GEOMETRIC ERGODICITY IN A WEIGHTED SOBOLEV SPACE

BY ADITHYA DEVRAJ[1,*], IOANNIS KONTOYIANNIS[2] AND SEAN MEYN[1,**]

[1]*Department of Electrical & Computer Engineering, University of Florida,* *adithyamdevraj@ufl.edu;* **meyn@ece.ufl.edu*
[2]*Department of Engineering, Cambridge University, ik355@cam.ac.uk*

For a discrete-time Markov chain $X = \{X(t)\}$ evolving on $\mathbb{R}^\ell$ with transition kernel P, natural, general conditions are developed under which the following are established:

(i) The transition kernel P has a purely discrete spectrum, when viewed as a linear operator on a weighted Sobolev space $L_\infty^{v,1}$ of functions with norm,

$$\|f\|_{v,1} = \sup_{x \in \mathbb{R}^\ell} \frac{1}{v(x)} \max\{|f(x)|, |\partial_1 f(x)|, \dots, |\partial_\ell f(x)|\},$$

where $v \colon \mathbb{R}^\ell \to [1, \infty)$ is a Lyapunov function and $\partial_i := \partial/\partial x_i$.

(ii) The Markov chain is geometrically ergodic in $L_\infty^{v,1}$: There is a unique invariant probability measure π and constants $B < \infty$ and $\delta > 0$ such that, for each $f \in L_\infty^{v,1}$, any initial condition $X(0) = x$, and all $t \geq 0$:

$$|\mathsf{E}_x[f(X(t))] - \pi(f)| \leq B\|f\|_{v,1} e^{-\delta t} v(x),$$

$$\|\nabla \mathsf{E}_x[f(X(t))]\|_2 \leq B\|f\|_{v,1} e^{-\delta t} v(x),$$

where $\pi(f) = \int f \, d\pi$.

(iii) For any function $f \in L_\infty^{v,1}$ there is a function $h \in L_\infty^{v,1}$ solving Poisson's equation:

$$h - Ph = f - \pi(f).$$

Part of the analysis is based on an operator-theoretic treatment of the sensitivity process that appears in the theory of Lyapunov exponents. Relationships with topological coupling, in terms of the Wasserstein metric, are also explored.

1. Introduction. Consider a discrete-time Markov chain $X = \{X(t) : t \geq 0\}$ taking values in $\mathsf{X} = \mathbb{R}^\ell$, equipped with its associated Borel σ-field $\mathcal{B}$. Throughout the paper (except where explicitly noted otherwise, in particular see Section 3.2), the process X is assumed to be defined by the nonlinear state space model

$$(1) \qquad X(t+1) = a(X(t), N(t+1)), \quad t \in \mathbb{Z}_+,$$

where $N = \{N(t) : t = 0, 1, 2, \dots\}$ is a sequence of $\mathbb{R}^m$-valued, independent and identically distributed random variables, and $a : \mathbb{R}^{\ell \times m} \to \mathbb{R}^\ell$ is continuous, so that each realization $X(t)$ is a continuous function of $X(0) = x$.

The distribution of X is described by its initial state $X(0) = x \in \mathsf{X}$ and its transition semigroup: For any $t \geq 0$, $x \in \mathsf{X}$, $A \in \mathcal{B}$,

$$P^t(x, A) := \mathsf{P}_x\{X(t) \in A\} := \Pr\{X(t) \in A | X(0) = x\},$$

with the usual convention that P^1 is simply denoted P. For the Markov chain described by (1), it follows that $P(x, A) = \Pr\{a(x, N(1)) \in A\}$.

Recall that the kernel P^t acts as a linear operator on functions $f : X \to \mathbb{R}$ on the right and on signed measures ν on $(X, \mathcal{B})$ on the left, respectively, as

$$P^t f(x) = \int f(y) P^t(x, dy), \qquad \nu P^t(A) = \int \nu(dx) P^t(x, A), \quad x \in X, A \in \mathcal{B},$$

whenever the integrals exist. Also, for any signed measure ν on $(X, \mathcal{B})$ and any function $f : X \to \mathbb{R}$ we write $\nu(f) := \int f \, d\nu$, whenever the integral exists. In this paper, we constrain the domain of functions f to a Banach space defined with respect to a weighted L_∞ norm.

Specifically, given a fixed continuous function $v : X \to [1, \infty)$, the v-norm of any measurable function $f : X \to \mathbb{R}$ is denoted

$$(2) \qquad \qquad \|f\|_v := \sup_x \frac{|f(x)|}{v(x)},$$

and the corresponding Banach space L_∞^v is defined as, $L_\infty^v := \{f : X \to \mathbb{R} : \|f\|_v < \infty\}$. An analogous weighted norm is defined for signed measures μ on $(X, \mathcal{B})$ via

$$\|\mu\|_v := \sup \left\{ \frac{|\mu(h)|}{\|h\|_v} : h \in L_\infty^v, \|h\|_v \neq 0 \right\},$$

and we denote by $\mathcal{M}_1^v$ the space of signed measures μ with $\|\mu\|_v < \infty$.

The Markov chain X is v-*uniformly ergodic* [33, 41] whenever there exists a function v, a unique invariant probability measure π, and constants $b_0 < \infty$ and $0 < \rho_0 < 1$, such that, for each function $f \in L_\infty^v$,

$$(3) \qquad \qquad \left| \mathsf{E}[f(X(t)) \mid X(0) = x] - \pi(f) \right| \leq b_0 \rho_0^t \|f\|_v v(x), \quad t \geq 0,$$

where $\pi(f) = \int f \, d\pi$. It is well known that this is equivalent to the existence of a Lyapunov function satisfying the drift condition (V4) [33, 41].

1.1. *Motivation and background.* Let $c : X \to \mathbb{R}$ be a given function on the state space of X. One starting point for the classical study of the long-term behavior of X is the development of conditions for the existence of the mean ergodic limit

$$(4) \qquad \qquad \overline{c} := \lim_{n \to \infty} \frac{1}{n} \sum_{t=0}^{n-1} \mathsf{E}[c(X(t)) \mid X(0) = x],$$

and of the function

$$(5) \qquad \qquad h(x) := \sum_{t=0}^{\infty} \mathsf{E}[c(X(t)) - \overline{c} \mid X(0) = x], \quad x \in X,$$

which can be shown to be a solution of the associated *Poisson equation*

$$(6) \qquad \qquad h(x) - \mathsf{E}[h(X(1)) \mid X(0) = x] = c(x) - \overline{c}, \quad x \in X.$$

For example, if X is v-uniformly ergodic, then in addition to the convergence (3) of $P^t(x, \cdot)$ to its unique invariant probability measure π, the ergodic averages of $c(X(t))$ converge a.s. to $\overline{c} = \pi(c)$, and their associated central limit theorem variance is naturally expressed in terms of h [2, 41],

$$\sigma^2 = \pi(h^2 - (Ph)^2).$$

Moreover, if $c \in L_\infty^v$ then h is also in L_∞^v [25].

A closely related object of interest is the collection, for each $\alpha \in (0, 1)$, of the functions

$$(7) \qquad \qquad h_\alpha(x) := \sum_{t=0}^{\infty} \alpha^t \mathsf{E}[c(X(t)) \mid X(0) = x], \quad x \in X,$$

where each h_α can be viewed as the result of the action of the resolvent kernel,

$$R_\alpha := \sum_{t=0}^{\infty} \alpha^t P^t,$$

on the function c. Again, under v-uniform ergodicity, $h_\alpha \in L_\infty^v$ for any $\alpha < 1$, whenever $c \in L_\infty^v$ [33].

The main goal of the present work is to develop natural conditions that guarantee appropriate *smoothness* properties of $\overline{c}$, h and h_α. In particular, as described next, we show that the derivative of $P^t c(x) = \mathsf{E}[c(X(t))|X(0) = x]$ with respect to the initial condition $X(0) = x$ converges to zero; we provide series representations, analogous to (5) and (7), for the *derivatives* of h and h_α, and we also obtain bounds for those derivatives.

In addition to the theoretical interest of these results, we are also motivated in part by related questions and applications in stochastic control. In that context, c is viewed as a *one-step cost function*, α is the *discount factor*, $\overline{c}$ is the *average cost*, $h(x)$ is the *relative value function* and $h_\alpha(x)$ is the *total discounted cost*. The present results provide a foundation for a new approach to approximate dynamic programming developed in [13], to gain approximation for the feedback particle filter [38, 44], and to efficient temporal difference learning algorithms [12].

1.2. *Overview of main results.* Suppose that the function a appearing in (1) is continuously differentiable. This justifies the following definition of the $\ell \times \ell$ *sensitivity process* $S = \{S(t) : t \geq 0\}$, whose (i, j) component is defined at time t by

$$(8) \qquad S_{i,j}(t) := \frac{\partial X_i(t)}{\partial X_j(0)}, \quad 1 \leq i, j \leq \ell.$$

From (1), the sensitivity process evolves according to the random linear system

$$(9) \qquad S(t+1) = A(t+1)S(t), \qquad S(0) = I,$$

where $A^T(t) := \nabla_x a(X(t-1), N(t))$.

For any function $f \in C^1$ and all $t \geq 0$, we write

$$(10) \qquad \nabla^S f(X(t)) := S^T(t)\nabla f(X(t)).$$

It follows from the chain rule that this coincides with the gradient of $f(X(t))$ with respect to the initial condition $X(0)$. This interpretation of (10) motivates the introduction of a new semigroup $\{Q^t : t \geq 0\}$ of operators acting on measurable functions $g : \mathsf{X} \to \mathbb{R}^\ell$: For $t \geq 1$,

$$(11) \qquad Q^t g(x) := \mathsf{E}[S^T(t)g(X(t)) \mid X(0) = x],$$

and $Q^0 g = g$. Provided we can exchange the gradient and the expectation, and writing as usual $\mathsf{E}_x(\cdot)$ for the conditional expectation $\mathsf{E}(\cdot|X(0) = x)$

$$\frac{\partial}{\partial x_i} \mathsf{E}_x[f(X(t))] = \mathsf{E}_x[[\nabla^S f(X(t))]_i],$$

which implies that

$$\nabla P^t f(x) = \mathsf{E}_x[\nabla^S f(X(t))] = Q^t \nabla f(x).$$

Main results. The main contribution of this paper is the justification of the above manipulations, within an appropriate Banach space setting. Specifically, for all functions $c : \mathsf{X} \to \mathbb{R}$ in an appropriate space, we identify general, natural conditions under which the following are established:

(i) Not only does $P^t c$ converge to $\bar{c} = \pi(c)$ as $t \to \infty$, but also the gradient $\nabla P^t c$ of $P^t c$ with respect to the initial condition $X(0) = x$ converges to zero, at a uniformly geometric rate; cf. Theorem 2.1.

(ii) The solution h of the Poisson equation defined in (5) is differentiable, and the following representation is obtained in Theorem 2.3 for its gradient:

$$\nabla h = \Omega \nabla c := \sum_{t=0}^{\infty} Q^t \nabla c,$$

where $\{Q^t\}$ is the semigroup defined in (11) in terms of the sensitivity process.

