23, Chapter 6], convergence in Wasserstein distance of order \( p \) implies convergence of moments up to \( p \)th order: For every continuous function \( f : \mathbb{R}_+ \rightarrow \mathbb{R} \) with \( |f(x)| \leq 1 + x^p \) for \( x \geq 0 \), there exists a constant \( C(f) > 0 \) such that
\[
|\langle \nu_1, f \rangle - \langle \nu_2, f \rangle| \leq C(f) \mathcal{W}_p^{p}(\nu_1, \nu_2).
\]

Therefore, convergence of measures in Wasserstein distance with rate \( e^{-kt} \) (obtained in the main theorems) implies convergence with test function \( f \) with rate \( e^{-kpt} \). Convergence in \( V \)-norm for the function \( V(x) = e^{\lambda x} \) implies convergence of all moments. Indeed, a simple calculus exercise shows (and we will make use of it in the proofs) that for every \( p \geq 1 \) there exists an \( a(p) > 0 \) such that \( 1 + x^p \leq a(p)V(x), x \geq 0 \). Thus for the \( f \) as above, we get:
\[
|\langle \nu_1, f \rangle - \langle \nu_2, f \rangle| \leq a(p) \| \nu_1 - \nu_2 \|_V.
\]

Therefore, convergence in the norm \( \| \cdot \|_V \) with rate \( e^{-xt} \) implies convergence with test function \( f \) with the same rate. To find whether convergence in the \( \| \cdot \|_V \)-norm or Wasserstein distance is faster, we need to compare \( kp \) with \( \lambda \). Convergence rate in Theorem 1 is with the same rate as in [22] for the norm \( \| \cdot \|_V \). However, in Theorem 2 convergence rate can be faster than in [22] for the norm \( \| \cdot \|_V \), if only \( K > k(\lambda) \).

Example 3. For \( G = 0 \) (including Example 1), Theorem 2 does not improve upon Theorem 1. But if \( G < 0 \), then Theorem 2 can give a better rate of convergence than Theorem 1.

For example, for reflected Ornstein–Uhlenbeck process (without jumps)
\[
dX(t) = -a(m + X(t)) dt + \sigma dW(t) + d\ell(t), \quad t \geq 0,
\]
with constants \( a, m, \sigma > 0 \). Then \( G := -a \), and we can take
\[
\lambda = \frac{am}{\sigma^2}, \quad k(\lambda) = \frac{a^2m^2}{2\sigma^2}, \quad K := \frac{a^2m^2}{2\sigma^2p} + a.
\]

This convergence rate \( pk \) in Wasserstein distance is faster than the rate \( k(\lambda) \) of convergence in the norm \( \| \cdot \|_V \) from [22]:
\[
a^2m^2 \frac{1}{2\sigma^2} + ap = pk > \frac{a^2m^2}{2\sigma^2}.
\]
4. Proofs

4.1. Proof of Theorem 1. (a) The idea is the same as in [14, 22]: Couple two copies \( X_1 \) and \( X_2 \) of this processes starting from \( x_1 \) and \( x_2 \). Without loss of generality, assume \( 0 \leq x_1 \leq x_2 \). It was shown in [22, Lemma 4.2] that this process is stochastically ordered if the family of measures \((\nu_x)_{x \geq 0}\) is stochastically ordered, which is assumed by Assumption 2. Thus we can assume \( X_1(t) \leq X_2(t) \) for all \( t \geq 0 \). Take \( \tau := \inf\{t \geq 0 \mid X_2(t) = 0\} \), then \( X_1(\tau) = 0 \) and we can assume \( X_1(t) = X_2(t) \) for \( t > \tau \). Using calculus, we compute

\[
(10) \quad x^p \leq ae^{\lambda x} = aV(x), \quad x \geq 0, \quad a := (pe/\lambda)^p. 
\]

Since \( X_2(t) \geq |X_2(t) - X_1(t)| \) for all \( t \geq 0 \), \( X_1(t) = X_2(t) \) for \( t \geq \tau \), by \((10)\),

\[
(11) \quad \mathbb{E} \left| X_1(t) - X_2(t) \right|^p = \mathbb{E} \left[ 1_{\{\tau > t\}} \right] \mathbb{E} \left| X_1(t) - X_2(t) \right|^p 
\leq \mathbb{E} \left[ X_2^p(t) 1_{\{\tau > t\}} \right] \leq a \cdot \mathbb{E} \left[ V(X_2(t)) 1_{\{\tau > t\}} \right], \quad t \geq 0. 
\]

From [22, Section 5, (5.9)], the following process is an \((\mathcal{F}_t)_{t \geq 0}\)-supermartingale:

\[
(12) \quad (e^{k(\lambda)(t \wedge \tau)} V(X_2(t \wedge \tau)), t \geq 0). 
\]

