Electronic Journal of Probability

Vol. 10 (2005), Paper no. 3, pages 61-123.

Journal URL
http://www.math.washington.edu/∼ejpecp/

Large Deviations Asymptotics and the Spectral Theory of Multiplicatively Regular Markov Processes

I. Kontoyiannis[1]
Division of Applied Mathematics and Dept of Computer Science
Brown University, Box F, 182 George St., Providence, RI 02912, USA
Email: `yiannis@dam.brown.edu`
Web: `www.dam.brown.edu/people/yiannis/`

and

S.P. Meyn[2]
Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
Email: `meyn@uiuc.edu` Web: `black.csl.uiuc.edu/∼meyn/`

Abstract

In this paper we continue the investigation of the spectral theory and exponential asymptotics of primarily discrete-time Markov processes, following Kontoyiannis and Meyn [32]. We introduce a new family of nonlinear Lyapunov drift criteria, which characterize distinct subclasses of geometrically ergodic Markov processes in terms of simple inequalities for the nonlinear generator. We concentrate primarily on the class of *multiplicatively regular* Markov processes, which are characterized via simple conditions similar to (but weaker than) those of Donsker-Varadhan. For any such process $\mathbf{\Phi} = \{\Phi(t)\}$ with transition kernel P on a general state space X, the following are obtained.

Spectral Theory: For a large class of (possibly unbounded) functionals $F : \mathsf{X} \to \mathbb{C}$, the kernel $\widehat{P}(x, dy) = e^{F(x)}P(x, dy)$ has a discrete spectrum in an appropriately defined Banach space. It follows that there exists a "maximal" solution $(\lambda, \check{f})$ to the *multiplicative Poisson equation*, defined as the eigenvalue problem $\widehat{P}\check{f} = \lambda\check{f}$. The functional $\Lambda(F) = \log(\lambda)$ is convex, smooth, and its convex dual Λ^* is convex, with compact sublevel sets.

Multiplicative Mean Ergodic Theorem: Consider the partial sums $\{S_t\}$ of the process with respect to any one of the functionals $F(\Phi(t))$ considered above. The normalized mean $\mathsf{E}_x[\exp(S_t)]$ (and not the logarithm of the mean) converges to $\check{f}(x)$ exponentially fast, where $\check{f}$ is the above solution of the multiplicative Poisson equation.

[1]I.K. was supported in part by a Sloan Foundation Research Fellowship and by NSF grant #0073378-CCR.
[2]S.M. was spported in part by NSF grant ECS 99-72957.

Multiplicative regularity: The Lyapunov drift criterion under which our results are derived is equivalent to the existence of regeneration times with finite exponential moments for the partial sums $\{S_t\}$, with respect to any functional F in the above class.

Large Deviations: The sequence of empirical measures of $\{\Phi(t)\}$ satisfies a large deviations principle in the "τ^{W_0}-topology," a topology finer that the usual τ-topology, generated by the above class of functionals F on X which is strictly larger than $L_\infty(\mathsf{X})$. The rate function of this LDP is Λ^*, and it is shown to coincide with the Donsker-Varadhan rate function in terms of relative entropy.

Exact Large Deviations Asymptotics: The above partial sums $\{S_t\}$ are shown to satisfy an exact large deviations expansion, analogous to that obtained by Bahadur and Ranga Rao for independent random variables.

Keywords: Markov process, large deviations, entropy, stochastic Lyapunov function, empirical measures, nonlinear generator, large deviations principle.

AMS 2000 Subject Classification: 60J10, 60J25, 37A30, 60F10, 47H99.

Submitted to EJP on December 16, 2003. Final version accepted on December 14, 2004.

Contents

1 Introduction and Main Results

Let $\boldsymbol{\Phi} = \{\Phi(t) : t \in \mathbb{T}\}$ be a Markov processes taking values in a Polish state space X, equipped with its associated Borel σ-field $\mathcal{B}$. The time index $\mathbb{T}$ may be discrete, $\mathbb{T} = \mathbb{Z}_+$, or continuous $\mathbb{T} = \mathbb{R}_+$, but we specialize to the discrete-parameter case after Section 1.1.

The distribution of $\boldsymbol{\Phi}$ is determined by its initial state $\Phi(0) = x \in \mathsf{X}$, and the transition semigroup $\{P^t : t \in \mathbb{T}\}$, where in discrete time all kernels P^t are powers of the 1-step transition kernel P. Throughout the paper we assume that $\boldsymbol{\Phi}$ is *ψ-irreducible* and *aperiodic*. This means that there is a σ-finite measure ψ on $(\mathsf{X}, \mathcal{B})$ such that, for any $A \in \mathcal{B}$ satisfying $\psi(A) > 0$ and any initial condition x,

$$P^t(x, A) > 0, \qquad \text{for all } t \text{ sufficiently large.}$$

Moreover, we assume that ψ is *maximal* in the sense that any other such ψ' is absolutely continuous with respect to ψ (written $\psi' \prec \psi$).

For a ψ-irreducible Markov process it is known that ergodicity is equivalent to the existence of a solution to the *Lyapunov drift criterion* (V3) below [34, 17]. Let $V : \mathsf{X} \to (0, \infty]$ be an extended-real valued function, with $V(x_0) < \infty$ for at least one $x_0 \in \mathsf{X}$, and write $\mathcal{A}$ for the (extended) generator of the semigroup $\{P^t : t \in \mathbb{T}\}$. This is equal to $\mathcal{A} = (P - I)$ in discrete time (where $I = I(x, dy)$ denotes the identity kernel $\delta_x(dy)$), and in continuous-time we think of $\mathcal{A}$ as a generalization of the classical differential generator $\mathcal{A} = \frac{d}{dt} P^t|_{t=0}$.