(iii) Similarly, for any $\alpha \in (0, 1)$, the following representation is obtained in Theorem 2.4 for the gradient of the total discounted cost h_α defined in (7):

$$\tag{12} \nabla h_\alpha = \Omega_\alpha \nabla c := \sum_{t=0}^{\infty} \alpha^t Q^t \nabla c.$$

1.3. *Prior research.*

Our main assumptions are minor variants of those used in much of the prior work on ergodic theory for ψ-irreducible Markov chains. In particular, the Lyapunov drift condition (DV3) is assumed throughout: For nonnegative, continuous functions $V : \mathsf{X} \to \mathbb{R}_+$, $W : \mathsf{X} \to [1, \infty)$, $\delta > 0$, and a compact set C:

$$\tag{13} \log \mathsf{E}\big[\exp\{V(X(t+1)) - V(X(t))\} \mid X(t) = x\big] \le -\delta W(x), \quad x \in C^c.$$

Condition (DV3) is an essential ingredient in much of the prior work of Donsker and Varadhan on large deviation theory for Markov models [15–17], and it is used to bound rates of convergence for a Markov chain for both mean and "multiplicative" ergodic theory in [4, 34, 36]. Also, the discrete-time counterpart of [27], Assumption 4, equation (15), implies (18) with V having bounded sublevel sets, and hence (DV3) for $W = V = \log v$.

The value of (DV3) is most clear when the sublevel sets of the function W are compact. In this case, a n-step transition kernel can be approximated by its truncation to a compact set arbitrarily closely in an associated induced operator norm [4, 34, 36]; see also [48, 50] on the implications of truncation approximations. The main assumptions of the paper summarized in Section 2.2 impose (DV3) and minor additional assumptions so that a truncation approximation is valid in the stronger norm used in this paper.

There has been increasing interest in finding connections between (DV3) and logarithmic Sobolev or Poincaré inequalities [10, 11, 26]. The implications of this and similar drift conditions are the main focus of [41]. In particular, in this monograph and subsequent papers [25, 33, 35], drift conditions are used to obtain existence and bounds on solutions to Poisson's equation. A log-Sobolev inequality is the condition used in [38] to establish the existence of a smooth solution to Poisson's equation for a diffusion.

The weaker drift condition (V4), which is known to characterize exponential ergodicity [18, 33, 41], is used to obtain bounds on solutions to Poisson's equation in [25]; see also [9]. In the special case of elliptic diffusions, and with Lyapunov function $v(x) = \|x\|^p$, the paper [42] extends [25] to obtain bounds on the growth rate of the gradient of the solution h.

In more recent work on ergodicity in a weighted Lipschitz norm, explicit bounds on the gradient are obtained in [20], and similar results follow (even if not stated explicitly) whenever explicit geometric bounds are obtained in the ergodic theorem with respect to this norm, as in [28].

Poisson's equation is one tool used in addressing parametric sensitivity in Markov chains, starting with the 50-year-old work of Schweitzer [47]; *infinitesimal perturbation analysis* is a well-known application of these techniques [8]. A modern treatment is contained in the very recent work [45]. The focus of this paper is on sensitivity with respect to the initial condition of the Markov chain rather than parametric uncertainty, so there is no obvious relationship with this prior research.

A history of topological notions of ergodicity up to 1993 is contained in Chapter 6 of [41]. In this context of this paper, the most relevant references involve equicontinuity of the transition semigroup—a condition introduced by Jamison [31]. The theory of Lyapunov exponents may be regarded as a technique to verify whether some form of equicontinuity holds. The sensitivity process defined in (8) is used to define the Lyapunov exponent

$$\Lambda_1 := \lim_{t \to \infty} \frac{1}{t} \log(\|\mathcal{S}(t)\|);$$

here and in (15,18), $\| \cdot \|$ can be taken to be any matrix norm. A negative exponent implies a topological notion of coupling: Suppose that Λ_1 is a negative constant, independent of the initial condition. If X and X' are two realizations of the Markov chain with different initial states, it follows from the mean value theorem that, with probability one,

$$\lim_{t \to \infty} \|X(t) - X'(t)\|_2 = 0, \tag{14}$$

and that this convergence is geometrically fast, with rate $e^{t\Lambda_1}$. Much of the earlier relevant research, including the study of the corresponding pth mean

$$\bar{\Lambda}_p := \lim_{t \to \infty} \frac{1}{t} \log(\mathsf{E}_x[\|\mathcal{S}(t)\|^p]), \tag{15}$$

is for diffusion processes in continuous time [1, 5, 37].

Verifiable conditions for a negative Lyapunov exponent are established in [3] for a class of hidden Markov models, and in [30] for a general class of stochastic sequences of the form (1); coupling results that suggest a negative Lyapunov exponent are established in [27] for a class of diffusions. In these papers, the main results are established *without* ψ-irreducibility. As discussed in [41], Section 6.4, in such cases it is impossible to establish convergence of the Markov semigroup in total variation, so it is natural to instead rely on topological notions of convergence or coupling.

The coupling result (14) depends of course on how the different processes X and X' are constructed. Since our main interest is in convergence of moments, it is natural to introduce an expectation in (14) and minimize over all couplings: for initial probability measures $X(0) \sim \mu_0$, $X'(0) \sim \mu_0'$, denote μ_t and μ_t' the distributions at time t:

$$\mu_t(\cdot) = \int \mu_0(dx) P^t(x, \cdot), \qquad \mu_t'(\cdot) = \int \mu_0'(dx) P^t(x, \cdot).$$

The L_1-Wasserstein distance between these two probability measures is denoted

$$W_1(\mu_t, \mu_t') = \inf_\Gamma \mathsf{E}[\|X - X'\|_2], \tag{16}$$

where the infimum is over all probability measures Γ on $\mathcal{B}(\mathsf{X} \times \mathsf{X})$ with marginals given by μ_1 and μ_t', and the expectation is taken with $(X, X') \sim \Gamma$; see [14, 20, 21, 27] for background. The Markov chain is ergodic in the L_1-Wasserstein metric if there exists an invariant measure π such that, for every $x \in \mathsf{X}$, with $\mu_0 = \delta_x$,

$$\lim_{t \to \infty} W_1(\mu_t, \pi) = 0. \tag{17}$$

Ergodicity in this sense has received a great deal of attention since the seminal work of [27]. The contraction bound imposed in Assumption 4 of [27] also resembles drift conditions imposed in the present work. For the finite-dimensional setting of this paper, this assumption implies that, for a function $v \geq 1$, constants $k < \infty$, $\rho < 1$, and some time $t_0 > 0$,

$$(18) \qquad \mathsf{E}_x[v(X(t_0))] + \mathsf{E}_x[\|\mathcal{S}(t_0)\|v(X(t_0))] \leq kv(x)^\rho, \quad x \in \mathsf{X}.$$

This is similar (and in fact stronger) than Condition (DV3) (13).

The bound (18) and other assumptions imply the desired coupling result in [27], Theorem 3.4, which is also shown to imply the ergodic limit in our Theorem 2.1. And we should note that the weighted Sobolev norm $\|f\|_{v,1}$ used in Theorem 2.1 and throughout in this paper (cf. (20) in Section 2 below), also appears as $\|f\|_{V^r}$ in [27], page 21.

The most recent survey is contained in [43], which was completed during a revision of the present paper. The main results of [43] extend those in [6, 19] to obtain drift conditions that ensure the limit (17) holds at a geometric rate. The most crucial assumption is a geometric contraction on a "coupling set," and a drift condition similar to condition (V4) of [33, 41] to ensure this coupling set is visited frequently enough.

The convergence result (17) holds under the conditions of this paper when the function v is bounded (see Section 2.2.3). It is likely that the boundedness assumption can be relaxed, and the coupling result (17) simultaneously strengthened under the assumptions of this paper.

The present approach is complementary to the recent literature on ergodicity in the L_1-Wasserstein metric. The conclusions are in some sense stronger than those presented here, since these papers do not require ψ-irreducibility, and [6, 19, 27, 43] each contain explicit bounds on the rate of convergence.

However, these strong conclusions require strong assumptions. For example, the contraction bound (18) (or those imposed in [43]) are often not easily verified in applications, and are unlike any assumption imposed in the present work.

In conclusion, although the key results of this paper are related in spirit to much of the prior work mentioned above, there are no formal implications, in either direction, to existing results that we are aware of. Also, rather than the norm of the sensitivity process as in the definition of $\bar{\Lambda}_1$, we obtain bounds on the expectation of the sensitivity process, showing, for example, that for all C^1 functions $g : \mathsf{X} \to \mathbb{R}$ in an appropriate Banach space, the following can be uniformly bounded above:

$$\limsup_{t \to \infty} \frac{1}{t} \log |\mathsf{E}_x\{[\mathcal{S}(t)\nabla g(X(t))]_i\}|, \quad 1 \leq i \leq \ell.$$

The relationship between the limit theorems established in this paper and classical Lyapunov exponents is a topic of current research.

The remainder of the paper is organized as follows: Section 2 summarizes the main results for the Markov chain (1); these results are obtained under a Lyapunov condition slightly stronger than what is assumed in [34]. Section 3 contains proofs of the main results, leaving technical results to the Appendix. Section 3.2 contains results for a general Markov chain, not necessarily admitting the representation (1).

2. Assumptions and main results. The four assumptions (A1)–(A4) introduced in this section include the existence of a Lyapunov function $V : \mathsf{X} \to (0, \infty)$ satisfying the drift condition (DV3) of [33, 34]; see condition (A4) below. We denote $v = e^V$, which is used to define the norm $\| \cdot \|_v$ in (2).

The weighted Sobolev spaces $L_\infty^{v,k}$ considered in this paper are based on a function-space norm that involves the derivatives of a function $f : \mathsf{X} \to \mathbb{R}$. For each $k \geq 1$, denote

$$(19) \qquad \|f\|_{v,k} = \max_{|\alpha| \leq k} \|D^\alpha f\|_v,$$

where the maximum is over all multiindices $\alpha \in \mathbb{Z}_+^\ell$ with $\sum_i \alpha_i \leq k$, and D^α is the corresponding partial derivative. For $k = 1$, this is a maximum over $\ell + 1$ terms

$$\|f\|_{v,1} = \max\{\|f\|_v, \|\partial_1 f\|_v, \ldots, \|\partial_\ell f\|_v\},\tag{20}$$

where ∂_i denotes the first partial derivative with respect to x_i, $\partial/\partial x_i$. For $k = 0$, we let $L_\infty^{v,0}$ denote the space $\{f \in L_\infty^v : f$ is continuous$\}$ with norm $\|\cdot\|_v$. For each $k \geq 1$, we also define the spaces

$$L_\infty^{v,k} := \{f : \mathsf{X} \to \mathbb{R} : D^\alpha f \in L_\infty^{v,0} \text{ for all } |\alpha| \leq k\},$$

equipped with the norm defined in (19). This introduces two new restrictions on any function $f \in L_\infty^{v,k}$: The kth partial derivatives of f must exist and be absolutely bounded by a constant times v. In addition, f and these derivatives must be continuous. In the special case $v \equiv 1$, the space $L_\infty^{v,k}$ coincides with the usual Sobolev space $W^{k,1}$. Throughout most of the paper, we restrict attention to the cases $k = 0$ and $k = 1$. In Proposition 3.1, we show that the normed spaces $L_\infty^{v,0}$ and $L_\infty^{v,1}$ are complete and, therefore, are Banach spaces.

