

# Geometric Ergodicity in a Weighted Sobolev Space

Adithya Devraj\*      Ioannis Kontoyiannis†      Sean Meyn‡

July 19, 2019

## Abstract

For a discrete-time Markov chain  $\mathbf{X} = \{X(t)\}$  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  $v: \mathbb{R}^\ell \rightarrow [1, \infty)$  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  $\pi$  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  $t \geq 0$ :

$$\left| \mathbb{E}_x[f(X(t))] - \pi(f) \right| \leq B e^{-\delta t} v(x), \quad \|\nabla \mathbb{E}_x[f(X(t))]\|_2 \leq B e^{-\delta t} v(x),$$

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.

**Keywords:** Markov chain, stochastic Lyapunov function, discrete spectrum, sensitivity process, weighted Sobolev space, Lyapunov exponent

**2010 AMS Subject Classification:** 60J05, 60J35, 37A30, 47H20.

---

\***Corresponding author.** Department of Electrical and Computer Engineering, University of Florida, Gainesville, USA. Email: [adithyamdevraj@ufl.edu](mailto:adithyamdevraj@ufl.edu)

†Department of Informatics, Athens University of Economics and Business, Patission 76, Athens 10434, Greece. Email: [yiannis@aueb.gr](mailto:yiannis@aueb.gr). I.K. was supported by the European Union and Greek National Funds through the Operational Program Education and Lifelong Learning of the National Strategic Reference Framework through the Research Funding Program Thales-Investing in Knowledge Society through the European Social Fund.

‡Department of Electrical and Computer Engineering, University of Florida, Gainesville, USA. Email: [meyn@ece.ufl.edu](mailto:meyn@ece.ufl.edu). S.M. and A.D. were supported by NSF grants EPCN-1609131 and CPS-1259040.

## 1 Introduction

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

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

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

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

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

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  $f : \mathsf{X} \rightarrow \mathbb{R}$  on the right and on signed measures  $\nu$  on  $(\mathsf{X}, \mathcal{B})$  on the left, respectively, as,

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

whenever the integrals exist. Also, for any signed measure  $\nu$  on  $(\mathsf{X}, \mathcal{B})$  and any function  $f : \mathsf{X} \rightarrow \mathbb{R}$  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  $L_\infty$  norm.

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

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

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

The Markov chain  $\mathbf{X}$  is  $v$ -uniformly ergodic [32, 25] whenever there exists a function  $v$ , a unique invariant probability measure  $\pi$ , and constants  $b_0 < \infty$  and  $0 < \rho_0 < 1$ , such that, for each function  $f \in L_\infty^v$ ,

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

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) of [31].

## 1.1 Motivation and background

Let  $c : \mathbf{X} \rightarrow \mathbb{R}$  be a given function on the state space of  $\mathbf{X}$ . One starting point for the classical study of the long-term behaviour of  $\mathbf{X}$  is the development of conditions for the existence of the mean ergodic limit,

$$\bar{c} := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=0}^{n-1} \mathbb{E}[c(X(t)) \mid X(0) = x], \quad (4)$$

and of the function,

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

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

$$h(x) - \mathbb{E}[h(X(1)) \mid X(0) = x] = c(x) - \bar{c}, \quad x \in \mathbf{X}. \quad (6)$$

For example, if  $\mathbf{X}$  is  $v$ -uniformly ergodic, then in addition to the convergence (3) of  $P^t(x, \cdot)$  to its unique invariant probability measure  $\pi$ , the ergodic averages of  $c(X(t))$  converge a.s. to  $\bar{c} = \pi(c)$ , and their associated central-limit-theorem variance is naturally expressed in terms of  $h$  [32, 2],

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

Moreover, if  $c \in L_{\infty}^v$  then  $h$  is also in  $L_{\infty}^v$  [17].

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

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

where each  $h_{\alpha}$  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  $c \in L_{\infty}^v$  [25].

The main goal of the present work is to develop natural conditions that guarantee appropriate *smoothness* properties of  $\bar{c}$ ,  $h$  and  $h_{\alpha}$ . In particular, as described next, we show that the derivative of  $P^t c(x) = \mathbb{E}[c(X(t)) \mid X(0) = x]$  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  $h_{\alpha}$ ; 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*,  $\alpha$  is the *discount factor*,  $\bar{c}$  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 [10], and to gain approximation for the feedback particle filter [29, 34].