Therefore, for all \( t \geq 0 \) we get:

\[
(13) \quad e^{k(\lambda)t} \mathbb{E} \left[ V(X_2(t)) 1_{\{\tau > t\}} \right] \leq \mathbb{E} \left[ e^{k(\lambda)(t \wedge \tau)} V(X_2(t \wedge \tau)) \right] \leq V(x_2).
\]

Combining \((11)\) and \((13)\), we get:

\[
(14) \quad \mathbb{E} \left| X_1(t) - X_2(t) \right|^p \leq ae^{-k(\lambda)t} V(x_2) \leq ae^{-k(\lambda)t} V(x_1 \vee x_2).
\]

Raising it to the power \(1/p\), we get:

\[
W_p(P^t(x_1, \cdot), P^t(x_2, \cdot)) \leq \left[ \mathbb{E} \left| X_1(t) - X_2(t) \right|^p \right]^{1/p} 
\leq a^{1/p} \exp \left[ -p^{-1}k(\lambda)t \right] V^{1/p}(x_2) = \lambda^{-1} pe \exp \left[ p^{-1}x_2 - p^{-1}k(\lambda)t \right] 
\leq \lambda^{-1} pe \exp \left[ p^{-1}x_1 \vee x_2 - p^{-1}k(\lambda)t \right].
\]

(b) We modify \((11)\) and again use \((10)\):

\[
(15) \quad \mathbb{E} \sup_{t \leq s \leq T} \left| X_1(s) - X_2(s) \right|^p = \mathbb{E} \left[ \sup_{t \leq s \leq T} \left| X_1(s) - X_2(s) \right|^p \cdot 1_{\{\tau > t\}} \right] 
\leq \mathbb{E} \left[ \sup_{t \leq s \leq T} X_2^p(s) \cdot 1_{\{\tau > t\}} \right] \leq a \cdot \mathbb{E} \left[ \sup_{t \leq s \leq T} V(X_2(t)) \cdot 1_{\{\tau > t\}} \right].
\]

We modify \((13)\) by applying Ville’s maximal inequality from [6, Exercise 4.8.2] to \((12)\):

\[
(16) \quad e^{k(\lambda)t} \cdot \mathbb{E} \left[ \sup_{t \leq s \leq T} V(X_2(t)) \cdot 1_{\{\tau > t\}} \right] \leq \mathbb{E} \left[ e^{k(\lambda)(t \wedge \tau)} V(X_2(t \wedge \tau)) \right] \leq V(x_2).
\]

Apply \((5)\) to combined \((15)\) and \((16)\) and complete the proof.

(c) In \((11)\) from the proof of (a), replace \( x_1 \vee x_2 \) with \( x_1 + x_2 \), and integrate the right-hand side with respect to \( x_1 \sim \rho_1 \) and \( x_2 \sim \rho_2 \). Thus \( V(x_1 \vee x_2) \leq V(x_1) + V(x_2) \). Then we get:

\[
\int_0^\infty \int_0^\infty [V(x_1) + V(x_2)] \rho_1(dx_1) \rho_2(dx_2) = (\rho_1, V) + (\rho_2, V).
\]

Then raise this estimate to the power \(1/p\), and complete the proof.

(d) Apply (c) to \( \rho_1 = \pi \) and \( \rho_2 = \delta_x \), with \( \delta_x P^t = P^t(x, \cdot) \) and \( \pi P^t = \pi \) for \( t \geq 0 \).
4.2. **Proof of Theorem 2.** Without loss of generality, we can assume $x_1 \leq x_2$. The following lemma is proved at the end of this subsection.

**Lemma 1.** We can couple $X_1$ and $X_2$ starting from $X_1(0) = x_1$ and $X_2(0) = x_2$ so that

$$0 \leq X_2(t) - X_1(t) \leq (x_2 - x_1)e^{Gt}, \quad t < \tau := \inf\{t \geq 0 \mid X_2(t) = 0\}.$$

(a) Adapt the proof of Theorem 1. Use (11) and Lemma 1

$$\mathcal{W}_p^p(P^t(x_1, \cdot), P^t(x_2, \cdot)) \leq \mathbb{E} [(X_2(t) - X_1(t))^p 1_{\{\tau > t\}}] \leq \mathbb{E} [(x_2 - x_1)^p e^{Gpt} 1_{\{\tau > t\}}].$$

Since $V(x) \geq 1$ for $x \geq 0$, we get:

$$e^{Gpt} \cdot \mathbb{E} [1_{\{\tau > t\}}] \leq e^{Gpt} \cdot \mathbb{E} [1_{\{\tau > t\}}V(X_2(t))] \leq e^{(Gp-k(\lambda))t} \cdot \mathbb{E} [e^{k(\lambda)(t\wedge\tau)} V(X_2(t \wedge \tau))].$$

By the supermartingale property of $(e^{k(\lambda)(t\wedge\tau)} V(t \wedge \tau), t \geq 0)$, we get:

$$\mathbb{E} [e^{k(\lambda)(t\wedge\tau)} V(X_2(t \wedge \tau))] \leq V(x_2).$$

Recall the definition of $K$. Raising this inequality to the power $1/p$, we complete the proof.