Consideration of the space $L_\infty^{v,1}$ requires the following assumptions on the evolution equations (1). Assumption (A1) ensures that the state at each time t is a continuously differentiable function of its initial condition $X(0) = x$, and justifies the representation of $\nabla^S f(X(t))$ in (10):

(A1)
(i) The process N does not depend upon the initial condition $X(0)$.
(ii) The function a is continuously differentiable in its first variable, with

$$\sup_{x,n} \|\nabla a(x, n)\| < \infty.$$

The notation $\|\cdot\|$ in (ii) can represent any matrix norm, and the jth column of the $\ell \times \ell$ matrix ∇a is equal to the gradient of a_j, so that

$$\left[\nabla a(x, n)\right]_{i,j} := \frac{\partial}{\partial x_i} a_j(x, n), \quad 1 \leq i, j \leq \ell.$$

2.1. *Irreducibility, densities and drift.* The general ergodic theory of Markov chains as developed in [41] involves two assumptions. The first is a generalization of irreducibility as defined for finite-state space Markov chains, and the second is a Foster–Lyapunov drift condition. The irreducibility conditions will hold under assumptions (A2) and (A3); the first is a density condition, and the second is a "reachability" assumption:

(A2)
For some $t_0 \geq 1$, the transition kernel admits a smooth density. That is, there is a continuously differentiable function p_{t_0} on $\mathsf{X} \times \mathsf{X}$ such that

$$P^{t_0}(x, A) = \int_A p_{t_0}(x, y)\, dy, \quad x \in \mathsf{X}, A \in \mathcal{B}.$$

Under (A2), there is in fact a density for *every* $t \geq t_0$, given by

$$p_t(x, y) = \int P^{t-t_0}(x, dz) p_{t_0}(z, y), \quad t \geq t_0, x, y \in \mathsf{X}.$$

The representation (1) implies the Feller property, that is, that the function $P^t f$ is continuous whenever f is continuous and bounded. Assumption (A2) implies the strong Feller property for P^t whenever $t \geq t_0$: The function $P^t f$ is continuous whenever f is measurable and bounded (cf. Lemma A.3 in the Appendix):

(A3)
There is a state $x_0 \in \mathsf{X}$ such that, for any $x \in \mathsf{X}$ and any open set O containing x_0, we have

$$P^t(x, O) > 0, \quad \text{for all } t \geq 0 \text{ sufficiently large.}$$

Under assumptions (A2) and (A3), the chain is ψ-irreducible and aperiodic, with $\psi(\cdot) := P^{t_0}(x_0, \cdot)$: For all $x \in \mathsf{X}$ and all $A \in \mathcal{B}$ such that $P^{t_0}(x_0, A) > 0$, we have

$$P^t(x, A) > 0, \quad \text{for all } t \geq 0 \text{ sufficiently large;}$$

see [41], Theorem 6.2.1.

Drift conditions are conveniently stated in terms of the *generator* for $\mathbf{X}$. In this discrete time setting, for measurable functions $f : \mathsf{X} \to \mathbb{R}$ the generator is defined as, $\mathcal{D}f := Pf - f$, that is,

$$\mathcal{D}f(x) := \mathsf{E}\big[f(X(t+1)) - f(X(t)) \mid X(t) = x\big], \quad x \in \mathsf{X},$$

for any f for which the expectation is defined for all x. Fleming's *nonlinear generator* [22–24, 34, 49] is defined via

$$(21) \qquad\qquad \mathcal{H}(F) := \log(Pe^F) - F,$$

for any measurable function F on X such that Pe^F exists.

We say [33, 34] that the *Lyapunov drift criterion* (DV3) *holds with respect to the Lyapunov function* $V : \mathsf{X} \to (0, \infty]$, if there exist a function $W : \mathsf{X} \to [1, \infty)$, a compact set $C \subset \mathsf{X}$, and constants $\delta > 0$, $b < \infty$, such that

$$(\text{DV3}) \qquad\qquad \mathcal{H}(V) \leq -\delta W + b\mathbb{I}_C.$$

In most of the subsequent results, the following strengthened version of (DV3) is assumed:

(A4) $\qquad$ Condition (DV3) holds with respect to functions V, W that are continuously differentiable and have compact sublevel sets. $\qquad\qquad\Big\}$

Recall that the sublevel sets of a function $F : \mathsf{X} \to \mathbb{R}_+$ are defined by

$$C_F(r) = \{x \in \mathsf{X} : F(x) \leq r\}, \quad r \geq 0.$$

2.2. *Results.* It is assumed throughout the remainder of this section that assumptions (A1)–(A4) hold. It follows that the Markov chain is v-uniformly ergodic, with $v = e^V$, so that (3) holds for a unique invariant probability measure π [34], Theorem 1.2. The first set of new results in this paper establish a similar conclusion in the Banach space $L_\infty^{v,1}$.

2.2.1. *Ergodicity in $L_\infty^{v,1}$.* The induced operator norm for a linear operator $\widehat{P} : L_\infty^v \to L_\infty^v$ is denoted

$$\|\!|\widehat{P}|\!\|_v := \sup\left\{ \frac{\|\widehat{P}f\|_v}{\|f\|_v} : f \in L_\infty^v, \|f\|_v \neq 0 \right\}.$$

On writing $\widetilde{P}^t = P^t - 1 \otimes \pi$ or, equivalently,

$$\widetilde{P}^t(x, A) = P^t(x, A) - \pi(A), \quad x \in \mathsf{X}, A \in \mathcal{B},$$

the bound (3) is expressed as

$$\|\!|\widetilde{P}^t|\!\|_v \leq b_0 \rho_0^t, \quad t \geq 0.$$

Similar notation is adopted for linear operators $\widehat{P} : L_\infty^{v,k} \to L_\infty^{v,k}$:

$$\|\!|\widehat{P}|\!\|_{v,k} := \sup\left\{ \frac{\|\widehat{P}f\|_{v,k}}{\|f\|_{v,k}} : f \in L_\infty^{v,k}, \|f\|_{v,k} \neq 0 \right\}.$$

Our main result here is the ergodicity of $\mathbf{X}$ in $L_\infty^{v,1}$. In fact, it is stated slightly more generally for all spaces $L_\infty^{v^\eta,1}$, defined as above with respect to the function $v^\eta = e^{\eta V}$, for any $\eta \in (0, 1]$.

THEOREM 2.1. *Under assumptions* (A1)–(A4), *for all* $\eta \in (0, 1]$, *there is* $b_0 < \infty$, $t_1 < \infty$ *and* $\varrho_0 < 1$ *such that*

$$(22) \qquad \qquad \left\| \tilde{P}^t \right\|_{v^\eta, 1} \leq b_0 \varrho_0^t, \quad t \geq t_1.$$

Consequently, for each $f \in L_\infty^{v^\eta, 1}$ *and* $t \geq t_1$,

$$\left| \mathsf{E}_x[f(X(t))] - \pi(f) \right| \leq b_0 \|f\|_{v^\eta, 1} \rho_0^t v^\eta(x), \quad and$$

$$(23)$$

$$\left| \frac{\partial}{\partial x_i} \mathsf{E}_x[f(X(t))] \right| \leq b_0 \|f\|_{v^\eta, 1} \rho_0^t v^\eta(x), \quad 1 \leq i \leq \ell.$$

The proof of the theorem, given in Section 3, is similar to the proof of v-uniform ergodicity in prior work [33]. Under assumptions (A1)–(A4), it is shown that the semigroup generated by $\tilde{P}$ has a discrete spectrum in $L_\infty^{v^\eta, 1}$, with spectral radius strictly bounded by unity.

In several of our subsequent results, we will need to restrict attention to the spaces $L_\infty^{v^\eta, 1}$ for η strictly less than 1. This is justified by the following proposition, stated here without proof; it is a simple consequence of the convexity of the operator $\mathcal{H}$.

PROPOSITION 2.2. *If the bound in condition* (DV3) *holds, then the same bound holds for any scaling by* $\eta \in (0, 1)$:

$$\mathcal{H}(\eta V) \leq -\delta \eta W + b \eta \mathbb{I}_C.$$

2.2.2. *Poisson's equation.* For $c \in L_\infty^v$, the sum (5) converges in L_∞^v, and h is a solution to the Poisson's equation (6): $h - Ph = c - \bar{c}$; cf. [25]. Under appropriate conditions, we show here that the gradient of h also exists.

Formal term-by-term differentiation of the definition of h in (5) yields

$$(24) \qquad \qquad \nabla h = \sum_{t=0}^\infty \nabla P^t c.$$

This will in fact follow from (22), once we could establish that $\left\| \tilde{P}^t \right\|_{v, 1}$ is finite for $t \geq t_1$. The proof of Theorem 2.3 is given in Section 3.3. Recall the definition of the semigroup $\{Q^t\}$ in (11).

THEOREM 2.3. *Suppose that* $c \in L_\infty^{v^\eta}$, *with* $\eta \in (0, 1]$. *Then the function* h *in* (5) *exists as an element of* $L_\infty^{v^\eta}$. *It is a solution to Poisson's equation* (6), *and it is unique among all functions in* $L_\infty^{v^\eta}$ *with* π-*mean equal to zero.*

If $\eta < 1$, *then we obtain the following additional conclusions*:

   (i) *If* $c \in L_\infty^{v^\eta, 0}$, *then* $h \in L_\infty^{v^\eta, 0}$.
   (ii) *If* $c \in L_\infty^{v^\eta, 1}$, *then* $h \in L_\infty^{v^\eta, 1}$, *with gradient given in* (24), *and also*

$$(25) \qquad \qquad \nabla h = \Omega \nabla c := \sum_{t=0}^\infty Q^t \nabla c.$$

Note that, in the theorem, the boundedness of Ω is only established on the space of functions of the form ∇f for some $f \in L_\infty^{v, 1}$. The first conclusion of the theorem (that h exists and uniquely solves Poisson's equation) has been established in [25, 41]; the remaining conclusions are new. The proof of (ii) is based on a representation of the gradient of the semigroup $\{P^t\}$: for $f \in L_\infty^{v^\eta, 1}$ with $\eta \in (0, 1)$, and $t \geq 1$:

$$(26) \qquad \qquad \nabla P^t f = Q^t \nabla f.$$

This and related results are given in Theorem 3.10.

Theorem 2.4, given next, states that exactly analogous results to those established in Theorem 2.3 for the solution h to Poisson's equation, can also be established for the function h_α, for any $\alpha \in (0, 1)$. Its proof is essentially the same as that of Theorem 2.3, and thus omitted.

THEOREM 2.4.   *Suppose that $c \in L_\infty^{v^\eta}$, with $\eta \in (0, 1]$. Then, for each $0 < \alpha < 1$, the function h_α in (7) exists as an element of $L_\infty^{v^\eta}$. It is a solution to the following fixed-point equation*:

(27) $$c + \alpha P h_\alpha - h_\alpha = 0,$$

and it is unique among all functions in $L_\infty^{v^\eta}$.
   If $\eta < 1$, then we obtain the following additional conclusions:

   (i) *If $c \in L_\infty^{v^\eta,0}$, then $h_\alpha \in L_\infty^{v^\eta,0}$.*
   (ii) *If $c \in L_\infty^{v^\eta,1}$, then $h_\alpha \in L_\infty^{v^\eta,1}$, with gradient given in (12)*:

(28) $$\nabla h_\alpha = \Omega_\alpha \nabla c := \sum_{t=0}^\infty \alpha^t Q^t \nabla c.$$

Once again, the fact that h_α is bounded and uniquely solves (27), follows from earlier work [40]. The proofs of parts (i) and (ii) are essentially identical to the proofs of the corresponding results in Theorem 2.3.

2.2.3. *Ergodicity and coupling.*   Let $L_\infty^{v,\mathrm{Lip}}$ denote the set of all measurable functions $f : \mathsf{X} \to \mathbb{R}$ for which the following norm is finite:

$$\|f\|_{v,\mathrm{Lip}} = \lim_{\varepsilon \to 0} \left(\sup_{x,y} \frac{1}{v(x)} \max\{|f(x)|, \varepsilon^{-1}|f(x) - f(y)|\} \right),$$

where the supremum is over all $x, y \in \mathbb{R}^\ell$ such that $\|y - x\| \le \varepsilon$. It is easily shown that $L_\infty^{v,\mathrm{Lip}}$ is a Banach space, and that $L_\infty^{v,1} \subset L_\infty^{v,\mathrm{Lip}}$, with consistent norm

$$\|f\|_{v,\mathrm{Lip}} = \|f\|_{v,1} \quad \text{for } f \in L_\infty^{v,1}.$$

A companion to Theorem 2.1 can be obtained, establishing a discrete spectrum for the semigroup on this larger Banach space. From this, a similar ergodic theorem is obtained.