## 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  $\mathcal{S} = \{\mathcal{S}(t) : t \geq 0\}$ , whose  $(i, j)$  component is defined at time  $t$  by:

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

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

$$\mathcal{S}(t+1) = \mathcal{A}(t+1)\mathcal{S}(t), \quad \mathcal{S}(0) = I, \quad (9)$$

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

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

$$\nabla^{\mathcal{S}} f(X(t)) := \mathcal{S}^T(t) \nabla f(X(t)). \quad (10)$$

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  $g : \mathbf{X} \rightarrow \mathbb{R}^\ell$ : For  $t \geq 1$ ,

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

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

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

which implies that:

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

**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  $c : \mathbf{X} \rightarrow \mathbb{R}$  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  $t \rightarrow \infty$ , 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  $h_\alpha$  defined in (7):

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

### 1.3 Prior research

The sensitivity process defined in (8) is used to define the Lyapunov exponent,

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

here and in (13,14),  $\|\cdot\|$  can be taken to be any matrix norm. A negative exponent implies a topological notion of coupling: Suppose that  $\Lambda_1$  is a negative constant, independent of the initial condition. If  $\mathbf{X}$  and  $\mathbf{X}'$  are two realizations of the Markov chain with different initial states, it follows from the mean value theorem that, with probability one,

$$\lim_{t \rightarrow \infty} \|\mathbf{X}(t) - \mathbf{X}'(t)\|_2 = 0,$$

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  $p$ th mean,

$$\bar{\Lambda}_p := \lim_{t \rightarrow \infty} \frac{1}{t} \log(\mathbb{E}_x[\|\mathcal{S}(t)\|^p]), \quad (13)$$

is for diffusion processes in continuous time [28, 5, 1].

Verifiable conditions for a negative Lyapunov exponent are established in [3] for a class of hidden Markov models, and in [21] for a general class of stochastic sequences of the form (1); coupling results that suggest a negative Lyapunov exponent are established in [19] for a class of diffusions. In these papers the main results are established *without*  $\psi$ -irreducibility. As discussed in [32, 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 prior work most closely related to our results is [19], although the development there is entirely for continuous-time processes. A central goal of [19] is to obtain coupling bounds in a topological sense for a class of stochastic Navier-Stokes equations, but results for a general stochastic flow on a Banach space are also derived. Assumption 4 of [19] imposes a contraction bound directly on the sensitivity process, which, in the finite-dimensional case implies that, for a function  $v \geq 1$ , constants  $k < \infty$ ,  $\rho < 1$ , and some time  $t_0 > 0$ ,

$$\mathbb{E}_x[v(\mathcal{X}(t_0))] + \mathbb{E}_x[\|\mathcal{S}(t_0)\|v(\mathcal{X}(t_0))] \leq kv(x)^\rho, \quad x \in \mathbf{X}. \quad (14)$$

This and other assumptions imply the desired coupling result in [19, 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. (17) in Section 2 below), also appears as  $\|f\|_{V^r}$  in [19, p. 21].

The present approach is complementary to [19]. Their conclusions are far stronger than those presented here: They obtain a very strong topological coupling of the process, and they do not require  $\psi$ -irreducibility. However, these strong conclusions require strong assumptions. Most significant is that their assumptions imply the contraction bound (14) that is not easily verified in applications, and is unlike any assumption imposed in the present work.

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. In particular, it is not known whether the assumptions imposed here can be used to establish any form of topological coupling. And, 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  $g: \mathbf{X} \rightarrow \mathbb{R}$  in an appropriate Banach space, the following can be uniformly bounded above:

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

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

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

$$\log \mathbb{E}[\exp\{V(X(t+1)) - V(X(t))\} \mid X(t) = x] \leq -\delta W(x), \quad x \in C^c. \quad (15)$$

Condition (DV3) is an essential ingredient in much of the prior work of Donsker and Varadhan on large deviation theory for Markov models [11, 12, 13], and it is used to bound rates of convergence for a Markov chain for both mean and “multiplicative” ergodic theory in [4, 26, 24]. Also, the discrete-time counterpart of [19, Assumption 4, equation (15)] implies (14) 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, an  $n$ -step transition kernel can be approximated by its truncation to a compact set arbitrarily closely in an associated induced operator norm [4, 26, 24]; see also [38, 39] 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 [18, 9, 8]. The implications of this and similar drift conditions are the main focus of [32]. In particular, in this monograph and subsequent papers [17, 25, 23], drift conditions are used to obtain existence and bounds on solutions to Poisson’s equation. A log-Sobolev inequality is the condition used in [29] to establish the existence of a smooth solution to Poisson’s equation for a diffusion. Somewhat more explicit sufficient conditions for a smooth solution to Poisson’s equation are obtained in [33] for elliptic diffusions.

Poisson’s equation is one tool used in addressing parametric sensitivity in Markov chains, starting with the 50-year-old work of Schweitzer [37]; *infinitesimal perturbation analysis* is a well-known application of these techniques [7]. A modern treatment is contained in the very recent work [35]. 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.

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 [26]. 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  $V : \mathsf{X} \rightarrow (0, \infty)$  satisfying the drift condition (DV3) of [25, 26]; 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  $f : \mathsf{X} \rightarrow \mathbb{R}$ . For each  $k \geq 1$  denote,

$$\|f\|_{v,k} = \max_{|\alpha| \leq k} \|D^\alpha f\|_v, \quad (16)$$

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

$$\|f\|_{v,1} = \max\left\{\|f\|_v, \|\partial_1 f\|_v, \dots, \|\partial_\ell f\|_v\right\}, \quad (17)$$

where  $\partial_i$  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  $\{f \in L_\infty^v : f \text{ is continuous}\}$  with norm  $\|\cdot\|_v$ . For each  $k \geq 1$  we also define the spaces,

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

equipped with the norm defined in (16). This introduces two new restrictions on any function  $f \in L_\infty^{v,k}$ : The  $k$ th 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).