(b) The proof is similar to the one from (a). Using (5), we get:

$$\mathcal{W}_p^p(P^{[t,T]}(x_1, \cdot), P^{[t,T]}(x_2, \cdot)) \leq \mathbb{E} \sup_{t \leq s \leq T} |X_1(t) - X_2(t)|^p.$$

Since $G \leq 0$, we get from Lemma 1

$$\sup_{t \leq s \leq T} |X_1(t) - X_2(t)|^p \leq \sup_{t \leq s \leq T} e^{Gps}|x_1 - x_2|^p = e^{Gpt}|x_1 - x_2|^p. \quad (17)$$

Taking expectation in (17), applying (3), raising to the power $1/p$, we complete the proof.

(c, d) The proofs are similar to that of (c, d) from Theorem 1.

**Proof of Lemma 1.** We use **piecing out**, a classic method for adding jumps to continuous processes, described in [10], [22]. Since the intensity $\Lambda := \nu_x(\mathbb{R}_+)$ is constant, we can couple the processes with simultaneous jumps $0 = \tau_0 < \tau_1 < \tau_2 < \ldots$ as Poisson process with rate $\Lambda$. Let us couple them so that the following is true:

$$X_1(t) \leq X_2(t), \quad t \geq 0, \quad \text{and for } k = 0, 1, \ldots, \text{if } \tau > \tau_{k+1}, \text{then}$$

$$d(X_2(t) - X_1(t)) \leq G(X_2(t) - X_1(t)) \, dt, \quad \tau_k \leq t < \tau_{k+1}; \quad X_2(\tau_{k+1}) - X_2(\tau_{k+1}-) \leq X_1(\tau_{k+1}) - X_1(\tau_{k+1}-). \quad (18)$$

Assume we proved (18) and (19). Use induction over $k$ to show the statement of the lemma for $\tau_k < t \leq \tau_{k+1}$, assuming this for $t = \tau_k$. By Gronwall’s lemma, since $X_2 - X_1$ is continuous,

$$X_2(t) - X_1(t) \leq e^{G(t-\tau_k)}(X_2(\tau_k) - X_1(\tau_k)), \quad \tau_k \leq t < \tau_{k+1}.$$

Combining this observation with the second result in (19), we complete the proof of induction step and with it the proof of Lemma 1. To prove (18) and the first line in (19), for $t \in (\tau_k, \tau_{k+1})$, use induction over $k = 0, 1, \ldots$. During $(\tau_k, \tau_{k+1})$, $X_1$ and $X_2$ behave as copies of a reflected diffusion with drift $g$ and diffusion $\sigma^2$. Thus

$$dX_i(t) = g(X_i(t)) \, dt + \sigma \, dW(t) + d\ell_i(t), \quad i = 1, 2,$$

where $\ell_i$ is a continuous nondecreasing process which can increase only when $X_i(t) = 0$, and $\ell_i(0) = 0$. We coupled them with the same driving Brownian motion $W$. Since $X_1(\tau_k) \leq X_2(\tau_k)$,
\( X_2(\tau_k) \) by the induction hypothesis, we get (18) for \( t \in (\tau_k, \tau_{k+1}) \). For \( t < \tau \), \( \ell_2(t) = 0 \), and \( \ell_1 \) is nondecreasing, thus \( d\ell_1(t) \geq 0 \). Assumption 4 about the function \( g \), together with \( X_1(t) \leq X_2(t) \), implies \( g(X_2(t)) - g(X_1(t)) \leq G(X_2(t) - X_1(t)) \). Therefore,

\[
d(X_2(t) - X_1(t)) = [g(X_2(t)) - g(X_1(t))] \, dt - d\ell_1(t) \\
\leq [g(X_2(t)) - g(X_1(t))] \, dt \leq G \, [X_2(t) - X_1(t)] \, dt.
\]

This proves the first line in (19). Letting \( t \uparrow \tau_{k+1} \) in \( X_1(t) \leq X_2(t) \), we get: \( y_1 := X_1(\tau_{k+1}^-) \leq y_2 := X_2(\tau_{k+1}^-) \). By Assumption 2 and Assumption 4 for jump measures, there is a coupling \((z_1, z_2)\) of normalized jump measures:

\[ z_1 \sim \Lambda^{-1} \nu_{y_1}, \quad z_2 \sim \Lambda^{-1} \nu_{y_2}, \quad z_2 \leq z_1, \quad z_1 + y_1 \leq z_2 + y_2, \]

Let \( X_i(\tau_{k+1}) := y_i + z_i, i = 1, 2 \) to prove (18) and the second line in (19) for \( t = \tau_{k+1} \).