THEOREM 2.5.   *Under assumptions (A1)–(A4), for all $\eta \in (0, 1]$, there is $b_0 < \infty$, $t_1 < \infty$ and $\varrho_0 < 1$ such that: for $f \in L_\infty^{v^\eta,\mathrm{Lip}}$*:

(29) $$\|P^t f - \pi(f)\|_{v^\eta,\mathrm{Lip}} \le b_0 \|f\|_{v^\eta,\mathrm{Lip}} \rho_0^t.$$

The major step in the proof is to demonstrate separability of P^t in the Banach space $L_\infty^{v^\eta,\mathrm{Lip}}$, for sufficiently large $t \ge 1$. The proof is similar to Theorem 2.1 and is omitted.
   A topic for future research is to relax the gradient bound in (A1), and instead assume a Lipschitz bound on the functions $\{a_j(\cdot, n)\}$ that is uniform in n. It is likely that the conclusions of Theorem 2.5 hold under this weaker assumption.

3. Spectral theory.

PROPOSITION 3.1.   *For any function $v : \mathsf{X} \to [1, \infty)$, the normed spaces L_∞^v, $L_\infty^{v,0}$ and $L_\infty^{v,1}$ are each Banach spaces.*

The proof of Proposition 3.1 is contained in Section 3.3. The following subsection concerns spectral theory for an operator acting on one of these spaces.

3.1. *Separability.*   A linear operator $T: L_\infty^{v,k} \to L_\infty^{v,k}$ has *finite rank* if there are functions $\{s_i\} \subset L_\infty^{v,k}$, measures $\{v_j\} \subset \mathcal{M}_1^v$, and constants $\{m_{ij}\}$ such that

$$(30) \qquad T = \sum_{i,j=1}^{N} m_{ij} s_i \otimes v_j,$$

where $[s \otimes v](x, dy) := s(x)v(dy)$, and $N < \infty$. We say that a linear operator $\widehat{P}: L_\infty^{v,k} \to L_\infty^{v,k}$ is *separable in* $L_\infty^{v,k}$ if, for each $\varepsilon > 0$, there is a finite-rank linear operator T such that $\||\widehat{P} - T\||_{v,k} \leq \varepsilon$.

The *spectrum* $\mathcal{S}(\widehat{P}) \subset \mathbb{C}$ of a linear operator $\widehat{P}: L_\infty^{v,k} \to L_\infty^{v,k}$ is the set of $z \in \mathbb{C}$ such that the inverse $[Iz - \widehat{P}]^{-1}$ does not exist as a bounded linear operator on $L_\infty^{v,k}$.

The *spectral radius* of the semigroup $\{\widehat{P}^n\}$ is denoted

$$(31) \qquad \xi_v(\widehat{P}) := \lim_{n \to \infty} \||\widehat{P}^n\||_v^{1/n}.$$

An element $z_0 \in \mathcal{S}(\widehat{P})$ is called a *pole of (finite) multiplicity n* if, for some $\varepsilon_1 > 0$:

(i) z_0 is isolated, $\{z \in \mathcal{S}(\widehat{P}) : |z - z_0| \leq \varepsilon_1\} = \{z_0\}$;
(ii) The associated projection operator $\mathcal{P}$ has finite rank, where

$$(32) \qquad \mathcal{P} := \frac{1}{2\pi i} \int_{\partial\{z:|z-z_0|\leq\varepsilon_1\}} [Iz - \widehat{P}]^{-1} dz.$$

For background, see the decomposition theorem in [46], Theorem 4.4, page 421.

The linear operator $\widehat{P}: L_\infty^{v,k} \to L_\infty^{v,k}$ has a *discrete spectrum in* $L_\infty^{v,k}$ if, for any compact set $C \subset \mathbb{C} \setminus \{0\}$, its spectrum $\mathcal{S}$ has the property that $\mathcal{S} \cap C$ is finite and contains only poles of finite multiplicity.

The above definition of separability is an extension of separability in L_∞^v for a linear operator $\widehat{P}: L_\infty^v \to L_\infty^v$ as defined in [34], which requires that we can find, for each $\varepsilon > 0$ a positive kernel of the form (30) in which $\{s_i\} \subset L_\infty^v$ and $\||\widehat{P} - T\||_{v,k} \leq \varepsilon$. Besides the consideration of the Banach spaces $L_\infty^{v,k}$, the definition here differs with [34] in two respects: First, positivity of T is not assumed since the kernel $\widehat{P}$ may not be positive. Second, in this prior work it was assumed that each function s_i and measure v_j had support on a compact set. This is not necessary here or in the technical results of [34].

The following theorem is an extension of Theorem 3.5 of [34] and it provides the fundamental connection between separability and ergodicity.

THEOREM 3.2 (Separability $\Rightarrow$ Discrete spectrum).   *If the linear operator $\widehat{P}: L_\infty^{v,k} \to L_\infty^{v,k}$ is bounded and $\widehat{P}^{t_1}: L_\infty^{v,k} \to L_\infty^{v,k}$ is separable in $L_\infty^{v,k}$ for some $t_1 \geq 1$ and $k \geq 0$, then $\widehat{P}$ has a discrete spectrum in $L_\infty^{v,k}$.*

As in the prior work [34], separability of the t-step transition kernel is established in two steps: First, it is shown that it can be approximated by its truncation to a compact set, and then the truncated kernel is shown to be separable.

A smooth truncation is the first step in the present paper: A transition density is approximated using Bernstein polynomials to establish that the truncation is separable in $L_\infty^{v,1}$. To simplify notation, consider first a C^1 function $\varphi: [0, 1]^N \to \mathbb{R}$, with $N \geq 2$. For an integer $m \geq 2$, the Bernstein approximation is given by

$$\varphi_m(z) = \sum_{j_1,\ldots,j_n=0}^{m} \varphi\left(\frac{j_1}{m}, \ldots, \frac{j_m}{m}\right) \prod_{i=1}^{n} \binom{m}{j_i} z^{j_i}(1 - z_i)^{m-j_i}, \quad z \in [0, 1]^N.$$

The proof of the following can be found in [32] for the special case $N = 2$; also see [7, 29] for related results, and [39] for a more recent discussion of the general case.

LEMMA 3.3. *The Bernstein polynomials provide the following uniform approximation for any C^1 function $\varphi\colon [0, 1]^N \to \mathbb{R}$:*

$$\lim_{m\to\infty} \sup_z \|\varphi(z) - \varphi_m(z)\|_2 = \lim_{m\to\infty} \sup_{z,i} \left\| \frac{\partial}{\partial z_i} \varphi(z) - \frac{\partial}{\partial z_i} \varphi_m(z) \right\|_2 = 0.$$

The truncated transition kernel will play the role of $\widehat{P}$ in the following lemma; its proof is given in Section A.3.

LEMMA 3.4. *Suppose $\widehat{P}$ has a density r with respect to probability measure μ: For each $x \in \mathsf{X}$ and $A \in \mathcal{B}$, $\widehat{P}(x, A) = \int_A r(x, y)\mu(dy)$. Suppose moreover that the density $r : \mathbb{R}^\ell \times \mathbb{R}^\ell \to \mathbb{R}^+$ is C^1 with compact support. Then $\widehat{P}$ is separable in $L_\infty^{v,1}$.*

We thus have a roadmap to prove the main results. First, we consider the case of Markov chains that may not have the representation (1).

3.2. *Separability implies ergodicity for general chains.* In this subsection only, we consider a general Markov chain evolving on $\mathsf{X} = \mathbb{R}^\ell$, not necessarily of the form (1). The goal is to generalize the results of Section 2, and also provide an overview of the proofs of the main results surveyed there.

Theorem 3.5 states that separability in $L_\infty^{v,1}$ implies ergodicity in this weighted Sobolev space. Sufficient conditions for separability are provided after the proof of the theorem.

THEOREM 3.5. *Suppose that the Markov chain $\mathbf{X}$ with transition kernel P satisfies the following conditions, for a continuous function $v\colon \mathsf{X} \to [1, \infty)$: It is v-uniformly ergodic, so that (3) holds for each $f \in L_\infty^v$. And, for some $t_1 \geq 1$, $\|\!|P^t\|\!|_{v,1} < \infty$ for $t \geq t_1$, and P^{t_1} is separable in $L_\infty^{v,1}$.*

Then the following conclusions hold:

(i) *The Markov chain is "ergodic in $L_\infty^{v,1}$": There is $b_0 < \infty$ and $\varrho_0 < 1$ such that*

$$\|\!|\widetilde{P}^t\|\!|_{v,1} \leq b_0 \varrho_0^t, \quad t \geq t_1.$$

(ii) *If, in addition, $\|\!|P^t\|\!|_{v,1} < \infty$ for $t \geq 1$, then, for any function $c \in L_\infty^{v,1}$, there is a solution to Poisson's equation $h \in L_\infty^{v,1}$, with gradient given in (24):*

$$\nabla h = \sum_{t=0}^{\infty} \nabla P^t c.$$

Part (ii) of the theorem is based on the following.

LEMMA 3.6. *For $\eta \in (0, 1]$ suppose that $\{g_n\} \subset L_\infty^{v^\eta, 1}$ satisfy $\sup_n \|g_n\|_{v^\eta, 1} < \infty$ and the following limits hold pointwise for continuous functions g and ζ:*

$$\lim_{n\to\infty} g_n(x) = g(x), \qquad \lim_{n\to\infty} \nabla g_n(x) = \zeta(x), \quad x \in \mathsf{X}.$$

Then $\nabla g = \zeta$ and $g \in L_\infty^{v^\eta, 1}$.

PROOF. For each n, i, α and x, we have

$$g_n(x + \alpha e^i) - g_n(x) = \int_0^\alpha \partial_i g_n(x + t e^i)\, dt,$$

where, as before, ∂_i denotes the partial derivative with respect to the ith coordinate. Letting $n \to \infty$ gives

$$g(x + \alpha e^i) - g(x) = \int_0^\alpha \zeta_i(x + te^i)\,dt.$$

Continuity of ζ implies that $\nabla g = \zeta$. $\square$

PROOF OF THEOREM 3.5. It follows from Theorem 3.2 that P has a discrete spectrum in $L_\infty^{v,1}$, and hence this is also true for P^{t_1}. Furthermore, it is straightforward to see that the spectrum of $\widetilde{P}^{t_1}$ in $L_\infty^{v,1}$ is a subset of its spectrum in L_∞^v: $S_{v,1}(\widetilde{P}^{t_1}) \subseteq S_v(\widetilde{P}^{t_1})$. Denote the respective spectral radii by $\xi_{v,1}(\widetilde{P}^{t_1})$ and $\xi_v(\widetilde{P}^{t_1})$ (recall the definition (31)). We obviously have $\xi_{v,1}(\widetilde{P}^{t_1}) \leq \xi_v(\widetilde{P}^{t_1})$. Also, under v-uniform ergodicity we have $\xi_v(\widetilde{P}^{t_1}) < 1$ [34], Theorem 2.4.