- (i) The process  $\mathbf{N}$  does not depend upon the initial condition  $X(0)$ .  
(ii) The function  $a$  is continuously differentiable in its first variable, with:
- $$\left. \begin{array}{l} \sup_{x,n} \|\nabla a(x,n)\| < \infty. \end{array} \right\} \quad (\mathbf{A1})$$

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

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

### 2.1 Irreducibility, densities and drift

The general ergodic theory of Markov chains as developed in [32] 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:

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,

$$P^{t_0}(x, A) = \int_A p_{t_0}(x, y) dy, \quad x \in \mathsf{X}, \quad A \in \mathcal{B} \quad \left. \vphantom{\int_A} \right\} \quad (\text{A2})$$

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

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

The representation (1) implies the Feller property, i.e., 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.

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,

$$P^t(x, O) > 0, \quad \text{for all } t \geq 0 \text{ sufficiently large.} \quad \left. \vphantom{P^t} \right\} \quad (\text{A3})$$

Under assumptions (A2) and (A3), the chain is  $\psi$ -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 [32, Theorem 6.2.1].

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

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

for any  $f$  for which the expectation is defined for all  $x$ . Fleming’s *nonlinear generator* [16, 14, 40, 15, 26] is defined via,

$$\mathcal{H}(F) := \log(Pe^F) - F, \quad (18)$$

for any measurable function  $F$  on  $\mathsf{X}$  such that  $Pe^F$  exists.

We say [25, 26] that the *Lyapunov drift criterion* (DV3) holds with respect to the Lyapunov function  $V: \mathsf{X} \rightarrow (0, \infty]$ , if there exist a function  $W: \mathsf{X} \rightarrow [1, \infty)$ , a compact set  $C \subset \mathsf{X}$ , and constants  $\delta > 0$ ,  $b < \infty$ , such that,

$$\mathcal{H}(V) \leq -\delta W + b\mathbb{1}_C. \quad (\text{DV3})$$

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

Condition (DV3) holds with respect to functions  $V, W$  that are continuously differentiable and have compact sublevel sets.  $\left. \vphantom{\text{Condition}} \right\} \quad (\text{A4})$



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

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

## 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  $\pi$  [26, 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  $\widehat{P}: L_\infty^v \rightarrow L_\infty^v$  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 ,

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

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  $\widehat{P}: L_\infty^{v,k} \rightarrow L_\infty^{v,k}$ :

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

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  $\rho_0 < 1$  such that:*

$$\|\widetilde{P}^t\|_{v^\eta,1} \leq b_0 \rho_0^t, \quad t \geq t_1. \tag{19}$$

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

$$|\mathbb{E}_x[f(X(t))] - \pi(f)| \leq b_0 \rho_0^t \|f\|_{v^\eta,1} v^\eta(x), \tag{20}$$

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

The proof of the theorem, given in Section 3, is similar to the proof of  $v$ -uniform ergodicity in prior work [25]. Under assumptions (A1)–(A4) it is shown that the semigroup generated by  $\widetilde{P}$  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  $\eta$  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  $L_\infty^v$ , and  $h$  is a solution to the Poisson's equation (6):  $h - Ph = c - \bar{c}$ ; cf [17]. 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:

$$\nabla h = \sum_{t=0}^{\infty} \nabla P^t c. \quad (21)$$

This will in fact follow from (19), once we could establish that  $\|\tilde{P}^t\|_{v,1}$  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  $\pi$ -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 (21), and also,*

$$\nabla h = \Omega \nabla c := \sum_{t=0}^{\infty} Q^t \nabla c. \quad (22)$$

Note that, in the theorem, the boundedness of  $\Omega$  is only established on the space of functions of the form  $\nabla f$  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 [17, 32]; 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  $t \geq 1$ :

$$\nabla P^t f = Q^t \nabla f. \quad (23)$$

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  $h_\alpha$ , 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  $h_\alpha$  in (7) exists as an element of  $L_\infty^{v,\eta}$ . It is a solution to the following fixed point equation,*

$$c + \alpha Ph_\alpha - h_\alpha = 0 \quad (24)$$

*and it is unique among all functions in  $L_\infty^{v,\eta}$ .*

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

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

$$\nabla h_\alpha = \Omega_\alpha \nabla c := \sum_{t=0}^{\infty} \alpha^t Q^t \nabla c \quad (25)$$

Once again, the fact that  $h_\alpha$  is bounded and uniquely solves (24), follows from earlier work [31]. The proofs of parts (i) and (ii) are essentially identical to the proofs of the corresponding results in Theorem 2.3.