**References**

[1] Dominique Bakry, Patrick Cattiaux, Arnaud Guillin (2008). Rate of Convergence of Ergodic Continuous Markov Chains: Lyapunov vs Poincare. *Journal of Functional Analysis* **254**, 727–754.

[2] Yana Belopolskaya, Guodong Pang, Andrey Sarantsev, Yuri Suhov (2019). Stationary Distributions and Convergence for M/M/1 Queues in Interactive Random Environment. To appear in *Queueing Systems: Theory and Applications*.

[3] Oleg Butkovski (2014). Subgeometric Rates of Convergence of Markov Processes in the Wasserstein Metric. *Annals of Applied Probability* **24**, 526–552.

[4] Patrick Cattiaux, Arnaud Guillin (2014). Semi Log-Concave Markov Diffusions. *Lecture Notes in Mathematics* **2123**, 231–292, Springer.

[5] Douglas Down, Sean P. Meyn, Richard L. Tweedie (1995). Exponential and Uniform Ergodicity of Markov Processes. *Annals of Probability* **23**, 1671–1691.

[6] Richard Durrett (2019). *Probability: Theory and Examples*. 5th ed., Cambridge University Press.

[7] Pierre-Olivier Goffard, Andrey Sarantsev (2019). Exponential Convergence Rate of Ruin Probabilities for Level-Dependent Levy-Driven Risk Processes. *Journal of Applied Probability* **56**, 1244–1268.

[8] Martin Hairer, Jonathan Mattingly (2011). Yet Another Look at Harris’ Ergodic Theorem for Markov Chains. *Seminar on Stochastic Analysis, Random Fields and Applications* **6**, 109–117.

[9] Tomoyuki Ichiba, Andrey Sarantsev (2019). Convergence and Stationary Distributions for Walsh Diffusions. *Bernoulli* **25**, 2439–2478.

[10] Nobuyuki Ikeda, Masao Nanasawa, Shinzo Watanabe (1966). A Construction of Markov Processes by Piecing Out. *Proceedings of the Japan Academy of Sciences* **42**, 370–375.

[11] Torgny Lindvall (1983). On Coupling of Diffusion Processes. *Journal of Applied Probability* **20**, 82–93.

[12] Torgny Lindvall (2002). *Lectures on the Coupling Method*. Dover.

[13] Torgny Lindvall, L. C. G. Rogers (1986). Coupling of Multidimensional Diffusions by Reflection. *Annals of Probability* **14**, 860–872.

[14] Robert B. Lund, Sean P. Meyn, Richard L. Tweedie (1996). Computable Exponential Convergence Rates for Stochastically Ordered Markov Processes. *Annals of Applied Probability* **6**, 218–237.

[15] Sean P. Meyn, Richard L. Tweedie (1993). Stability of Markovian Processes II. Continuous-Time Processes and Sampled Chains. *Advances in Applied Probability* **25**, 487–517.

[16] Sean P. Meyn, Richard L. Tweedie (1993). Stability of Markovian Processes III. Foster-Lyapunov Criteria for Continuous-Time Markov Processes. *Advances in Applied Probability* **25**, 518–548.

[17] Sean P. Meyn, Richard L. Tweedie (1994). Computable Bounds for Geometric Convergence Rates of Markov Chains. *Annals of Applied Probability* **4**, 981–1011.

[18] Soumik Pal (2012). Concentration for Multidimensional Diffusions and their Boundary Local Times. *Probability Theory and Related Fields* **154**, 225–254.

[19] Soumik Pal and Andrey Sarantsev (2019). A Note on Transportation Cost Inequalities for Diffusions with Reflections. *Electronic Communications in Probability* **24**, # 21.
[20] **Jeffrey S. Rosenthal** (1995). Minorization Conditions and Convergence Rates for Markov Chain Monte Carlo. *Journal of the American Statistical Association* **90**, 558–566.

[21] **Paul-Marie Samson** (2000). Concentration of Measure Inequalities for Markov Chains and Φ-Mixing Processes. *Annals of Probability* **28**, 416–461.

[22] **Andrey Sarantsev** (2016). Explicit Rates of Exponential Convergence for Reflected Jump-Diffusions on the Half-Line. *Latin American Journal of Probability & Mathematical Statistics* **13**, 1069–1093.

[23] **Cedric Villani** (2008). *Optimal Transport: Old and New*. Grundlehren der Mathematischen Wissenschaften **338**, Springer.

**University of Nevada in Reno, Department of Mathematics and Statistics**

*E-mail address: asarantsev@unr.edu*