Recall that a function $s\colon \mathsf{X} \to \mathbb{R}_+$ and a probability measure ν on $(\mathsf{X}, \mathcal{B})$ are called *small* if for some measure m on $\mathbb{Z}$ with finite mean we have

$$\sum_{t \geq 0} P^t(x, A)\, m(t) \geq s(x)\nu(A), \qquad x \in \mathsf{X},\ A \in \mathcal{B}.$$

A set C is called *small* if $s = \epsilon \mathbb{I}_C$ is a small function for some $\epsilon > 0$. Also recall that an arbitrary kernel $\widehat{P} = \widehat{P}(x, dy)$ acts linearly on functions $f : \mathsf{X} \to \mathbb{C}$ and measures ν on $(\mathsf{X}, \mathcal{B})$, via

$$\widehat{P} f\,(\,\cdot\,) = \int_{\mathsf{X}} \widehat{P}(\cdot, dy) f(y) \quad \text{and} \quad \nu \widehat{P}\,(\,\cdot\,) = \int_{\mathsf{X}} \nu(dx) \widehat{P}(x, \,\cdot\,), \quad \text{respectively.} \tag{1}$$

We say that *the Lyapunov drift condition* (V3) *holds with respect to the Lyapunov function* V [34], if:

$$\left.\begin{array}{l}\text{For a function } W\colon \mathsf{X} \to [1, \infty), \text{ a small set } C \subset \mathsf{X}, \text{ and constants } \delta > 0,\ b < \infty, \\[1ex] \qquad \mathcal{A}V \leq -\delta W + b\mathbb{I}_C\,, \qquad \text{on } S_V := \{x : V(x) < \infty\}.\end{array}\right\} \quad \textbf{(V3)}$$

Condition (V3) implies that the set S_V is absorbing (and hence full), so that $V(x) < \infty$ a.e. $[\psi]$; see [34, Proposition 4.2.3].

As in [34, 32], a central role in our development will be played by weighted L_∞ spaces: For any function $W\colon \mathsf{X} \to (0, \infty]$, define the Banach space of complex-valued functions,

$$L_\infty^W := \Big\{ g\colon \mathsf{X} \to \mathbb{C} \ \text{s.t.}\ \sup_x \frac{|g(x)|}{W(x)} < \infty \Big\}, \tag{2}$$

with associated norm $\|g\|_W := \sup_x |g(x)|/W(x)$. We write $\mathcal{B}^+$ for the set of functions $s : \mathsf{X} \to [0,\infty]$ satisfying $\psi(s) := \int s(x)\,\psi(dx) > 0$, and, with a slight abuse of notation, we write $A \in \mathcal{B}^+$ if $A \in \mathcal{B}$ and $\psi(A) > 0$ (i.e., the indicator function $\mathbb{I}_A$ is in $\mathcal{B}^+$). Also, we let $\mathcal{M}_1^W$ denote the Banach space of signed and possibly complex-valued measures μ on $(\mathsf{X},\mathcal{B})$ satisfying $\|\mu\|_W := \sup_{F\in L_\infty^W} |\mu|(F) < \infty$.

The following consequences of (V3) may be found in [34, Theorem 14.0.1].

Theorem 1.1 (Ergodicity) *Suppose that $\mathbf{\Phi}$ is a ψ-irreducible and aperiodic discrete-time chain, and that condition* (V3) *is satisfied. Then the following properties hold:*

1. (W-ergodicity) *The process is positive recurrent with a unique invariant probability measure $\pi \in \mathcal{M}_1^W$ and for all $x \in S_V$,*

$$\begin{aligned} \sup_{F\in L_\infty^W} \Big|P^t(x,F) - \pi(F)\Big| &\to 0, && t\to\infty, \\ \frac{1}{T}\sum_{t=0}^{T-1} F(\Phi(t)) &\to \pi(F) := \int F(y)\,\pi(dy), && T\to\infty,\ a.s.\ [\mathsf{P}_x] \quad F\in L_\infty^W, \end{aligned}$$

where P_x denotes the conditional distribution of $\mathbf{\Phi}$ given $\Phi(0) = x$.

2. (W-regularity) *For any $A \in \mathcal{B}^+$ there exists $c = c(A) < \infty$ such that*

$$\mathsf{E}_x\Big[\sum_{t=0}^{\tau_A-1} W(\Phi(t))\Big] \le \delta^{-1}V(x) + c, \qquad x \in \mathsf{X}.$$

where E_x is the expectation with respect to P_x, and the hitting times τ_A are defined as,

$$\tau_A := \inf\{t \ge 1 \,:\, \Phi(t) \in A\}, \qquad A \in \mathcal{B}. \tag{3}$$

3. (Fundamental Kernel) *There exists a linear operator $Z\colon L_\infty^W \to L_\infty^{V+1}$, the* fundamental kernel*, such that*

$$\mathcal{A}ZF = -F + \pi(F), \qquad F \in L_\infty^W\,.$$

That is, the function $\widehat{F} := ZF$ solves the Poisson equation*, $\mathcal{A}\widehat{F} = -F + \pi(F)$.*

1.1 Multiplicative Ergodic Theory

The ergodic theory outlined in Theorem 1.1 is based upon consideration of the semigroup of linear operators $\{P^t\}$ acting on the Banach space L_∞^W. In particular, the ergodic behavior of the corresponding Markov process can be determined via the generator $\mathcal{A}$ of this semigroup. In this paper we show that the foundations of the *multiplicative* ergodic theory and of the large deviations behavior of $\mathbf{\Phi}$ can be developed in analogy to the linear theory, by shifting attention from the semigroup of linear operators $\{P^t\}$ to the family of nonlinear, convex operators $\{\mathcal{W}^t\}$ defined, for appropriate G, by

$$\mathcal{W}^t G\,(x) := \log\Big(\mathsf{E}_x[e^{G(\Phi(t))}]\Big), \qquad x \in \mathsf{X}\,,\ t \in \mathbb{T}\,.$$

Formally, we would like to define the 'generator' $\mathcal{H}$ associated with $\{\mathcal{W}^t\}$ by letting $\mathcal{H} = (\mathcal{W} - I)$ in discrete time and $\mathcal{H} = \frac{d}{dt}\mathcal{W}^t|_{t=0}$ in continuous time. Observing that $\mathcal{W}^t G = \log(P^t e^G)$, in discrete time we have

$$\mathcal{H}G = (\mathcal{W} - I)G = \log(Pe^G) - G = \log(e^{-G}Pe^G),$$

and in continuous time we can similarly calculate,

$$\mathcal{H}G \;=\; \lim_{t\to 0} \frac{1}{t}\big[\mathcal{W}^t - I\big]G \;=\; \lim_{t\to 0}\frac{1}{t}\log\big(e^{-G}P^t e^G\big) \;=\; e^{-G}\mathcal{A}e^G\,,$$

whenever all the above limits exist. Rather than assume differentiability, we use these expressions as motivation for the following rigorous definition of the *nonlinear generator*,

$$\mathcal{H}(G) = \begin{cases} \log(e^{-G}Pe^G) & \text{discrete time } (\mathbb{T} = \mathbb{Z}_+); \\ e^{-G}\mathcal{A}e^G & \text{continuous time } (\mathbb{T} = \mathbb{R}_+), \end{cases} \tag{4}$$

when e^G is in the domain of the extended generator. In continuous time, this is Fleming's nonlinear generator; see [22] for a starting point, and [20, 21] for recent surveys.