The conclusion $\xi_{v,1}(\widetilde{P}^{t_1}) < 1$ immediately gives (i) for the t_1-skeleton chain: There is $b_1 < \infty$ and $\varrho_1 < 1$ such that

$$\left\||\widetilde{P}^{t_1 k}\right\||_{v,1} \leq b_1 \varrho_1^k, \quad k \geq 0.$$

Under the assumption that $\||P^t\||_{v,1} < \infty$ for $t \geq t_1$, we obtain

$$\left\||\widetilde{P}^{t_1(k+1)+i}\right\||_{v,1} \leq \left(\max_{0 \leq j < t_1} \left\||\widetilde{P}^{t_1+j}\right\||_{v,1}\right) b_1 \varrho_1^k \quad \text{for each } k \geq 0 \text{ and } 0 \leq i < t_1,$$

which implies (i).

Write $\tilde{c} := c - \bar{c}$. The ergodicity result (i) is equivalent to the following bound for each $c \in L_\infty^{v,1}$:

$$\max\{|P^t \tilde{c}(x)|, |\partial_1 P^t c(x)|, \ldots, |\partial_\ell P^t c(x)|\} \leq b_0 \rho_0^t \|c\|_{v,1} v(x), \quad t \geq t_1.$$

On defining, for each $n \geq 1$,

$$h_n = \sum_{t=0}^n P^t \tilde{c},$$

it follows that $h_n \to h$ in $L_\infty^{v,1}$ at the same rate, under the assumption $\||P^t\||_{v,1} < \infty$ for $t \geq 1$. And applying Lemma 3.6,

$$\nabla h = \lim_{n \to \infty} \nabla h_n = \lim_{n \to \infty} \sum_{t=0}^n \nabla P^t \tilde{c} = \sum_{t=0}^\infty \nabla P^t \tilde{c},$$

which completes the proof. $\square$

The next set of results provide conditions under which the assumptions of Theorem 3.5 hold. It is convenient to strengthen (A2) to $t_0 = 1$:

There is a continuously differentiable function p on $\mathsf{X} \times \mathsf{X}$ such that

(A2′)
$$P(x, A) = \int_A p(x, y)\,dy, \quad x \in \mathsf{X}, A \in \mathcal{B}.$$

Assumption (A3) is maintained, which together with (A2′) again implies that X is ψ-irreducible and aperiodic.

The final assumption invokes (DV3) and a similar condition for ∇P. The partial derivatives of the density are denoted

$$p_i'(x, y) := \frac{\partial}{\partial x_i} p(x, y), \quad 1 \leq i \leq \ell.$$

(A4′)

    (i) The transition kernel P satisfies (DV3) with respect to continuous functions V, W, a compact set $C \subset \mathsf{X}$, and constants $\delta > 0$, $b < \infty$.

    (ii) For each $x \in \mathsf{X}$, $1 \le i \le \ell$,

$$\log \int \left| p_i'(x, y) \right| \exp\big(V(y) - V(x)\big) \, dy \le -\delta W(x) + b \mathbb{I}_C(x).$$

The drift condition (DV3) is used here and in [34, 36] to truncate the transition kernel onto a compact subset of the state space. Denote for $n \ge 1$,

$$R_n = \big\{ x \in \mathbb{R}^\ell : |x_i| \le n, \, 1 \le i \le \ell \big\}.$$

The function χ_n will denote a smooth approximation of the indicator function on this set. This is based on a function $\chi_n^1 : \mathbb{R} \to [0, 1]$ satisfying $\chi_n^1(r) = 1$ for $|r| \le n$ and $\chi_n^1(r) = 0$ for $|r| \ge n + 1$. It is assumed that χ_n^1 is also C^1, with

$$\left| \frac{d}{dr} \chi_n^1(r) \right| \le 2, \quad \text{for all } r \in \mathbb{R}.$$

The choice is not unique, but fixed throughout the paper. In ℓ dimensions, define

$$\chi_n(x) := \prod_{i=1}^\ell \chi_n^1(x_i), \quad x \in \mathbb{R}^\ell.$$

This function is also C^1, equal to 1 on R_n, 0 on R_{n+1}^c, and

$$\left| \frac{\partial}{\partial x_i} \chi_n(x) \right| \le 2, \quad 1 \le i \le \ell, x \in \mathbb{R}^\ell.$$

The proof of Lemma 3.7 can be found in the Appendix.

LEMMA 3.7. *Suppose that assumptions* (A2′) *and* (A4′) *hold. Then*:

(i) P^2 *can be approximated by its truncation in* $L_\infty^{v,1}$:

$$\lim_{n \to \infty} \lim_{m \to \infty} \big\| P^2 - I_{\chi_n} P^2 I_{\chi_m} \big\|_{v,1} = 0.$$

(ii) *For each* n, *the kernel* $I_{\chi_n} P^2 I_{\chi_n}$ *is separable in* $L_\infty^{v,1}$.

The assumptions of Theorem 3.5 hold with $t_1 = 2$.

THEOREM 3.8. *Suppose a Markov chain with transition kernel* P *satisfies assumptions* (A2′), (A3) *and* (A4′). *Then*:

(i) $P : L_\infty^{v,1} \to L_\infty^{v,1}$;
(ii) P^2 *is separable in* $L_\infty^{v,1}$.

PROOF. The fact that $P : L_\infty^{v,1} \to L_\infty^{v,1}$ is a bounded linear operator follows from assumption (A4′), and Lemma 3.7 implies that P^2 is separable in $L_\infty^{v,1}$. $\square$

3.3. *Proofs.* We now return to the Markov chain described by (1). The proposition that follows provides much of the ammunition required to obtain a version of Theorem 3.8 for this model; see Theorem 3.10 below.

The next result concerns separability in L_∞^v: Proposition 3.9(i) follows from Lemma B.5 of [34], and the proof of part (ii) is similar. Recall the definition (11) of the semigroup $\{Q^t\}$, which maps $\mathbb{R}^\ell$-valued functions to $\mathbb{R}^\ell$-valued functions; let $Q_{i,j}^t$ denote the (i, j)-th component of Q^t, for $1 \le i, j \le \ell, t \ge 0$.

PROPOSITION 3.9. *Suppose assumptions* (A1)–(A4) *hold. Then, for all* $t \geq t_1$:

  (i) P^t *is separable in* L_∞^v.
 (ii) $Q_{i,j}^t$ *is separable in* L_∞^v *for all* $1 \leq i, j \leq \ell$.

Proposition 3.9 is extended in this paper to the weighted Sobolev Banach spaces $L_\infty^{v^\eta,0}$ and $L_\infty^{v^\eta,1}$. The proof of Theorem 3.10 is contained in the Appendix.

THEOREM 3.10. *If assumptions* (A1)–(A4) *hold, then*:

(i) *For all* $t \geq t_1$ *and* $\eta \in (0, 1]$:

    (a) $P^t : L_\infty^{v^\eta,k} \to L_\infty^{v^\eta,k}$ *for* $k = 0, 1$.
    (b) $Q_{i,j}^t : L_\infty^{v^\eta,0} \to L_\infty^{v^\eta,0}$ *for all* $1 \leq i, j \leq \ell$.
    (c) $\nabla P^t f = Q^t \nabla f$, *if* $f \in L_\infty^{v^\eta,1}$.
    (d) P^t *is separable in* $L_\infty^{v^\eta,k}$ *for* $k = 0$ *and* $k = 1$.

(ii) *Results* (a)–(c) *hold for all* $t \geq 1$, *if* $\eta \in (0, 1)$.

PROOF OF THEOREM 2.1. Theorem 3.10(i) states that under assumptions (A1)–(A4), P^t is separable in $L_\infty^{v^\eta,1}$, for all $t \geq t_1$ and $\eta \in (0, 1]$. It also states that $P^t : L_\infty^{v^\eta,1} \to L_\infty^{v^\eta,1}$. Theorem 3.2 then implies that P^t has a discrete spectrum in $L_\infty^{v^\eta,1}$. Theorem 3.5 implies the desired conclusion:

$$\||\widetilde{P}^t\||_{v^\eta,1} \leq b_0 \varrho_0^t, \quad t \geq t_1. \qquad \square$$

PROOF OF THEOREM 2.3. As before, let $\tilde{c} = c - \overline{c}$. It is obvious that h in (5) is a solution to Poisson equation, and that its mean is zero. To establish uniqueness, suppose that $h \in L_\infty^v$ is any solution with mean zero. We iterate Poisson's equation to obtain

$$P^n h = h - \sum_{t=0}^{n-1} P^t \tilde{c}.$$

Since $h \in L_\infty^v$ with mean zero, we have $\|P^n h\|_v \to 0$ as $n \to \infty$, which establishes that h is equal to the infinite sum in (5). This establishes the first assertions of the theorem.

To prove (i), we fix $\eta \in (0, 1)$ and $c \in L_\infty^{v^\eta,0}$. We have as before that $\|P^t \tilde{c}\|_{v^\eta} \to 0$ as $t \to \infty$, and consequently $h \in L_\infty^{v^\eta}$. It remains to show that h is continuous.

Recall from Theorem 3.10 that $P^t : L_\infty^{v^\eta,0} \to L_\infty^{v^\eta,0}$ for each t. Since v is assumed to have compact sublevel sets, it follows that $\{P^t \tilde{c} : t \geq 0\}$ are continuous functions that converge to zero uniformly geometrically fast on compact subsets of X. This establishes continuity of h.

The proof of (ii) requires conclusion (ii)(c) of Theorem 3.10: For $c \in L_\infty^{v^\eta,1}$, $\eta \in (0, 1)$ and $t \geq 1$,

$$(Q^t \nabla c)_i = \partial_i P^t c.$$

Theorem 2.1 implies a geometric bound on the right-hand side: For $t \geq t_1$,

$$|\partial_i P^t c(x)| = |\partial_i \mathsf{E}_x[c(X(t))]| \leq b_0 \rho_0^t \|c\|_{v^\eta,1} v^\eta(x), \quad 1 \leq i \leq \ell.$$

Define for each $n \geq 1$

$$h_n = \sum_{t=0}^n P^t \tilde{c}.$$

Since $P^t : L_\infty^{v^\eta,1} \to L_\infty^{v^\eta,1}$ for each $t \geq 1$, for any finite n, we have $h_n \in L_\infty^{v^\eta,1}$. Moreover, since $P^t \tilde{c} \to 0$ in $L_\infty^{v^\eta,1}$ as $t \to \infty$ at a geometric rate, it follows that as $n \to \infty$, $h_n \to h$ in $L_\infty^{v^\eta,1}$ at the same rate.

In particular,

$$\nabla h = \lim_{n \to \infty} \nabla h_n = \lim_{n \to \infty} \sum_{t=0}^{n} \nabla P^t \tilde{c} = \sum_{t=0}^{\infty} Q^t \nabla c = \Omega \nabla c,$$

as claimed. □

APPENDIX

A.1. Operator bounds. We begin with sufficient conditions for the identity (26). We first require the following corollary to Lemma 3.6.

LEMMA A.1. *Suppose that for some $t \geq 1$ and $\eta \in (0, 1]$,*

(33) $$P^t : L_\infty^{v^\eta,0} \to L_\infty^{v^\eta,0}, \qquad Q_{i,j}^t : L_\infty^{v^\eta,0} \to L_\infty^{v^\eta,0}, \quad 1 \leq i, j \leq \ell.$$

Then $P^t : L_\infty^{v^\eta,1} \to L_\infty^{v^\eta,1}$ and (26) holds on $L_\infty^{v^\eta,1}$:

$$\nabla P^t f = Q^t \nabla f, \quad f \in L_\infty^{v^\eta,1}.$$

PROOF. For any function $f \in L_\infty^{v^\eta,1}$ and $n \geq 1$, let $f_n = \mathcal{X}_n f$. The function f_n and its partial derivatives are continuous, and $\sup_n \|f_n\|_{v^\eta,1} < \infty$. We have $\lim_{n \to \infty} \nabla f_n = \nabla f$, where the limit is continuous by assumption.