### 3 Spectral Theory

**Proposition 3.1.** *For any function  $v: X \rightarrow [1, \infty)$ , the normed spaces  $L_\infty^v$ ,  $L_\infty^{v,0}$  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  $T: L_\infty^{v,k} \rightarrow L_\infty^{v,k}$  has *finite rank* if there are functions  $\{s_i\} \subset L_\infty^{v,k}$ , measures  $\{\nu_j\} \subset \mathcal{M}_1^v$ , and constants  $\{m_{ij}\}$  such that,

$$T = \sum_{i,j=1}^N m_{ij} s_i \otimes \nu_j \quad (26)$$

where  $[s \otimes \nu](x, dy) := s(x)\nu(dy)$ , and  $N < \infty$ . We say that a linear operator  $\widehat{P}: L_\infty^{v,k} \rightarrow L_\infty^{v,k}$  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  $\widehat{P}: L_\infty^{v,k} \rightarrow L_\infty^{v,k}$  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

$$\xi_v(\widehat{P}) := \lim_{n \rightarrow \infty} \|\widehat{P}^n\|_v^{1/n}. \quad (27)$$

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:

$$\mathcal{P} := \frac{1}{2\pi i} \int_{\partial\{z: |z - z_0| \leq \varepsilon_1\}} [Iz - \widehat{P}]^{-1} dz. \quad (28)$$

For background see the decomposition theorem in [36, Theorem 4.4, p. 421].

The linear operator  $\widehat{P} : L_\infty^{v,k} \rightarrow L_\infty^{v,k}$  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  $L_\infty^v$  for a linear operator  $\widehat{P} : L_\infty^v \rightarrow L_\infty^v$  as defined in [26], which requires that we can find, for each  $\varepsilon > 0$  a positive kernel of the form (26) 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 [26] 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  $\nu_j$  had support on a compact set. This is not necessary here or in the technical results of [26].

The following theorem is an extension of Theorem 3.5 of [26] and it provides the fundamental connection between separability and ergodicity.

**Theorem 3.2** (Separability  $\Rightarrow$  Discrete spectrum). *If the linear operator  $\widehat{P} : L_\infty^{v,k} \rightarrow L_\infty^{v,k}$  is bounded and  $\widehat{P}^{t_1} : L_\infty^{v,k} \rightarrow L_\infty^{v,k}$  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 [26], 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  $\varphi : [0, 1]^N \rightarrow \mathbb{R}$ , with  $N \geq 2$ . For an integer  $m \geq 2$ , the Bernstein approximation is given by:

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

The proof of the following can be found in [22] for the special case  $N = 2$ ; also see [6, 20] for related results, and [30] 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  $\varphi : [0, 1]^N \rightarrow \mathbb{R}$ :*

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

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  $\mu$ : For each  $x \in X$  and  $A \in \mathcal{B}$ ,  $\widehat{P}(x, A) = \int_A r(x, y) \mu(dy)$ . Suppose moreover that the density  $r : \mathbb{R}^\ell \times \mathbb{R}^\ell \rightarrow \mathbb{R}^+$  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  $v: \mathsf{X} \rightarrow [1, \infty)$ : 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*

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

(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 (21):*

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

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

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

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

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

*Proof.* For each  $n, i, \alpha$  and  $x$  we have,

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

where, as before,  $\partial_i$  denotes the partial derivative with respect to the  $i$ th coordinate. Letting  $n \rightarrow \infty$  gives,

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

Continuity of  $\zeta$  implies that  $\nabla g = \zeta$ . □

*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  $\tilde{P}^{t_1}$  in  $L_\infty^{v,1}$  is a subset of its spectrum in  $L_\infty^v$ :  $S_{v,1}(\tilde{P}^{t_1}) \subseteq S_v(\tilde{P}^{t_1})$ . Denote the respective spectral radii by  $\xi_{v,1}(\tilde{P}^{t_1})$  and  $\xi_v(\tilde{P}^{t_1})$  (recall the definition (27)). We obviously have  $\xi_{v,1}(\tilde{P}^{t_1}) \leq \xi_v(\tilde{P}^{t_1})$ . Also, under  $v$ -uniform ergodicity we have  $\xi_v(\tilde{P}^{t_1}) < 1$  [26, Theorem 2.4].

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

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

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

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

which implies (i).

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

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

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

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

it follows that  $h_n \rightarrow h$  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,

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

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,

$$P(x, A) = \int_A p(x, y) dy, \quad x \in \mathsf{X}, \quad A \in \mathcal{B}. \quad \left. \vphantom{\int_A} \right\} \text{(A2')}$$

Assumption (A3) is maintained, which together with (A2') again implies that  $\mathsf{X}$  is  $\psi$ -irreducible and aperiodic.

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

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

- (i) The transition kernel  $P$  satisfies (DV3) with respect to continuous functions  $V, W$ , a compact set  $C \subset \mathbf{X}$ , and constants  $\delta > 0, b < \infty$ .
- (ii) For each  $x \in \mathbf{X}$ ,  $1 \leq i \leq \ell$ ,
- $$\log \int |p'_i(x, y)| \exp(V(y) - V(x)) dy \leq -\delta W(x) + b \mathbb{I}_C(x).$$
- (A4')

The drift condition (DV3) is used here and in [26, 27] to truncate the transition kernel onto a compact subset of the state space. Denote for  $n \geq 1$ ,

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

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

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

The choice is not unique, but fixed throughout the paper. In  $\ell$  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

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

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  $L_\infty^{v,1}$ :

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

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

*Proof.* The fact that  $P : L_\infty^{v,1} \rightarrow L_\infty^{v,1}$  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  $L_\infty^v$ : Proposition 3.9 (i) follows from Lemma B.5 of [26], 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  $1 \leq i, j \leq \ell$ ,  $t \geq 0$ .