In this paper our main focus will be on the following 'multiplicative' analog of (V3), where the role of the generator is now played by the nonlinear generator $\mathcal{H}$. We say that the *Lyapunov drift criterion* (DV3) *holds with respect to the Lyapunov function* $V : \mathsf{X} \to (0,\infty]$, if:

$$\left.\begin{array}{c} \text{For a function } W\colon \mathsf{X} \to [1,\infty), \text{ a small set } C \subset \mathsf{X}, \text{ and constants } \delta > 0,\ b < \infty, \\[1em] \mathcal{H}(V) \le -\delta W + b\mathbb{I}_C\,, \qquad \text{on } S_V. \end{array}\right\} \quad \textbf{(DV3)}$$

[This condition was introduced in [32], under the name (mV3).] Under either condition (V3) or (DV3), we let $\{C_W(r)\}$ denote the sublevel sets of W:

$$C_W(r) = \{y : W(y) \le r\}, \qquad r \in \mathbb{R}. \tag{5}$$

The main assumption in many of our results below will be that $\boldsymbol{\Phi}$ satisfies (DV3), and also that the transition kernels satisfy a mild continuity condition: We require that they possess a density with respect to some reference measure, uniformly over all initial conditions x in the sublevel set $C_W(r)$ of W. These assumptions are formalized in condition (DV3+) below.

(DV3+)

(i) The Markov process $\boldsymbol{\Phi}$ is ψ-irreducible, aperiodic, and it satisfies condition (DV3) with some Lyapunov function $V : \mathsf{X} \to [1,\infty)$;

(ii) There exists $T_0 > 0$ such that, for each $r < \|W\|_\infty$, there is a measure β_r with $\beta_r(e^V) < \infty$ and $\mathsf{P}_x\{\Phi(T_0) \in A,\ \tau_{C_W^c(r)} > T_0\} \le \beta_r(A)$ for all $x \in C_W(r)$, $A \in \mathcal{B}$.

Condition (DV3+) captures the essential ingredients of the large deviations conditions imposed by Donsker and Varadhan in their pioneering work [14, 15, 16], and is in fact somewhat weaker than those conditions. In Section 2 an extensive discussion of this assumption is given, its relation to several well-known conditions in the literature is described in detail. In particular, part (ii) of condition (DV3+) [to which we will often refer as the "density assumption" in (DV3+)] is generally the weaker of the two assumptions.

In most of our results we assume that the function W in (DV3) is unbounded, $\|W\|_\infty := \sup_x |W(x)| = \infty$. When this is the case, we let $W_0 : \mathsf{X} \to [1,\infty)$ be a fixed function in L_∞^W, whose growth at infinity is strictly slower than W in the sense that

$$\lim_{r\to\infty} \sup_{x\in\mathsf{X}} \Big[\frac{W_0(x)}{W(x)}\, \mathbb{I}_{\{W(x)>r\}}\Big] = 0\,. \tag{6}$$

Below we collect, from various parts of the paper, the "multiplicative" ergodic results we derive from (DV3+), in analogy to the "linear" ergodic-theoretic results stated in Theorem 1.1.

Theorem 1.2 (Multiplicative Ergodicity) *Suppose that the discrete-time chain $\boldsymbol{\Phi}$ satisfies condition* (DV3+) *with W unbounded, and let $W_0 \in L_\infty^W$ be as in* (6)*. Then the following properties hold:*

1. (W-multiplicative ergodicity) *The process is positive recurrent with a unique invariant probability measure π satisfying, for some $\eta > 0$,*

$$\pi(e^{\eta V}) < \infty \text{ and } \quad \pi(e^{\eta W}) < \infty.$$

For any real-valued $F \in L_\infty^{W_0}$, there exist $\check{F} \in L_\infty^V$, $\Lambda(F) \in \mathbb{C}$, and constants $b_0 > 0$, $B_0 < \infty$, such that

$$\left| \mathsf{E}_x \left[\exp\Big(\sum_{t=0}^{T-1} [F(\Phi(t))] - \Lambda(F)] \Big) \right] - e^{\check{F}(x)} \right| \le e^{\eta V(x) + B_0 - b_0 T}\,, \tag{7}$$

for all $T \ge 1,\ x \in \mathsf{X}$.

2. (W-multiplicative regularity) *For any $A \in \mathcal{B}^+$ there exist constants $\eta = \eta(A) > 0$ and $c = c(A) < \infty$, such that*

$$\log\Big(\mathsf{E}_x \Big[\exp\Big(\eta \sum_{t=0}^{\tau_A - 1} W(\Phi(t)) \Big) \Big] \Big) \le V(x) + c, \qquad x \in \mathsf{X}.$$

3. (Multiplicative Fundamental 'Kernel') *There exists a nonlinear operator $\mathcal{G}\colon L_\infty^{W_0} \to L_\infty^V$, the* multiplicative fundamental kernel*, such that the function $\check{F}$ in (1.) can be expressed as $\check{F} = \mathcal{G}(F)$ for real-valued $F \in L_\infty^{W_0}$, and $\check{F}$ solves the* multiplicative Poisson equation*,*

$$\mathcal{H}(\check{F}) = -F + \Lambda(F)\,. \tag{8}$$

Proof. Assumption (DV3) combined with Theorem 2.2 implies that $\boldsymbol{\Phi}$ is geometrically ergodic (equivalently, V_0-uniformly ergodic) for some Lyapunov function $V_0\colon \mathsf{X} \to [1,\infty)$, hence the process is also positive recurrent. Moreover, $v_{\eta_0} := e^{\eta_0 V} \in L_\infty^{V_0}$ for some $0 < \eta_0 < 1$. By the geometric ergodic theorem of [34] it follows that $\pi(v_{\eta_0}) < \infty$.