We apply Lemma 3.6 with $g_n = P^t f_n$. To verify the conditions of the lemma, first observe that $\{g_n\}$ converges to $g = P^t f$ by dominated convergence. The limiting function g is continuous by (33). From (11), it follows that $\nabla g_n = Q^t \nabla f_n$ and, since each $Q_{i,j}^t$ is a bounded linear operator, it follows that $\sup_n \|g_n\|_{v^\eta,1} < \infty$. The final requirement of the lemma is convergence of the gradients. This follows from a second application of dominated convergence:

$$\zeta(x) := \lim_{n \to \infty} \nabla g_n(x) = \lim_{n \to \infty} \int Q^t(x, dy) \nabla f_n(y) = Q^t \nabla f(x).$$

Lemma 3.6 then implies the desired conclusion that $\nabla P^t f = Q^t \nabla f$. This identity combined with (33) then implies that $P^t : L_\infty^{v^\eta,1} \to L_\infty^{v^\eta,1}$. □

A second application of Lemma 3.6 is in the proof of Proposition 3.1.

PROOF OF PROPOSITION 3.1. The proof only requires that each of these function spaces is complete. This is an elementary exercise in the case of L_∞^v, and then $L_\infty^{v,0}$. Completeness of $L_\infty^{v,1}$ is established here.

Suppose that $\{f_n\} \subset L_\infty^{v,1}$ is a Cauchy sequence. Since $L_\infty^{v,0}$ is a Banach space, it immediately follows that there are functions $\{f, \zeta_1, \ldots, \zeta_\ell\} \subset L_\infty^{v,0}$ such that

$$\lim_{n \to \infty} \|f_n - f\|_v = \lim_{n \to \infty} \|\partial_i f_n - \zeta_i\|_v = 0, \quad 1 \leq i \leq \ell.$$

Consequently, the assumptions of Lemma 3.6 hold, with $\zeta = \nabla f$ continuous. Moreover, these limits imply that convergence of $\{f_n\}$ to f holds in $L_\infty^{v,1}$, as required for completeness: $\lim_{n \to \infty} \|f_n - f\|_{v,1} = 0$. □

A.2. Truncations. Several truncation bounds are obtained here. The following notation will be useful: For any operator Z on L_∞^v, $L_\infty^{v,0}$ or $L_\infty^{v,1}$, we write $Z_n \xrightarrow{v} Z$, if

$$\lim_{n\to\infty} \||Z_n - Z\||_v = 0.$$

The elementary observation stated below without proof, is used to avoid establishing one-sided truncation bounds; recall the definition of the functions $\{\chi_n\}$ in Section 3.2.

LEMMA A.2. *Suppose that Z is a bounded linear operator on a Banach space of functions on* X, *with induced operator norm $\||\cdot\||$. If Z can be approximated by its truncation on both sides,*

$$\lim_{n\to\infty} \||Z - I_{\chi_n} Z I_{\chi_n}\|| = 0,$$

then Z can be approximated by its truncation on either side:

$$\lim_{n\to\infty} \||Z - Z I_{\chi_n}\|| = \lim_{n\to\infty} \||Z - I_{\chi_n} Z\|| = 0.$$

PROOF OF LEMMA 3.7.   It is only necessary to prove (i), since the implication (i) $\Rightarrow$ (ii) follows from Lemma 3.4.

Assumption (A2$'$) along with part (i) of (A4$'$) implies that the transition kernel can be approximated by its left truncation: In L_∞^v, we have $I_{\chi_n} P \xrightarrow{v} P$ (see [34], Lemma B.4), and hence

$$(34) \qquad\qquad (I_{\chi_n} P)^2 \xrightarrow{v} P^2.$$

Furthermore, assumptions (A2$'$) and (A4$'$) imply a bound of the form

$$\left| I_{\chi_n} P I_{\chi_n}(x, A) \right| \le \beta_n^0(A), \quad 1 \le i \le \ell, x \in \mathsf{X}, A \in \mathcal{B},$$

where β_n^0 is a positive measure with compact support, and hence $\beta_n^0(v) < \infty$. Therefore, for all $x \in \mathsf{X}$ and $A \subset \mathcal{B}$,

$$\left| (I_{\chi_n} P)^2(x, A) \right| = \left| \int_{R_{n+1}} \left\{ \chi_n(x) P(x, dy)\, \chi_n(y) P(y, A) \right\} \right|$$

$$\le \int_{R_{n+1}} \beta_n^0(dy) P(y, A) := \beta_n(A),$$

with $\beta_n(v) < \infty$ since both $\||P\||_v$ and $\beta_n^0(v)$ are finite. Therefore, for any $f \in L_\infty^v$,

$$\left\| (I_{\chi_n} P)^2 I_{\chi_m} f - P^2 f \right\|_v \le \left\| (I_{\chi_n} P)^2 (1 - \chi_m) f \right\|_v + \left\| (I_{\chi_n} P)^2 f - P^2 f \right\|_v$$

$$\le \beta_n(v I_{R_m^c}) \| f \|_v + \left\| (I_{\chi_n} P)^2 f - P^2 f \right\|_v,$$

and applying (34),

$$\lim_{m,n\to\infty} \left\|\left| (I_{\chi_n} P)^2 I_{\chi_m} - P^2 \right\|\right|_v = 0.$$

To complete the proof of (i), it remains to be shown that, for each $1 \le i \le \ell$,

$$(35) \qquad\qquad \lim_{m,n\to\infty} \left\|\left| \partial_i (I_{\chi_n} P)^2 I_{\chi_m} - \partial_i P^2 \right\|\right|_v = 0,$$

where, again, ∂_i is shorthand for $\partial/\partial x_i$. The proof follows exactly the same steps as before: Assumption (A2$'$) and part (ii) of (A4$'$) imply that P_i' can be truncated on the left: $I_{\chi_n} P_i' \xrightarrow{v} P_i'$ for each i; that is,

$$\lim_{n\to\infty} \||\partial_i P - I_{\chi_n} \partial_i P\||_v = 0, \quad 1 \le i \le \ell.$$

From this, and the prior conclusion $I_{\chi_n} P \xrightarrow{v} P$, we obtain

(36) $$\lim_{n\to\infty} \left\| \partial_i (I_{\chi_n} P)^2 f - \partial_i P^2 f \right\|_v = 0.$$

Furthermore, the two assumptions imply a bound of the form

$$\left| \partial_i I_{\chi_n} P I_{\chi_n}(x, A) \right| \le \gamma_n^0(A), \quad 1 \le i \le \ell, x \in \mathsf{X}, A \in \mathcal{B},$$

where γ_n^0 is a positive measure with compact support, and hence $\gamma_n^0(v) < \infty$. Therefore, for all $x \in \mathsf{X}$, $A \subset \mathcal{B}$ and $1 \le i \le \ell$,

$$\left| \partial_i (I_{\chi_n} P)^2(x, A) \right| = \left| \int_{R_{n+1}} \frac{\partial}{\partial x_i} \{ \chi_n(x) P(x, dy) \chi_n(y) P(y, A) \} \right|$$

$$\le \int_{R_{n+1}} \gamma_n^0(dy) P(y, A) := \gamma_n(A).$$

It follows that for all $f \in L_\infty^v$, and $1 \le i \le \ell$,

$$\left\| \partial_i (I_{\chi_n} P)^2 I_{\chi_m} f - \partial_i P^2 f \right\|_v$$

$$\le \left\| \partial_i (I_{\chi_n} P)^2 (1 - \chi_m) f \right\|_v + \left\| \partial_i (I_{\chi_n} P)^2 f - \partial_i P^2 f \right\|_v$$

$$\le \gamma_n(v I_{R_m^c}) \| f \|_v + \left\| \partial_i (I_{\chi_n} P)^2 f - \partial_i P^2 f \right\|_v.$$

Combining this with (36) implies that (35) holds, and this completes the proof of part (i), as required. $\square$

The next results concern the nonlinear state space model. Lemma A.3 follows directly from the assumptions. Recall the discussion of the (strong) Feller property in Section 2.1. As before, $Q_{i,j}^t$ denotes the (i, j)-th component of Q^t, for $1 \le i, j \le \ell, t \ge 0$.

LEMMA A.3. *Suppose that assumptions* (A1)–(A4) *hold, and let Z_t denote any one of the kernels P^t or $Q_{i,j}^t$ with $t \ge 1$ and $1 \le i, j \le \ell$.*

(i) *The* Feller property *holds for Z_t, for $t \ge 1$ and the* strong Feller property *holds for Z_t, when $t \ge t_0$. Moreover, the following stronger properties hold:*

$$Z_t : L_\infty^{v^\eta, 0} \to L_\infty^{v^\eta, 0}, \quad t \ge 1,$$

$$Z_t : L_\infty^{v^\eta} \to L_\infty^{v^\eta, 0}, \quad t \ge t_0.$$

(ii) *For each $n \ge 1$ and $\eta \in (0, 1]$,*

$$Z_t I_{\chi_n} : L_\infty^{v^\eta, 0} \to L_\infty^{v^\eta, 0}, \quad t \ge 1,$$

$$Z_t I_{\chi_n} : L_\infty^{v^\eta} \to L_\infty^{v^\eta, 0}, \quad t \ge t_0.$$

(iii) *Suppose that for some $\eta \in (0, 1]$, $t \ge 1$ and every $g \in L_\infty^{v^\eta}$,*

$$\lim_{n\to\infty} Z_t I_{\chi_n} g = Z_t g,$$

where the convergence is uniform on compact subsets of X. Then

$$Z_t : L_\infty^{v^\eta, 0} \to L_\infty^{v^\eta, 0}, \quad and$$

$$Z_t : L_\infty^{v^\eta} \to L_\infty^{v^\eta, 0}, \quad provided \ t \ge t_0.$$

The proof of the next result is also elementary.

LEMMA A.4. *Suppose the conclusions of Proposition* 3.9 *are true, that is, for each $t \geq t_1$:*

(a) *P^t is separable in L_∞^v;*
(b) *$Q_{i,j}^t$ is separable in L_∞^v for any pair $1 \leq i, j \leq \ell$.*

Then the kernels P^t and Q^t can be approximated by their truncations:

(i) *$\lim_{n \to \infty} \| P^t - I_{\chi_n} P^t I_{\chi_n} \|_{v,1} = 0$;*
(ii) *$\lim_{n \to \infty} \| Q_{i,j}^t - I_{\chi_n} Q_{i,j}^t I_{\chi_n} \|_v = 0$ for any pair $1 \leq i, j \leq \ell$.*

PROOF. The fact that the kernels can be approximated in L_∞^v by their truncations for each $t \geq t_1$ follows directly from the assumption that they are separable: We have

$$
\begin{aligned}
I_{\chi_n} P^t I_{\chi_n} &\xrightarrow{\;v\;} P^t \\
I_{\chi_n} Q_{i,j}^t I_{\chi_n} &\xrightarrow{\;v\;} Q_{i,j}^t, \quad 1 \leq i, j \leq \ell.
\end{aligned}
$$

(37)

In particular, part (ii) is immediate.