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

- (i)  $P^t$  is separable in  $L_\infty^v$ .
- (ii)  $Q_{i,j}^t$  is separable in  $L_\infty^v$  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)  $P^t : L_\infty^{v^\eta, k} \rightarrow L_\infty^{v^\eta, k}$  for  $k = 0, 1$ .
  - (b)  $Q_{i,j}^t : L_\infty^{v^\eta, 0} \rightarrow L_\infty^{v^\eta, 0}$  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  $P^t : L_\infty^{v^\eta, 1} \rightarrow L_\infty^{v^\eta, 1}$ . 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:

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

□

*Proof of Theorem 2.3.* As before, let  $\tilde{c} = c - \bar{c}$ . 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  $\|P^n h\|_v \rightarrow 0$  as  $n \rightarrow \infty$ , 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  $\|P^t \tilde{c}\|_{v^\eta} \rightarrow 0$  as  $t \rightarrow \infty$  and consequently  $h \in L_\infty^{v^\eta}$ . It remains to show that  $h$  is continuous.

Recall from Theorem 3.10 that  $P^t: L_\infty^{v^\eta, 0} \rightarrow L_\infty^{v^\eta, 0}$  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  $t \geq t_1$ ,

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

Define for each  $n \geq 1$ ,

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

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

In particular,

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

as claimed. □

## A Appendix

### A.1 Operator bounds

We begin with sufficient conditions for the identity (23). We first require the following corollary to Lemma 3.6:

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

$$P^t: L_\infty^{v^\eta, 0} \rightarrow L_\infty^{v^\eta, 0}, \quad Q_{i,j}^t: L_\infty^{v^\eta, 0} \rightarrow L_\infty^{v^\eta, 0}, \quad 1 \leq i, j \leq \ell. \quad (29)$$

*Then,  $P^t: L_\infty^{v^\eta, 1} \rightarrow L_\infty^{v^\eta, 1}$  and (23) 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  $f_n = \chi_n f$ . The function  $f_n$  and its partial derivatives are continuous, and  $\sup_n \|f_n\|_{v^\eta, 1} < \infty$ . We have  $\lim_{n \rightarrow \infty} \nabla f_n = \nabla f$ , 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 (29). 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:

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

Lemma 3.6 then implies the desired conclusion, that  $\nabla P^t f = Q^t \nabla f$ . This identity combined with (29) then implies that  $P^t: L_\infty^{v^\eta, 1} \rightarrow L_\infty^{v^\eta, 1}$ .  $\square$

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  $L_\infty^v$ , and thence  $L_\infty^{v, 0}$ . 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  $\{f, \zeta_1, \dots, \zeta_\ell\} \subset L_\infty^{v, 0}$  such that

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

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:  $\lim_{n \rightarrow \infty} \|f_n - f\|_{v, 1} = 0$ .  $\square$

## A.2 Truncations

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

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

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  $Z$  it can be approximated by its truncation on both sides,*

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

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

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

*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  $L_\infty^v$  we have  $I_{\chi_n} P \xrightarrow{v} P$  (see [26, Lemma B.4]), and hence:

$$(I_{\chi_n} P)^2 \xrightarrow{v} P^2. \quad (30)$$

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

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

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

$$\begin{aligned} |(I_{\chi_n} P)^2(x, A)| &= \left| \int_{R_{n+1}} \left\{ \chi_n(x) P(x, dy) \chi_n(y) P(y, A) \right\} \right| \\ &\leq \int_{R_{n+1}} \beta_n^0(dy) P(y, A) := \beta_n(A), \end{aligned}$$

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

$$\begin{aligned} \|(I_{\chi_n} P)^2 I_{\chi_m} f - P^2 f\|_v &\leq \|(I_{\chi_n} P)^2 (1 - \chi_m) f\|_v + \|(I_{\chi_n} P)^2 f - P^2 f\|_v \\ &\leq \beta_n(v I_{R_m^c}) \|f\|_v + \|(I_{\chi_n} P)^2 f - P^2 f\|_v, \end{aligned}$$

and applying (30),

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

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

$$\lim_{m,n \rightarrow \infty} \|\partial_i(I_{\chi_n}P)^2 I_{\chi_m} - \partial_i P^2\|_v = 0, \quad (31)$$

where, again,  $\partial_i$  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,

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

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

$$\lim_{n \rightarrow \infty} \|\partial_i(I_{\chi_n}P)^2 f - \partial_i P^2 f\|_v = 0. \quad (32)$$

Furthermore, the two assumptions imply a bound of the form

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

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

$$\begin{aligned} |\partial_i(I_{\chi_n}P)^2(x, A)| &= \left| \int_{R_{n+1}} \frac{\partial}{\partial x_i} \left\{ \chi_n(x) P(x, dy) \chi_n(y) P(y, A) \right\} \right| \\ &\leq \int_{R_{n+1}} \gamma_n^0(dy) P(y, A) := \gamma_n(A). \end{aligned}$$