Under (DV3), the stochastic process $\mathbf{m} = \{m(t)\}$ defined below is a super-martingale with respect to $\mathcal{F}_t = \sigma\{\Phi(s) : 0 \le s \le t\}$, $t \ge 0$,

$$m(t) := \exp\Big(V(\Phi(t)) + \sum_{s=0}^{t-1}[\delta W(\Phi(s)) - b\mathbb{I}_C(\Phi(s))]\Big), \quad t \ge 0\,. \tag{9}$$

From the super-martingale property and Jensen's inequality we obtain the bound,

$$\mathsf{E}_x\Big[\exp\Big(\eta_0 V(\Phi(t)) - \eta_0 b + \sum_{s=0}^{t-1}\eta_0\delta W(\Phi(s))\Big)\Big] < v_{\eta_0}(x)\,, \qquad x \in \mathsf{X}.$$

which gives the desired bound in (1.), where $\eta := \delta\eta_0$. The multiplicative ergodic limit (7) follows from Theorem 3.1 (iii). The existence of an inverse $\mathcal{G}$ to $\mathcal{H}$ is given in Proposition 3.6, which establishes the bound $\check{F} \in L_\infty^V$ stated in (1.), as well as result (3.).

Theorem 2.5 shows that (DV3) actually characterizes W-multiplicative regularity, and provides the bound in (2.). $\square$

As in [32], central to our development is the observation that the multiplicative Poisson equation (8) can be written as an eigenvalue problem. In discrete-time with $\Lambda = \Lambda(F)$, (8) becomes $(e^F P)e^{\check{F}} = e^\Lambda e^{\check{F}}$, or, writing $f = e^F$, $\check{f} = e^{\check{F}}$ and $\lambda = e^\Lambda$, we obtain the eigenvalue equation,

$$P_f\check{f} = \lambda\check{f}, \qquad \text{for the kernel } P_f(x,dy) := f(x)P(x,dy).$$

The assumptions of Theorem 1.2 are most easily illustrated in continuous time. Consider the following diffusion model on $\mathbb{R}$, sometimes referred to as the *Smoluchowski equation*. For a given potential $u : \mathbb{R} \to \mathbb{R}_+$, this is defined by the stochastic differential equation

$$dX(t) = -u_x(X(t))\,dt + \sigma dW(t)\,, \tag{10}$$

where $u_x := \frac{d}{dx}u$, and $\boldsymbol{W} = \{W(t) : t \ge 0\}$ is a standard Brownian motion. On C^2, the extended generator $\mathcal{A}$ of $\boldsymbol{X} = \{X(t) : t \ge 0\}$ coincides with the differential generator given by,

$$\mathcal{A} = \tfrac{1}{2}\sigma^2\frac{d^2}{dx^2} - u_x\frac{d}{dx}\,. \tag{11}$$

When $\sigma > 0$ this is an *elliptic* diffusion, so that the semigroup $\{P^t\}$ has a family of smooth, positive densities $P^t(x,dy) = p(x,y;t)dy$, $x,y \in \mathbb{R}$ [33]. Hence the Markov process $\boldsymbol{X}$ is ψ-irreducible, with ψ equal to Lebesgue measure on $\mathbb{R}$.

A special case is the one-dimensional Ornstein-Uhlenbeck process,

$$dX(t) = -\delta X(t)\,dt + \sigma dW(t)\,, \tag{12}$$

where the corresponding potential function is $u(x) = \frac{1}{2}\delta x^2$, $x \in \mathbb{R}$.

Proposition 1.3 *The Smoluchowski equation satisfies* (DV3+) *with $V = 1 + u\sigma^{-2}$ and $W = 1 + u_x^2$, provided the potential function $u\colon \mathbb{R} \to \mathbb{R}_+$ is C^2 and satisfies:*

(a) $\displaystyle\lim_{|x|\to\infty} u(x) = \infty$;

(b) $\lim_{|x|\to\infty} \frac{(u_x(x))^2}{|u_{xx}(x)|} = \infty, \qquad \liminf_{|x|\to\infty} (u_x(x))^2 > 0.$

PROOF. Let $V = 1 + u\sigma^{-2}$. We then have,

$$\begin{aligned}\mathcal{H}(V) := e^{-V}\mathcal{A}e^{V} &= e^{-V}\Big\{-u_x\Big(e^V\sigma^{-2}u_x\Big) + \tfrac{1}{2}\sigma^2\Big(e^V[u_{xx}\sigma^{-2} + \sigma^{-4}u_x^2]\Big)\Big\}\\ &= -\tfrac{1}{2}\sigma^{-2}u_x^2 + \tfrac{1}{2}u_{xx}\,.\end{aligned}$$

It is thus clear that the desired drift conditions hold. The proof is complete since $P^t(x, dy)$ possesses a continuous density $p(x, y; t)$ for each $t > 0$: We may take $T_0 = 1$, and for each r we take β_r equal to a constant times Lebesgue measure on $C_W(r)$. □

Proposition 1.3 does not admit an exact generalization to discrete-time models. However, the discrete-time one-dimensional Ornstein-Uhlenbeck process,

$$X(t+1) - X(t) = -\delta X(t) + W(t+1)\,, \qquad t \ge 0,\ X(0) \in \mathbb{R}, \tag{13}$$

does satisfy the conclusions of the proposition, again with $V = 1 + \epsilon_0 x^2$ for some $\epsilon_0 > 0$, when $\delta > 0$ and $\boldsymbol{W}$ is an i.i.d. Gaussian process with positive variance.

Notation. Often in the transition from ergodic results to their multiplicative counterparts we have to take exponentials of the corresponding quantities. In order to make this correspondence transparent we have tried throughout the paper to follow, as consistently as possible, the convention that the exponential version of a quantity is written as the corresponding lower case letter. For example, above we already had $f = e^F$, $\check{f} = e^{\check{F}}$ and $\lambda = e^{\Lambda}$.

1.2 Large Deviations

From now on we restrict attention to the discrete-time case.

Part 1 of Theorem 1.2 extends the *multiplicative mean ergodic theorem* of [32] to the larger class of (possibly unbounded) functionals $F \in L_\infty^{W_0}$. In this section we assume that (DV3+) holds with an unbounded function W, and we let a function $W_0 \in L_\infty^W$ be chosen as in (6).

For $n \ge 1$, let L_n denote the empirical measures induced by $\boldsymbol{\Phi}$ on $(\mathsf{X}, \mathcal{B})$,

$$L_n := \frac{1}{n}\sum_{t=0}^{n-1} \delta_{\Phi(t)} \qquad n \ge 1\,, \tag{14}$$

and write $\langle\cdot,\cdot\rangle$ for the usual inner product; for μ a measure and G a function, $\langle\mu, G\rangle = \mu(G) := \int G(y)\,\mu(dy)$, whenever the integral exists. Then, from Theorem 3.1 it follows that for any real-valued $F \in L_\infty^{W_0}$ and any $a \in \mathbb{R}$ we have the following version of the multiplicative mean ergodic theorem,

$$\exp\Big(-n\Lambda(aF)\Big)\mathsf{E}_x\Big[\exp\Big(an\langle L_n, F\rangle\Big)\Big] \;\to\; \check{f}_a(x)\,, \qquad n\to\infty,\ x \in \mathsf{X}\,, \tag{15}$$

where $\check{f}_a := e^{\mathcal{G}(aF)}$ is the eigenfunction constructed in part 3 of Theorem 1.2, corresponding to the function aF.