To complete the proof of (i), it remains to be shown that there is a vanishing sequence $\{\varepsilon(n)\}$ such that, for any function $f \in L_\infty^{v,1}$,

$$
\left\| \partial_i \{ P^t f \} - \partial_i \{ I_{\chi_n} P^t I_{\chi_n} f \} \right\|_v \leq \varepsilon(n) \| f \|_{v,1}.
$$

Lemma A.3 along with (37) implies that the assumptions of Lemma A.1 are satisfied (with $t \geq t_1$):

$$
P^t : L_\infty^{v^\eta,0} \to L_\infty^{v^\eta,0}, \qquad Q_{i,j}^t : L_\infty^{v^\eta,0} \to L_\infty^{v^\eta,0}, \quad 1 \leq i, j \leq \ell.
$$

Now, applying the product rule gives

$$
\partial_i \{ I_{\chi_n} P^t I_{\chi_n} f \} = \{ \partial_i \chi_n \} \{ P^t I_{\chi_n} f \} + I_{\chi_n} (Q^t \nabla (f \chi_n))_i,
$$

with the second term justified applying Lemma A.1.

The first term can be bounded

$$
\left\| \{ \partial_i \chi_n \} \{ P^t I_{\chi_n} f \} \right\|_v \leq \varepsilon_1(n) \| f \|_v \leq \varepsilon_1(n) \| f \|_{v,1},
$$

where

$$
\varepsilon_1(n) = \Big(\max_i \| \partial_i \chi_n \|_\infty \Big) \| I_{R_n^c} P^t \|_v \leq 2 \| I_{R_n^c} P^t \|_v.
$$

The premultiplication by $I_{R_n^c}$ is justified since $\nabla \chi_n = 0$ on R_n. Equation (37) along with Lemma A.2 implies that $\| I_{R_n^c} P^t \|_v \to 0$ as $n \to \infty$, and hence $\lim_{n \to \infty} \varepsilon_1(n) = 0$. Therefore,

$$
\begin{aligned}
\left\| \partial_i \{ P^t f \} - \partial_i \{ I_{\chi_n} P^t I_{\chi_n} f \} \right\|_v &\leq \varepsilon_1(n) \| f \|_{v,1} \\
&\quad + \left\| (Q^t \nabla f)_i - I_{\chi_n} (Q^t \nabla (f \chi_n))_i \right\|_v.
\end{aligned}
$$

Once more applying equation (37), it is straightforward to see that there is a vanishing sequence $\{\varepsilon_2(n)\}$ such that

$$
\left\| (Q^t \nabla f)_i - I_{\chi_n} (Q^t \nabla (f \chi_n))_i \right\|_v \leq \varepsilon_2(n) \| f \|_{v,1}, \quad n \geq 1.
$$

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

The justification of the representation (25) for ∇h requires a different set of truncation arguments.

LEMMA A.5. *Let $\eta \in (0, 1)$. For each $t \geq 1$, the kernels P^t and Q^t can be approximated in $L_\infty^{v^\eta, 1}$ by their truncations on the right*:

(i) $P^t I_{\chi_n} \longrightarrow_{v^\eta} P^t$;
(ii) $Q_{i,j}^t I_{\chi_n} \longrightarrow_{v^\eta} Q_{i,j}^t$, *for any pair $1 \leq i, j \leq \ell$.*

PROOF. Let $\eta \in (0, 1)$, take $f \in L_\infty^{v^\eta, 1}$, and let $f_n := I_{\chi_n} f$. Then, for all $t \geq 1$,

$$\left| P^t f(x) - P^t f_n(x) \right| \leq \left| \int_{R_n^c} P^t(x, dy) f_n(y) \right|$$

$$\leq \|f\|_{v^\eta} \int_{R_n^c} P^t(x, dy) v(y) \left(\frac{v^\eta(y)}{v(y)} \right)$$

$$\leq \|f\|_{v^\eta} \left[\sup_{y' \in R_n^c} v^{\eta-1}(y') \right] \int_{R_n^c} P^t(x, dy) v(y)$$

$$\leq \|f\|_{v^\eta} \varepsilon(n),$$

where $\varepsilon(n) \to 0$ as $n \to \infty$. The last step follows from the fact that $\|\|P^t\|\|_v < \infty$ under (DV3), and $v(x) \to \infty$ as $\|x\|_2 \to \infty$ because v has compact sublevel sets under assumption (A4).

Under assumption (A1), and using the same arguments as above, we have

$$\lim_{n \to \infty} \|Q_{i,j}^t f - Q_{i,j}^t f_n\|_{v^\eta} = 0, \quad \text{for all } 1 \leq i, j \leq \ell. \qquad \square$$

The following strengthening of the Feller property is another step in the proof of Theorem 3.10.

PROPOSITION A.6. *Under assumptions (A1)–(A4)*:

(i) *For all $t \geq t_1$, and $\eta = 1$,*

(38) $$P^t: L_\infty^{v^\eta, 0} \to L_\infty^{v^\eta, 0}, \qquad Q_{i,j}^t: L_\infty^{v^\eta, 0} \to L_\infty^{v^\eta, 0}, \quad 1 \leq i, j \leq \ell.$$

(ii) *The conclusions (38) hold for all $t \geq 1$ when $\eta \in (0, 1)$.*

PROOF. Lemma A.4(i) along with Lemma A.2 implies that for any function $g \in L_\infty^{v, 0}$ and all $t \geq t_1$ we have

$$\lim_{n \to \infty} P^t I_{\chi_n} g = P^t g,$$

where the convergence is uniform on compact subsets of X. It then follows from Lemma A.3 that $P^t: L_\infty^{v^\eta, 0} \to L_\infty^{v^\eta, 0}$ for any $\eta \in (0, 1]$.

Similarly, using Lemma A.4(ii),

$$\lim_{n \to \infty} Q_{i,j}^t I_{\chi_n} g = Q_{i,j}^t g,$$

for any $g \in L_\infty^{v, 0}$. This again implies that $Q_{i,j}^t: L_\infty^{v^\eta, 0} \to L_\infty^{v^\eta, 0}$, $\eta \in (0, 1]$, from Lemma A.3. This completes the proof of part (i) of the proposition.

The proof of part (ii) follows exactly in the same manner, using Lemma A.5 (instead of Lemma A.4) along with Lemma A.3. $\square$

PROOF OF THEOREM 3.10. First, consider part (i). Proposition A.6 establishes (b), and part of (a): For all $t \geq t_1$ and $\eta \in (0, 1]$,

(39) $$P^t: L_\infty^{v^\eta, 0} \to L_\infty^{v^\eta, 0}, \quad \text{and} \quad Q_{i,j}^t: L_\infty^{v^\eta, 0} \to L_\infty^{v^\eta, 0}, \quad 1 \leq i, j \leq \ell.$$

Applying Lemma A.1, we obtain the remainder of (a), and also (c).

Assumption (A2) implies that P^{t_1} has a density which is C^1. Furthermore, from Lemma A.4, we conclude that under assumptions (A1)–(A4), P^{t_1} can be approximated by its truncation $I_{\chi_n} P^{t_1} I_{\chi_n}$ in $L_\infty^{v,1}$. Lemma 3.4 therefore completes the proof of (d).

Next, consider part (ii). Proposition A.6 again establishes (b). Part (ii) of Proposition A.6 states that (39) holds for each $t \geq 1$ and $\eta \in (0, 1)$. Consequently, (a) and (c) follow as before by applying Lemma A.1. □

A.3. Separability and Bernstein polynomials.

PROOF OF LEMMA 3.4.   Let $r_v(x, y) := r(x, y)v(y)$. For any function $g \in L_\infty^v$, we have

$$\widehat{P}g(x) = \int r_v(x, y)g(y)v^{-1}(y)\mu(dy).$$

Since v is assumed to be C^1, r_v is also C^1 with compact support.

Choose $n \geq 1$ such that $r_v(x, y) = 0$ on $(R_n \times R_n)^c$. Therefore, for any given $\varepsilon > 0$ there exists a Bernstein's polynomial $r_v^{\varepsilon_0}$ such that, for all $(x, y) \in R_{n+1} \times R_{n+1}$,

$$|r_v(x, y) - r_v^{\varepsilon_0}(x, y)| \leq \varepsilon, \quad \text{and}$$

$$\left|\frac{\partial}{\partial x_i} r_v(x, y) - \frac{\partial}{\partial x_i} r_v^{\varepsilon_0}(x, y)\right| \leq \varepsilon, \quad 1 \leq i \leq \ell.$$

The approximating polynomial can be expressed in the suggestive form

$$r_v^{\varepsilon_0}(x, y) = \sum_{i=1}^{N} s_i^0(x) r_i^0(y).$$

Truncating the approximation smoothly as $r_v^\varepsilon(x, y) = \chi_n(x)\chi_n(y)r_v^{\varepsilon_0}(x, y)$, we obtain a function supported on $R_{n+1} \times R_{n+1}$,

$$r_v^\varepsilon(x, y) = \sum_{i=1}^{N} s_i(x) r_i(y),$$

with $s_i = \chi_n s_i^0$ and $r_i = \chi_n r_i^0$. It is then straightforward that

$$\sup_{x,y}|r_v(x, y) - r_v^\varepsilon(x, y)| \leq \varepsilon, \quad \text{and}$$

$$\sup_{x,y}\left|\frac{\partial}{\partial x_i} r_v(x, y) - \frac{\partial}{\partial x_i} r_v^\varepsilon(x, y)\right| \leq \varepsilon, \quad 1 \leq i \leq \ell,$$

where the suprema are over $(x, y) \in \mathsf{X} \times \mathsf{X}$.

The following approximating kernel has finite rank:

$$T_\varepsilon(x, dy) = r_v^\varepsilon(x, y)v^{-1}(y)\mu(dy).$$

We also have

$$|\widehat{P}g(x) - T_\varepsilon g(x)| \leq \int |r_v(x, y) - r_v^\varepsilon(x, y)|\left|\frac{g(y)}{v(y)}\right|\mu(dy)$$

$$\leq \sup_{x,y}|r_v(x, y) - r_v^\varepsilon(x, y)|\sup_z\left|\frac{g(z)}{v(z)}\right|$$

$$\leq \varepsilon\|g\|_v,$$

and

$$\left| \frac{\partial}{\partial x_i} \widehat{P} g(x) - \frac{\partial}{\partial x_i} T_\varepsilon g(x) \right|$$

$$= \left| \frac{\partial}{\partial x_i} \int \Delta_r^\varepsilon(x, y) \frac{g(y)}{v(y)} \mu(dy) \right|$$

$$= \left| \lim_{\delta \to 0} \frac{1}{\delta} \int \left(\Delta_r^\varepsilon(x + \delta e^i, y) - \Delta_r^\varepsilon(x, y) \right) \frac{g(y)}{v(y)} \mu(dy) \right|,$$

where $\Delta_r^\varepsilon = r_v - r_v^\varepsilon$, and e^i denotes the ith basis vector in $\mathbb{R}^\ell$.