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

$$\begin{aligned} \|\partial_i(I_{\chi_n}P)^2 I_{\chi_m} f - \partial_i P^2 f\|_v &\leq \|\partial_i(I_{\chi_n}P)^2(1 - \chi_m) f\|_v + \|\partial_i(I_{\chi_n}P)^2 f - \partial_i P^2 f\|_v \\ &\leq \gamma_n(v I_{R_m^c}) \|f\|_v + \|\partial_i(I_{\chi_n}P)^2 f - \partial_i P^2 f\|_v \end{aligned}$$

Combining this with (32) implies that (31) 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  $1 \leq i, j \leq \ell$ ,  $t \geq 0$ .

**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  $t \geq 1$  and  $1 \leq i, j \leq \ell$ .*

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

$$\begin{aligned} Z_t &: L_\infty^{v^\eta, 0} \rightarrow L_\infty^{v^\eta, 0}, \quad t \geq 1, \\ Z_t &: L_\infty^{v^\eta} \rightarrow L_\infty^{v^\eta, 0}, \quad t \geq t_0. \end{aligned}$$

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

$$\begin{aligned} Z_t I_{\chi_n} &: L_\infty^{v^\eta, 0} \rightarrow L_\infty^{v^\eta, 0}, \quad t \geq 1, \\ Z_t I_{\chi_n} &: L_\infty^{v^\eta} \rightarrow L_\infty^{v^\eta, 0}, \quad t \geq t_0. \end{aligned}$$

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

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

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

$$\begin{aligned} Z_t &: L_\infty^{v^\eta, 0} \rightarrow L_\infty^{v^\eta, 0}, \\ \text{and } Z_t &: L_\infty^{v^\eta} \rightarrow L_\infty^{v^\eta, 0}, \quad \text{provided } t \geq t_0. \end{aligned}$$

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

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

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

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

$$\begin{aligned} I_{\chi_n} P^t I_{\chi_n} &\xrightarrow{v} P^t \\ I_{\chi_n} Q_{i,j}^t I_{\chi_n} &\xrightarrow{v} Q_{i,j}^t, \quad 1 \leq i, j \leq \ell. \end{aligned} \tag{33}$$

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  $f \in L_\infty^{v,1}$ ,

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

Lemma A.3 along with (33) implies that the assumptions of Lemma A.1 are satisfied (with  $t \geq t_1$ ):

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

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 bounded,

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

where,

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

The pre-multiplication by  $I_{R_n^c}$  is justified since  $\nabla \chi_n = 0$  on  $R_n$ . Equation (33) along with Lemma A.2 implies that  $\|I_{R_n^c} P^t\|_v \rightarrow 0$  as  $n \rightarrow \infty$ , and hence  $\lim_{n \rightarrow \infty} \varepsilon_1(n) = 0$ . Therefore,

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

Once more applying equation (33), it is straightforward to see that there is a vanishing sequence  $\{\varepsilon_2(n)\}$  such that:

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

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

The justification of the representation (22) for  $\nabla h$  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)  $P^t I_{\chi_n} \xrightarrow{v^\eta} P^t$ ;
- (ii)  $Q_{i,j}^t I_{\chi_n} \xrightarrow{v^\eta} Q_{i,j}^t$ , 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  $t \geq 1$ ,

$$\begin{aligned} |P^t f(x) - P^t f_n(x)| &\leq \left| \int_{R_n^c} P^t(x, dy) f_n(y) \right| \\ &\leq \|f\|_{v^\eta} \int_{R_n^c} P^t(x, dy) v(y) \left( \frac{v^\eta(y)}{v(y)} \right) \\ &\leq \|f\|_{v^\eta} \left[ \sup_{y' \in R_n^c} v^{\eta-1}(y') \right] \int_{R_n^c} P^t(x, dy) v(y) \\ &\leq \|f\|_{v^\eta} \varepsilon(n), \end{aligned}$$

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

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

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

$\square$

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  $\eta = 1$ ,*

$$P^t: L_\infty^{v^\eta, 0} \rightarrow L_\infty^{v^\eta, 0}, \quad Q_{i,j}^t: L_\infty^{v^\eta, 0} \rightarrow L_\infty^{v^\eta, 0}, \quad 1 \leq i, j \leq \ell. \quad (34)$$

(ii) *The conclusions (34) hold for all  $t \geq 1$  when  $\eta \in (0, 1)$ .*  $\square$

*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,

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

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

Similarly, using Lemma A.4 (ii),

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

for any  $g \in L_\infty^{v, 0}$ . This again implies that  $Q_{i,j}^t: L_\infty^{v^\eta, 0} \rightarrow L_\infty^{v^\eta, 0}$ ,  $\eta \in (0, 1]$ , 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  $\eta \in (0, 1]$ ,

$$P^t: L_\infty^{v^\eta, 0} \rightarrow L_\infty^{v^\eta, 0}, \quad \text{and} \quad Q_{i,j}^t: L_\infty^{v^\eta, 0} \rightarrow L_\infty^{v^\eta, 0}, \quad 1 \leq i, j \leq \ell. \quad (35)$$