In Section 5, strong large deviations results for the sequence of empirical measures $\{L_n\}$ are derived from the multiplicative mean ergodic theorem in (15), using standard techniques [9, 7, 12]. First we show that, for any initial condition $x \in \mathsf{X}$, the sequence $\{L_n\}$ satisfies a large deviations principle (LDP) in the space $\mathcal{M}_1$ of all probability measures on $(\mathsf{X}, \mathcal{B})$ equipped with the τ^{W_0}*-topology*, that is, the topology generated by the system of neighborhoods

$$N_F(c, \epsilon) := \big\{\nu \in \mathcal{M}_1 : |\nu(F) - c| < \epsilon\big\}, \quad \text{for real-valued } F \in L_\infty^{W_0},\ c \in \mathbb{R},\ \epsilon > 0\,. \tag{16}$$

Moreover, the rate function $I(\nu)$ that governs this LDP is the same as the Donsker-Varadhan rate function, and can be characterized in terms of relative entropy,

$$I(\nu) := \inf H(\nu \odot \check{P} \| \nu \odot P)\,,$$

where the infimum is over all transition kernels $\check{P}$ for which ν is an invariant measure, $\nu \odot \check{P}$ denotes the bivariate measure $[\nu \odot \check{P}](dx, dy) := \nu(dx)\check{P}(x, dy)$ on $(\mathsf{X} \times \mathsf{X}, \mathcal{B} \times \mathcal{B})$, and $H(\,\cdot\,\|\,\cdot\,)$ denotes the relative entropy,

$$H(\mu\|\nu) = \begin{cases} \int d\mu \log \frac{d\mu}{d\nu}, & \text{when} \frac{d\mu}{d\nu} \text{ exists} \\ \infty, & \text{otherwise.} \end{cases} \tag{17}$$

[Throughout the paper we follow the usual convention that the infimum of the empty set is $+\infty$.] As we discuss in Section 2.6 and Section 5, the density assumption in (DV3+) (ii) is weaker than the continuity assumptions of Donsker and Varadhan, but it cannot be removed entirely.

Further, the precise convergence in (15) leads to exact large deviations expansions analogous to those obtained by Bahadur and Ranga Rao [1] for independent random variables, and to the local expansions established in [32] for geometrically ergodic chains. For real-valued, non-lattice functionals $F \in L_\infty^{W_0}$, in Theorem 5.3 we obtain the following: For $c > \pi(F)$ and $x \in \mathsf{X}$,

$$\mathsf{P}_x\Big\{\sum_{t=0}^{n-1} F(\Phi(t)) \ge nc\Big\} \ \sim\ \frac{\check{f}_a(x)}{a\sqrt{2\pi n\sigma_a^2}} e^{-nJ(c)}, \quad n \to \infty, \tag{18}$$

where $a \in \mathbb{R}$ is chosen such that $\frac{d}{da}\Lambda(aF) = c$, $\check{f}_a(x)$ is the eigenfunction appearing in the multiplicative mean ergodic theorem (15), $\sigma_a^2 = \frac{d^2}{da^2}\Lambda(aF)$, and the exponent $J(c)$ is given in terms of $I(\nu)$ as

$$J(c) := \inf\big\{I(\nu) : \nu \text{ is a probability measure on } (\mathsf{X}, \mathcal{B}) \text{ satisfying } \nu(F) \ge c\big\}\,. \tag{19}$$

A corresponding expansion is given for lattice functionals.

These large deviations results extend the classical Donsker-Varadhan LDP [14, 15] in several directions: First, our conditions are weaker. Second, when (DV3+) holds with an unbounded function W, the τ^{W_0}-topology is finer and hence stronger than either the topology of weak convergence, or the τ-topology, with respect to which the LDP for the empirical measures $\{L_n\}$ is usually established [24, 4, 13]. Third, apart from the LDP we also obtain precise large deviations expansions as in (18) for the partial sums with respect to (possibly unbounded) functionals $F \in L_\infty^{W_0}$.

Following the Donsker-Varadhan papers, a large amount of work has been done in establishing large deviations properties of Markov chains under a variety of different assumptions; see [12, 13] for detailed treatments. Under conditions similar to those in this paper, Ney and Nummelin have proved "pinned" large deviations principles in [37, 38]. In a different vein, under much weaker assumptions (essentially under irreducibility alone) de Acosta [10] and Jain [28] have proved general large deviations lower bounds, but these are, in general, not tight.

One of the first places where the Feller continuity assumption of Donsker and Varadhan was relaxed is Bolthausen's work [4]. There, a very stringent condition on the chain is imposed, often referred to in the literature as Stroock's uniform condition (U). In Section 2.5 we argue that (U) is much more restrictive than the conditions we impose in this paper. In particular, condition (U) implies Doeblin recurrence as well as the density assumption in (DV3+) (ii).

More recently, Eichelsbacher and Schmock [19] proved an LDP for the empirical measures of Markov chains, again under the uniform condition (U). This LDP is proved in a strict subset of $\mathcal{M}_1$, and with respect to a topology finer than the usual τ-topology and similar in spirit to the τ^{W_0} topology introduced here. In addition to (U), the results of [19] require strong integrability conditions that are *a priori* hard to verify: In the above notation, in [19] it is assumed that for at least one unbounded function $W_0 : \mathsf{X} \to \mathbb{R}$, we have $\mathsf{E}_x[\exp\{a|W_0(\Phi(n))|\}] < \infty$, uniformly over $n \geq 1$, for all real $a > 0$. This assumption is closely related to our condition (DV3), and, as we show in Section 3, (DV3) in particular provides a means for identifying a natural class of functions W_0 satisfying this bound.

2 Structural Assumptions

There is a wide range of interrelated tools that have been used to establish large deviations properties for Markov processes and to develop parts of the corresponding multiplicative ergodic theory. Most of these tools rely on a functional-analytic setting within which spectral properties of the process are examined. A brief survey of these approaches is given in [32], where the main results relied on the *geometric ergodicity* of the process. In this section we show how the assumptions used in prior work may be expressed in terms of the drift criteria introduced here and describe the operator-theoretic setting upon which all our subsequent results will be based.