Since, both r_v and r_v^ε are C^1, the mean value theorem gives

$$\frac{1}{\delta} \left| \Delta_r^\varepsilon(x + \delta e^i, y) - \Delta_r^\varepsilon(x, y) \right| = \left| \frac{\partial}{\partial x_i} \Delta_r^\varepsilon(\overline{x}_i, y) \right|,$$

for some $\overline{x}_i \in (x, x + \delta e^i)$. The right-hand side is uniformly bounded over all $\delta \in (0, 1]$, and thus, by dominated convergence,

$$\left| \frac{\partial}{\partial x_i} \widehat{P} g(x) - \frac{\partial}{\partial x_i} T_\varepsilon g(x) \right|$$

$$\leq \int \limsup_{\delta \to 0} \frac{1}{\delta} \left| \Delta_r^\varepsilon(x + \delta e^i, y) - \Delta_r^\varepsilon(x, y) \right| \left| \frac{g(y)}{v(y)} \right| \mu(dy)$$

$$\leq \sup_{x, y} \left| \frac{\partial}{\partial x_i} r_v(x, y) - \frac{\partial}{\partial x_i} r_v^\varepsilon(x, y) \right| \| g \|_v$$

$$\leq \varepsilon \| g \|_v.$$

This completes the proof of separability of $\widehat{P}$ in $L_\infty^{v,1}$. $\square$

REFERENCES

[1] ARNOLD, L. and CRAUEL, H. (1991). Random dynamical systems. In *Lyapunov Exponents* (*Oberwolfach*, 1990). *Lecture Notes in Math.* **1486** 1–22. Springer, Berlin. MR1178943 https://doi.org/10.1007/BFb0086654

[2] ASMUSSEN, S. and GLYNN, P. W. (2007). *Stochastic Simulation*: *Algorithms and Analysis. Stochastic Modelling and Applied Probability* **57**. Springer, New York. MR2331321

[3] ATAR, R. and ZEITOUNI, O. (1997). Lyapunov exponents for finite state nonlinear filtering. *SIAM J. Control Optim.* **35** 36–55. MR1430282 https://doi.org/10.1137/S0363012994272046

[4] BALAJI, S. and MEYN, S. P. (2000). Multiplicative ergodicity and large deviations for an irreducible Markov chain. *Stochastic Process. Appl.* **90** 123–144. MR1787128 https://doi.org/10.1016/S0304-4149(00)00032-6

[5] BAXENDALE, P. H. (1989). Lyapunov exponents and relative entropy for a stochastic flow of diffeomorphisms. *Probab. Theory Related Fields* **81** 521–554. MR0995809 https://doi.org/10.1007/BF00367301

[6] BUTKOVSKY, O. (2014). Subgeometric rates of convergence of Markov processes in the Wasserstein metric. *Ann. Appl. Probab.* **24** 526–552. MR3178490 https://doi.org/10.1214/13-AAP922

[7] BUTZER, P. L. (1953). On two-dimensional Bernstein polynomials. *Canad. J. Math.* **5** 107–113. MR0052573 https://doi.org/10.4153/cjm-1953-014-2

[8] CAO, X.-R. and CHEN, H.-F. (1997). Perturbation realization, potentials, and sensitivity analysis of Markov processes. *IEEE Trans. Automat. Control* **42** 1382–1393. MR1472856 https://doi.org/10.1109/9.633827

[9] CATTIAUX, P., CHAFAÏ, D. and GUILLIN, A. (2012). Central limit theorems for additive functionals of ergodic Markov diffusions processes. *ALEA Lat. Am. J. Probab. Math. Stat.* **9** 337–382. MR3069369

[10] CATTIAUX, P. and GUILLIN, A. (2017). Hitting times, functional inequalities, Lyapunov conditions and uniform ergodicity. *J. Funct. Anal.* **272** 2361–2391. MR3603301 https://doi.org/10.1016/j.jfa.2016.10.003

[11] CATTIAUX, P., GUILLIN, A., WANG, F.-Y. and WU, L. (2009). Lyapunov conditions for super Poincaré inequalities. *J. Funct. Anal.* **256** 1821–1841. MR2498560 https://doi.org/10.1016/j.jfa.2009.01.003

[12] DEVRAJ, A., KONTOYIANNIS, I. and MEYN, S. (2018). Differential temporal difference learning. arXiv preprint. Available at arXiv:1812.11137.

[13] DEVRAJ, A. M. and MEYN, S. P. (2016). Differential TD learning for value function approximation. In *55th Conference on Decision and Control* 6347–6354.

[14] DJELLOUT, H., GUILLIN, A. and WU, L. (2004). Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Ann. Probab.* **32** 2702–2732. MR2078555 https://doi.org/10.1214/009117904000000531

[15] DONSKER, M. D. and VARADHAN, S. R. S. (1975). Asymptotic evaluation of certain Markov process expectations for large time. I. II. *Comm. Pure Appl. Math.* **28** 1–47; ibid. **28**, 279–301. MR0386024 https://doi.org/10.1002/cpa.3160280102

[16] DONSKER, M. D. and VARADHAN, S. R. S. (1976). Asymptotic evaluation of certain Markov process expectations for large time. III. *Comm. Pure Appl. Math.* **29** 389–461. MR0428471 https://doi.org/10.1002/cpa.3160290405

[17] DONSKER, M. D. and VARADHAN, S. R. S. (1983). Asymptotic evaluation of certain Markov process expectations for large time. IV. *Comm. Pure Appl. Math.* **36** 183–212. MR0690656 https://doi.org/10.1002/cpa.3160360204

[18] DOWN, D., MEYN, S. P. and TWEEDIE, R. L. (1995). Exponential and uniform ergodicity of Markov processes. *Ann. Probab.* **23** 1671–1691. MR1379163

[19] DURMUS, A. and MOULINES, É. (2017). Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *Ann. Appl. Probab.* **27** 1551–1587. MR3678479 https://doi.org/10.1214/16-AAP1238

[20] EBERLE, A., GUILLIN, A. and ZIMMER, R. (2019). Quantitative Harris-type theorems for diffusions and McKean–Vlasov processes. *Trans. Amer. Math. Soc.* **371** 7135–7173. MR3939573 https://doi.org/10.1090/tran/7576

[21] EDWARDS, D. A. (2011). On the Kantorovich–Rubinstein theorem. *Expo. Math.* **29** 387–398. MR2861765 https://doi.org/10.1016/j.exmath.2011.06.005

[22] FENG, J. (1999). Martingale problems for large deviations of Markov processes. *Stochastic Process. Appl.* **81** 165–216. MR1694569 https://doi.org/10.1016/S0304-4149(98)00104-5

[23] FENG, J. and KURTZ, T. G. (2006). *Large Deviations for Stochastic Processes. Mathematical Surveys and Monographs* **131**. Amer. Math. Soc., Providence, RI. MR2260560 https://doi.org/10.1090/surv/131

[24] FLEMING, W. H. (1977/78). Exit probabilities and optimal stochastic control. *Appl. Math. Optim.* **4** 329–346. MR0512217 https://doi.org/10.1007/BF01442148

[25] GLYNN, P. W. and MEYN, S. P. (1996). A Liapounov bound for solutions of the Poisson equation. *Ann. Probab.* **24** 916–931. MR1404536 https://doi.org/10.1214/aop/1039639370

[26] GONG, F. and WU, L. (2006). Spectral gap of positive operators and applications. *J. Math. Pures Appl.* (9) **85** 151–191. MR2199011 https://doi.org/10.1016/j.matpur.2004.11.004

[27] HAIRER, M. and MATTINGLY, J. C. (2008). Spectral gaps in Wasserstein distances and the 2D stochastic Navier–Stokes equations. *Ann. Probab.* **36** 2050–2091. MR2478676 https://doi.org/10.1214/08-AOP392

[28] 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 https://doi.org/10.1007/s00440-009-0250-6

[29] HAUSSMANN, W. and POTTINGER, P. (1976). On multivariate approximation by continuous linear operators. In *Constructive Theory of Functions of Several Variables: Proceedings of a Conference Held at Oberwolfach April 25–May 1* (W. Schempp and K. Zeller, eds.) 101–108. Springer, Berlin.

[30] HUANG, J., KONTOYIANNIS, I. and MEYN, S. P. (2002). The ODE method and spectral theory of Markov operators. In *Proceedings of the Workshop Held at the University of Kansas, Lawrence, Kansas, October 18–20, 2001* (T. E. Duncan and B. Pasik-Duncan, eds.) *Lecture Notes in Control and Information Sciences* **280**. Springer, Berlin.

[31] JAMISON, B. (1970). Irreducible Markov operators on $C(S)$. *Proc. Amer. Math. Soc.* **24** 366–370. MR0250046 https://doi.org/10.2307/2036363

[32] KINGSLEY, E. H. (1951). Bernstein polynomials for functions of two variables of class $C^{(k)}$. *Proc. Amer. Math. Soc.* **2** 64–71. MR0042548 https://doi.org/10.2307/2032622

[33] KONTOYIANNIS, I. and MEYN, S. P. (2003). Spectral theory and limit theorems for geometrically ergodic Markov processes. *Ann. Appl. Probab.* **13** 304–362. MR1952001 https://doi.org/10.1214/aoap/1042765670

[34] KONTOYIANNIS, I. and MEYN, S. P. (2005). Large deviations asymptotics and the spectral theory of multiplicatively regular Markov processes. *Electron. J. Probab.* **10** 61–123. MR2120240 https://doi.org/10.1214/EJP.v10-231

[35] KONTOYIANNIS, I. and MEYN, S. P. (2012). Geometric ergodicity and the spectral gap of non-reversible Markov chains. *Probab. Theory Related Fields* **154** 327–339. MR2981426 https://doi.org/10.1007/s00440-011-0373-4

[36] KONTOYIANNIS, I. and MEYN, S. P. (2017). Approximating a diffusion by a finite-state hidden Markov model. *Stochastic Process. Appl.* **127** 2482–2507. MR3660880 https://doi.org/10.1016/j.spa.2016.11.004

[37] KUNITA, H. (1990). *Stochastic Flows and Stochastic Differential Equations. Cambridge Studies in Advanced Mathematics* **24**. Cambridge Univ. Press, Cambridge. MR1070361

[38] LAUGESEN, R. S., MEHTA, P. G., MEYN, S. P. and RAGINSKY, M. (2015). Poisson's equation in nonlinear filtering. *SIAM J. Control Optim.* **53** 501–525. MR3311867 https://doi.org/10.1137/13094743X

[39] MAJER, P. (2012). Multivariate Bernstein polynomials for approximation of derivatives (answer). *MathOverflow*. Available at https://mathoverflow.net/q/111257.

[40] MEYN, S. (2008). *Control Techniques for Complex Networks*. Cambridge Univ. Press, Cambridge. MR2372453

[41] MEYN, S. and TWEEDIE, R. L. (2009). *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2509253 https://doi.org/10.1017/CBO9780511626630

[42] PARDOUX, E. and VERETENNIKOV, A. Y. (2001). On the Poisson equation and diffusion approximation. I. *Ann. Probab.* **29** 1061–1085. MR1872736 https://doi.org/10.1214/aop/1015345596

[43] QIN, Q. and HOBERT, J. P. (2019). Geometric convergence bounds for Markov chains in Wasserstein distance based on generalized drift and contraction conditions. ArXiv E-prints. Available at arXiv:1902.02964.

[44] RADHAKRISHNAN, A., DEVRAJ, A. and MEYN, S. (2016). Learning techniques for feedback particle filter design. In *55th Conference on Decision and Control* 5453–5459.

[45] RHEE, C.-H. and GLYNN, P. (2017). Lyapunov conditions for differentiability of Markov chain expectations: The absolutely continuous case. ArXiv e-prints.

[46] RIESZ, F. and NAGY, B. S. (1955). *Functional Analysis*. Ungar, New York. Translated from the 2d French by Leo F. Boron.

[47] SCHWEITZER, P. J. (1968). Perturbation theory and finite Markov chains. *J. Appl. Probab.* **5** 401–413. MR0234527 https://doi.org/10.2307/3212261

[48] WU, L. (2000). Uniformly integrable operators and large deviations for Markov processes. *J. Funct. Anal.* **172** 301–376. MR1753178 https://doi.org/10.1006/jfan.1999.3544

[49] WU, L. (2001). Large and moderate deviations and exponential convergence for stochastic damping Hamiltonian systems. *Stochastic Process. Appl.* **91** 205–238. MR1807683 https://doi.org/10.1016/S0304-4149(00)00061-2

[50] WU, L. M. (1995). Large deviations for Markov processes under superboundedness. *C. R. Acad. Sci. Paris Sér. I Math.* **321** 777–782. MR1354725