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 (35) holds for each  $t \geq 1$  and  $\eta \in (0, 1)$ . Consequently, results (a) and (c) follow as before by applying Lemma A.1.  $\square$

### 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  $(x, y) \in R_{n+1} \times R_{n+1}$ ,

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

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,

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

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

The following approximating kernel has finite rank:

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

We also have,

$$\begin{aligned} \left| \widehat{P}g(x) - T_\varepsilon g(x) \right| & \leq \int \left| r_v(x, y) - r_v^\varepsilon(x, y) \right| \left| \frac{g(y)}{v(y)} \right| \mu(dy) \\ & \leq \sup_{x, y} \left| r_v(x, y) - r_v^\varepsilon(x, y) \right| \sup_z \left| \frac{g(z)}{v(z)} \right| \\ & \leq \varepsilon \|g\|_v, \end{aligned}$$

and,

$$\begin{aligned} \left| \frac{\partial}{\partial x_i} \widehat{P}g(x) - \frac{\partial}{\partial x_i} T_\varepsilon g(x) \right| & = \left| \frac{\partial}{\partial x_i} \int \Delta_r^\varepsilon(x, y) \frac{g(y)}{v(y)} \mu(dy) \right| \\ & = \left| \lim_{\delta \rightarrow 0} \frac{1}{\delta} \int \left( \Delta_r^\varepsilon(x + \delta e^i, y) - \Delta_r^\varepsilon(x, y) \right) \frac{g(y)}{v(y)} \mu(dy) \right|, \end{aligned}$$

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



Since, both  $r_v$  and  $r_v^\varepsilon$  are  $C^1$ , the mean value theorem gives,

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

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

$$\begin{aligned} \left| \frac{\partial}{\partial x_i} \widehat{P}g(x) - \frac{\partial}{\partial x_i} T_\varepsilon g(x) \right| &\leq \int \limsup_{\delta \rightarrow 0} \frac{1}{\delta} \left| \Delta_r^\varepsilon(x + \delta e^i, y) - \Delta_r^\varepsilon(x, y) \right| \left| \frac{g(y)}{v(y)} \right| \mu(dy) \\ &\leq \sup_{x, y} \left| \frac{\partial}{\partial x_i} r_v(x, y) - \frac{\partial}{\partial x_i} r_v^\varepsilon(x, y) \right| \|g\|_v \\ &\leq \varepsilon \|g\|_v. \end{aligned}$$

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

## References

- [1] L. Arnold and H. Crauel. Random dynamical systems. In L. Arnold, H. Crauel, and J.-P. Eckmann, editors, *Lyapunov Exponents*, volume 1486 of *Lecture Notes in Mathematics*, pages 1–22. Springer Berlin Heidelberg, 1991.
- [2] S. Asmussen and P. W. Glynn. *Stochastic Simulation: Algorithms and Analysis*, volume 57 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, New York, 2007.
- [3] R. Atar and O. Zeitouni. Lyapunov exponents for finite state nonlinear filtering. *SIAM J. Control Optim.*, 35(1):36–55, 1997.
- [4] S. Balaji and S. P. Meyn. Multiplicative ergodicity and large deviations for an irreducible Markov chain. *Stoch. Proc. Applns.*, 90(1):123–144, 2000.
- [5] P. H. Baxendale. Lyapunov exponents and relative entropy for a stochastic flow of diffeomorphisms. *Prob. Theory Related Fields*, 81(4):521–554, 1989.
- [6] P. L. Butzer. On two-dimensional Bernstein polynomials. *Canadian Journal of Mathematics*, 5:107–113, 1953.
- [7] X.-R. Cao and H.-F. Chen. Potentials, perturbation realization, and sensitivity analysis of Markov processes. *IEEE Trans. Automat. Control*, 42:1382–1393, 1997.
- [8] P. Cattiaux and A. Guillin. Hitting times, functional inequalities, Lyapunov conditions and uniform ergodicity. *Journal of Functional Analysis*, 272(6):2361–2391, 2017.
- [9] P. Cattiaux, A. Guillin, F.-Y. Wang, and L. Wu. Lyapunov conditions for logarithmic Sobolev and super Poincaré inequality. Available on Math. ArXiv 0712.0235, 2007.
- [10] A. M. Devraj and S. P. Meyn. Differential TD learning for value function approximation. In *55th Conference on Decision and Control*, pages 6347–6354, Dec 2016.
- [11] M. Donsker and S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time. I. II. *Comm. Pure Appl. Math.*, 28:1–47; *ibid.* **28** (1975), 279–301, 1975.
- [12] M. Donsker and S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time. III. *Comm. Pure Appl. Math.*, 29(4):389–461, 1976.
- [13] M. Donsker and S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time. IV. *Comm. Pure Appl. Math.*, 36(2):183–212, 1983.
- [14] J. Feng. Martingale problems for large deviations of Markov processes. *Stoch. Proc. Applns.*, 81:165–212, 1999.
- [15] J. Feng and T. G. Kurtz. *Large Deviations for Stochastic Processes*, volume 131 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, U.S.A., 2006.