2.1 Drift Conditions

Recall that the *(extended) generator* $\mathcal{A}$ of $\mathbf{\Phi}$ is defined as follows: For a function $g : \mathsf{X} \to \mathbb{C}$, we write $\mathcal{A}g = h$ if for each initial condition $\Phi(0) = x \in \mathsf{X}$ the process $\ell(t) := \sum_{s=0}^{t-1} h(\Phi(s)) - g(\Phi(t))$, $t \geq 1$, is a local martingale with respect to the natural filtration $\{\mathcal{F}_t = \sigma(\Phi(s), 0 \leq s \leq t) : t \geq 1\}$. In discrete time, the extended generator is simply $\mathcal{A} = P - I$, and its domain contains all measurable functions on X.

The following drift conditions are considered in [34] in discrete time,

$$
\begin{array}{llcl}
\textbf{(V2)} & \mathcal{A}V & \leq & -\delta + b\mathbb{I}_C \\
\textbf{(V3)} & \mathcal{A}V & \leq & -\delta W + b\mathbb{I}_C \\
\textbf{(V4)} & \mathcal{A}V & \leq & -\delta V + b\mathbb{I}_C \,,
\end{array}
$$

where in each case C is small, $V \colon \mathsf{X} \to (0,\infty]$ is finite a.e. $[\psi]$, and $b < \infty$, $\delta > 0$ are constants. We further assume that W is bounded below by unity in (V3), and that V is bounded from below by unity in (V4). It is easy to see that (V2)–(V4) are stated in order of increasing strength: (V4) $\Rightarrow$ (V3) $\Rightarrow$ (V2).

Analogous multiplicative versions of these drift criteria are defined as follows,

$$\begin{array}{lrcl} \textbf{(DV2)} & \mathcal{H}V & \leq & -\delta + b\mathbb{I}_C \\ \textbf{(DV3)} & \mathcal{H}V & \leq & -\delta W + b\mathbb{I}_C \\ \textbf{(DV4)} & \mathcal{H}V & \leq & -\delta V + b\mathbb{I}_C \,, \end{array}$$

where $\mathcal{H}$ is the nonlinear generator defined in (4). The following implications follow easily from the definitions:

Proposition 2.1 *For each $k = 2,3,4$, the drift condition* (DVk) *implies* (Vk).

PROOF. We provide a proof only for $k = 3$ since all are similar. Under (DV3), $Pe^V \leq e^{V-W+b\mathbb{I}_C}$. Jensen's inequality gives $e^{PV} \leq Pe^V$, and taking logarithms gives (V3). $\square$

We find that Proposition 2.1 gives a poor bound in general. Theorem 2.2 shows that (DV2) actually implies (V4). Its proof is given in the Appendix, after the proof of Theorem 2.5.

Theorem 2.2 ((DV2) $\Rightarrow$ (V4)) *Suppose $\mathbf{\Phi}$ is ψ-irreducible and aperiodic. If* (DV2) *holds for some $V \colon \mathsf{X} \to (0,\infty]$, then* (V4) *holds for some V_0 which is* equivalent *to $v_\eta := e^{\eta V}$ for some $\eta > 0$, in the sense that,*

$$V_0 \in L_\infty^{v_\eta} \quad \textit{and} \quad v_\eta \in L_\infty^{V_0} \,.$$

2.2 Spectral Theory Without Reversibility

The spectral theory described in this paper and in [32] is based on various operator semigroups $\{\widehat{P}^n : n \in \mathbb{Z}_+\}$, where each $\widehat{P}^n$ is the nth composition of a possibly non-positive kernel $\widehat{P}$. Examples are the transition kernel P; the multiplication kernel $I_G(x,dy) = G(x)\delta_x(dy)$. for a given function G; the *scaled kernel* defined by

$$P_f(x,dy) := f(x)P(x,dy)\,, \tag{20}$$

for any function $F \colon \mathsf{X} \to \mathbb{C}$ with $f = e^F$; and also the *twisted kernel*, defined for a given function $h \colon \mathsf{X} \to (0,\infty)$ by

$$\check{P}_h(x,A) := [I_{Ph}^{-1} P I_h]\,(x,A) = \frac{\int_A P(x,dy)h(y)}{Ph(x)} \qquad x \in \mathsf{X},\ A \in \mathcal{B}. \tag{21}$$

This is a *probabilistic* kernel (i.e., a positive kernel with $\check{P}_h(x,\mathsf{X}) = 1$ for all x) provided $Ph\,(x) < \infty$, $x \in \mathsf{X}$. It is a generalization of the twisted kernel considered in [32], where the function h was taken as $h = \check{f}$ for a specially constructed $\check{f}$. It may also be regarded as a version of Doob's h-transform [40].

The most common approach to spectral decompositions for probabilistic semigroups $\{P^n\}$ is to impose a reversibility condition [23, 5, 41]. The motivation for this assumption comes from the L_2 setting in which these problems are typically posed, and the well-known fact that the semigroup $\{P^n\}$ is then self-adjoint. We avoid a Hilbert space setting here and instead consider the weighted L_∞ function spaces defined in (2); cf. [30, 31, 25, 35, 32].

The weighting function is determined by the particular drift condition satisfied by the process. In particular, under (DV3) it follows from the convexity of $\mathcal{H}$ (see Proposition 4.4) that for any $0 < \eta \le 1$ we have the bound,

$$\mathcal{H}(\eta V) \le -\delta\eta W + b\eta\mathbb{I}_C\,, \qquad \text{on } S_V\,, \tag{22}$$

which may be equivalently expressed as $Pv_\eta \le e^{\eta[-\delta W + b\mathbb{I}_C]}v_\eta$, where $v_\eta := e^{\eta V}$. This bound implies that $P_f\colon L_\infty^{v_\eta} \to L_\infty^{v_\eta}$ is a bounded linear operator for any function f satisfying $\|F^+\|_W \le \eta\delta$ (where $F^+ := \max(F, 0)$), and any $0 \le \eta \le 1$.

Under any one of the above Lyapunov drift criteria, we will usually consider the function v defined in terms of the corresponding Lyapunov function V on X via $v = e^V$. For any such function $v\colon \mathsf{X} \to [1, \infty)$ and any linear operator $\widehat{P}\colon L_\infty^v \to L_\infty^v$, we denote the induced operator norm by,

$$|||\widehat{P}|||_v := \sup\Big\{\frac{\|\widehat{P}h\|_v}{\|h\|_v} : h \in L_\infty^v,\ \|h\|_v \ne 0\Big\}. \tag{23}$$

The *spectrum* $\mathcal{S}(\widehat{P}) \subset \mathbb{C}$ of $\widehat{P}$ 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_∞^v. We let $\xi = \xi(\{\widehat{P}^n\})$ denote the *spectral radius* of the semigroup $\{\widehat{P}^n\}$,

$$\xi(\{\widehat{P}^n\}) := \lim_{n\to\infty} |||\widehat{P}^n|||_v^{1/n}. \tag{24}$$

In general, the quantities $|||\widehat{P}|||_v$ and ξ depend upon the particular weighting function v. If $\widehat{P}$ is a positive operator, then ξ is greater than or equal to the *generalized principal eigenvalue*, or *g.p.e.* (see e.g. [39]), and they are actually equal under suitable regularity assumptions (see [2, 32], and Proposition 2.8 below).

As in [32], we say that $\widehat{P}$ admits a *spectral gap* if there exists $\epsilon_0 > 0$ such that the set $\mathcal{S}(\widehat{P}) \cap \{z : |z| \ge \xi - \epsilon_0\}$ is finite and contains only poles of finite multiplicity; recall that $z_0 \in \mathcal{S}(\widehat{P})$ is a *pole of (finite) multiplicity* n if:

(i) z_0 is isolated in $\mathcal{S}(\widehat{P})$, i.e., for some $\epsilon_1 > 0$ we have $\{z \in \mathcal{S}(\widehat{P}) : |z - z_0| \le \epsilon_1\} = \{z_0\}$;

(ii) The associated projection operator

$$\widehat{Q} := \frac{1}{2\pi i}\int_{\partial\{z:|z-z_0|\le\epsilon_1\}} [Iz - \widehat{P}]^{-1} dz\,, \tag{25}$$

can be expressed as a finite linear combination of some $\{s_i\} \subset L_\infty^v$, $\{\nu_i\} \subset \mathcal{M}_1^v$,

$$\widehat{Q} = \sum_{i,j=0}^{n-1} m_{i,j}[s_i \otimes \nu_j]\,,$$

where $[s \otimes \nu](x, dy) := s(x)\nu(dy)$.

See [32, Sec. 4] for more details. Moreover, we say that $\widehat{P}$ is *v-uniform* if it admits a spectral gap and also there exists a unique pole $\lambda_\circ \in \mathcal{S}(\widehat{P})$ of multiplicity one, satisfying $|\lambda_\circ| = \xi(\{\widehat{P}^t\})$.

Recall that a Markov process $\mathbf{\Phi}$ is called *geometrically ergodic* [32] or equivalently *V-uniformly ergodic* [34] if it is positive recurrent, and the semigroup converges in the induced operator norm,

$$|\!|\!| P^n - \mathbf{1} \otimes \pi |\!|\!|_V \to 0, \qquad n \to \infty\,,$$

where $\mathbf{1}$ denotes the constant function $\mathbf{1}(x) \equiv 1$. It is known that this is characterized by condition (V4). Under this assumption, in [32] we proved that $\mathbf{\Phi}$ satisfies a "local" large deviations principle. In this paper under the stronger condition (DV3+) we show that these local results can be extended to a full large deviations principle.

The following result, taken from [32, Proposition 4.6], says that geometric ergodicity is equivalent to the existence of a spectral gap:

Theorem 2.3 (Spectral Gap & (V4)) *Let $\mathbf{\Phi}$ by a ψ-irreducible and aperiodic Markov chain.*

(*a*) *If $\mathbf{\Phi}$ is geometrically ergodic with Lyapunov function V, then its transition kernel P admits a spectral gap in L_∞^V and it is V-uniform.*

(*b*) *Conversely, if P is V_0-uniform, then $\mathbf{\Phi}$ is geometrically ergodic with respect to some Lyapunov function $V \in L_\infty^{V_0}$.*

Next we want to investigate the corresponding relationship between condition (DV3) and when the kernel P has a discrete spectrum in L_∞^v. First we establish an analogous 'near equivalence' between assumption (DV3) and the notion of v-separability, and in Theorem 3.5 we show that v-separability implies the discrete spectrum property.

For any $v\colon \mathsf{X} \to [1,\infty]$, finite a.e. $[\psi]$, we say that the linear operator $\widehat{P}\colon L_\infty^v \to L_\infty^v$ is *v-separable* if it can be approximated uniformly by kernels with finite-rank. That is, for each $\epsilon > 0$, there exists a finite-rank operator $\widehat{K}_\epsilon$ such that $|\!|\!| \widehat{P} - \widehat{K}_\epsilon |\!|\!|_v \le \epsilon$. Since the kernel $\widehat{K}_\epsilon$ has a finite-dimensional range space, we are assured of the existence of an integer $n \ge 1$, functions $\{s_i : 1 \le i \le n\} \subset L_\infty^v$, and probability measures $\{\nu_i : 1 \le i \le n\} \subset \mathcal{M}_1^v$, such that $\widehat{K}_\epsilon$ may be expressed,

$$\widehat{K}_\epsilon(x, dy) = \sum_{i=1}^n s_i \otimes \nu_i\,. \tag{26}$$

Note that the eigenvalues of $\widehat{K}_\epsilon$ may be interpreted as a *pseudo-spectrum*; see [8].

The following equivalence, established in the Appendix, illustrates the intimate relationship between the essential ingredients of the Donsker-Varadhan conditions, and the associated spectral theory as developed in this paper. Note that in Theorem 2.4 the density assumption from part (ii) of (DV3+) has been replaced by the more natural and weaker statement that $I_{C_W(r)} P^{T_0}$ is v-separable for all r.[3] The fact that this is indeed weaker than the assumption in (DV3) (ii) follows from Lemma B.3 in the Appendix. Applications of Theorem 2.4 to diffusions on $\mathbb{R}^n$ and refinements in this special case are developed in [26].

Theorem 2.4 (v-Separability & (DV3)) *Let $\mathbf{\Phi}$ be a ψ-irreducible and aperiodic Markov chain and let $T_0 > 0$ arbitrary. The following are equivalent:*

[3] The notation $I_A \widehat{P}$ for a set $A \in \mathcal{B}$ and a kernel $\widehat{P}$ is used to denote the kernel $\mathbb{I}_A(x)\widehat{P}(x, dy)$.