- [16] W. Fleming. Exit probabilities and optimal stochastic control. *App. Math. Optim.*, 4:329–346, 1978.
- [17] P. W. Glynn and S. P. Meyn. A Liapounov bound for solutions of the Poisson equation. *Ann. Probab.*, 24(2):916–931, 1996.
- [18] F. Gong and L. Wu. Spectral gap of positive operators and applications. *J. Math. Pures Appl.*, 85:151–191, 2006.
- [19] M. Hairer and J. C. Mattingly. Spectral gaps in Wasserstein distances and the 2D stochastic Navier-Stokes equations. *The Annals of Probability*, pages 2050–2091, 2008.
- [20] W. Haußmann and P. Pottinger. On multivariate approximation by continuous linear operators. In W. Schempp and K. Zeller, editors, *Constructive Theory of Functions of Several Variables: Proceedings of a Conference Held at Oberwolfach April 25 May 1, 1976*, pages 101–108, Berlin, Heidelberg, 1977. Springer Berlin Heidelberg.
- [21] J. Huang, I. Kontoyiannis, and S. P. Meyn. The ODE method and spectral theory of Markov operators. In T. E. Duncan and B. Pasik-Duncan, editors, *Proceedings of the workshop held at the University of Kansas, Lawrence, Kansas, October 18–20, 2001*, volume 280 of *Lecture Notes in Control and Information Sciences*, pages 205–222, Berlin, 2002. Springer-Verlag.
- [22] E. H. Kingsley. Bernstein polynomials for functions of two variables of class  $C^k$ . *Proc. of the American Mathematical Society*, 2(1):64–71, 1951.
- [23] I. Kontoyiannis and S. Meyn. Geometric ergodicity and the spectral gap of non-reversible Markov chains. *Prob. Theory Related Fields*, 154(1-2):327–339, 2012. 10.1007/s00440-011-0373-4.
- [24] I. Kontoyiannis and S. Meyn. Approximating a diffusion by a finite-state hidden Markov model. *Stochastic Processes and their Applications*, 127(8):2482–2507, 2017.
- [25] I. Kontoyiannis and S. P. Meyn. Spectral theory and limit theorems for geometrically ergodic Markov processes. *Ann. Appl. Probab.*, 13:304–362, 2003.
- [26] I. Kontoyiannis and S. P. Meyn. Large deviations asymptotics and the spectral theory of multiplicatively regular Markov processes. *Electron. J. Probab.*, 10(3):61–123 (electronic), 2005.
- [27] I. Kontoyiannis and S. P. Meyn. Approximating a diffusion by a Hidden Markov Model. <http://arxiv.org/abs/0906.0259>, and submitted for publication., June 2009.
- [28] H. Kunita. *Stochastic Flows and Stochastic Differential Equations*. Cambridge University Press, Cambridge, 1990.
- [29] R. S. Laugesen, P. G. Mehta, S. P. Meyn, and M. Raginsky. Poisson’s equation in nonlinear filtering. *SIAM J. Control Optim.*, 53(1):501–525, 2015.

- [30] P. Majer. Multivariate Bernstein polynomials for approximation of derivatives (answer). MathOverflow. URL: <https://mathoverflow.net/q/111257>, Nov. 02, 2012.
- [31] S. P. Meyn. *Control Techniques for Complex Networks*. Cambridge University Press, 2007. Pre-publication edition available online.
- [32] S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Cambridge University Press, Cambridge, second edition, 2009. Published in the Cambridge Mathematical Library. 1993 edition online.
- [33] E. Pardoux and Y. Veretennikov. On the Poisson equation and diffusion approximation. I. *Ann. Probab.*, 29(3):1061–1085, 07 2001.
- [34] A. Radhakrishnan, A. Devraj, and S. Meyn. Learning techniques for feedback particle filter design. In *55th Conference on Decision and Control*, pages 5453–5459, Dec 2016.
- [35] C.-H. Rhee and P. Glynn. Lyapunov conditions for differentiability of Markov chain expectations: the absolutely continuous case. *arXiv preprint arXiv:1707.03870*, 2017.
- [36] F. Riesz and B. S. Nagy. *Functional Analysis*. Ungar, New York, NY, 1955. Translated from the 2d French by Leo F. Boron.
- [37] P. J. Schweitzer. Perturbation theory and finite Markov chains. *J. Appl. Prob.*, 5:401–403, 1968.
- [38] L. Wu. Large deviations for Markov processes under superboundedness. *C. R. Acad. Sci Paris Série I*, 324:777–782, 1995.
- [39] L. Wu. Uniformly integrable operators and large deviations for Markov processes. *Journal of Functional Analysis*, 172(2):301–376, 2000.
- [40] L. Wu. Large and moderate deviations and exponential convergence for stochastic damping Hamiltonian systems. *Stoch. Proc. Applns.*, 91(2):205–238, 2001.