(*a*) *Condition* (DV3) *holds with* $V\colon \mathsf{X} \to [1,\infty)$*;* W *unbounded; and* $I_{C_W(r)}P^{T_0}$ *is* v*-separable for all* r*, where* $v = e^V$*.*

(*b*) *The kernel* P^{T_0} *is* v_0*-separable for some unbounded function* $v_0\colon \mathsf{X} \to [1,\infty)$*.*

We say that a linear operator $\widehat{P} : L_\infty^v \to L_\infty^v$ has a *discrete spectrum in* L_∞^v if its spectrum $\mathcal{S}$ has the property that $\mathcal{S} \cap K$ is finite, and contains only poles of finite multiplicity, for any compact set $K \subset \mathbb{C} \setminus \{0\}$. It is shown in Theorem 3.5 that the spectrum of P is discrete under the conditions of (b) above.

Taking a different operator-theoretic approach, Deuschel and Stroock [13] prove large deviations results for the empirical measures of stationary Markov chains under the condition of hypercontractivity (or hypermixing). In particular, their conditions imply that for some T_0, the kernel $P^{T_0}(x,dy)$ is a bounded linear operator from $L_2(\pi)$ to $L_4(\pi)$, with norm equal to 1.

2.3 Multiplicative Regularity

Recall the definition of the empirical measures in (14), and the hitting times $\{\tau_A\}$ defined in (3). The next set of results characterize the drift criterion (DV3) in terms of the following regularity assumptions:

Regularity

(i) A set $C \in \mathcal{B}$ is called *geometrically regular* if for any $A \in \mathcal{B}^+$ there exists $\eta = \eta(A) > 0$ such that

$$\sup_{x\in C} \mathsf{E}_x[\exp(\eta\tau_A)] < \infty.$$

The Markov process $\mathbf{\Phi}$ is called *geometrically regular* if there exists a geometrically regular set C, and $\eta > 0$ such that

$$\mathsf{E}_x[\exp(\eta\tau_C)] < \infty, \qquad x \in \mathsf{X}.$$

(ii) A set $C \in \mathcal{B}$ is called H*-multiplicatively regular* (H-m.-regular) if for any $A \in \mathcal{B}^+$, there exists $\eta = \eta(A) > 0$ satisfying,

$$\sup_{x\in C} \mathsf{E}_x\Bigl[\exp(\eta\tau_A\langle L_{\tau_A}, H\rangle)\Bigr] < \infty.$$

The Markov process $\mathbf{\Phi}$ is H*-m.-regular* if there exists an H-m.-regular set $C \in \mathcal{B}$, and $\eta > 0$ such that

$$\mathsf{E}_x\Bigl[\exp(\eta\tau_C\langle L_{\tau_C}, H\rangle)\Bigr] < \infty\,, \qquad x \in \mathsf{X}\,.$$

In [34, Theorem 15.0.1] a precise equivalence is given between geometric regularity and the existence of a solution to the drift inequality (V4). The following analogous result shows that (DV3) characterizes multiplicative regularity. A proof of Theorem 2.5 is included in the Appendix.

Theorem 2.5 (Multiplicative Regularity $\Leftrightarrow$ (DV3)) *For any $H\colon \mathsf{X} \to [1,\infty)$, the following are equivalent:*

(i) *$\mathbf{\Phi}$ is H-m.-regular;*

(ii) *The drift inequality* (DV3) *holds for some $V : \mathsf{X} \to (0,\infty)$ and with $H \in L_\infty^W$.*

If either of these equivalent conditions hold, then for any $A \in \mathcal{B}^+$, there exists $\epsilon > 0$, $1 \geq \eta > 0$, and $B < \infty$ satisfying,

$$\mathsf{E}_x\Big[\exp\Big(\epsilon\tau_A\langle L_{\tau_A}, H\rangle + \eta V(\Phi(\tau_A))\Big)\Big] \leq \exp(\eta V(x) + B), \qquad x \in \mathsf{X},$$

where V is the solution to (DV3) *in* (ii).

In a similar vein, in [44] the following condition is imposed for a diffusion on $\mathsf{X} = \mathbb{R}^n$:

$$\begin{array}{l}\textit{For any } n \geq 1 \textit{ there exists } K_n \subset \mathsf{X} \textit{ compact, such that for any} \\ \textit{compact set } K \subset \mathsf{X}, \\ \qquad\qquad\qquad \sup_{x\in K} \mathsf{E}_x[e^{n\tau_{K_n}}] < \infty.\end{array} \tag{27}$$

In [44, 42] it is shown that this condition is closely related to the existence of a solution to (DV3), where the function W is further assumed to have compact sublevel sets. Under these assumptions, and under continuity assumptions similar to those imposed in [43], it is possible to show that the operator P^n is *compact* for all $n > 0$ [42, Theorem 2.1], or [11, Lemma 3.4].

We show in Proposition 2.6 that the bound assumed in [44] always holds under (DV3+). We say that $G\colon \mathsf{X} \to \mathbb{R}_+$ is *coercive* if the sublevel set $\{x : G(x) \leq n\}$ is precompact for each $n \geq 1$. Coercive functions exist only when X is σ-compact.

Proposition 2.6 *Let $\mathbf{\Phi}$ be a ψ-irreducible and aperiodic Markov chain on X. Assume moreover that $\mathsf{X} = \mathbb{R}^n$; that condition* (DV3+) *holds with $V\colon \mathsf{X} \to [1,\infty)$ continuous; W unbounded; and the kernels $\{I_{C_W(r)}P^{T_0} : r \geq 1\}$ are v-separable for some $T_0 \geq 1$. Then, there exists a sequence of compact sets $\{K_n : n \geq 1\}$ satisfying (27).*

Proof. Lemma B.2 combined with Proposition C.7 implies that we may construct functions (V_1, W_1) from X to $[1,\infty)$, and a constant b_1 satisfying the following: $\sup\{V(x) : x \in C_{W_1}(r)\} < \infty$ for each r; $W_1, V_1 \in L_\infty^V$; W_1 is coercive; and $\mathcal{H}(V_1) \leq V_1 - W_1 + b_1$. Lemma C.8 combined with continuity of V then implies that (27) also holds, with $K_r =$ closure of $C_{W_1}(n_r)$ for some sequence of positive integers $\{n_r\}$. $\square$

Proposition 2.6 has a partial converse:

Proposition 2.7 *Suppose the chain $\mathbf{\Phi}$ is ψ-irreducible and aperiodic. Suppose moreover that $\mathsf{X} = \mathbb{R}^n$; that the support of ψ has non-empty interior; that P has the Feller property; and that there exists a sequence of compact sets $\{K_n : n \geq 1\}$ satisfying (27). Then Condition* (DV3) *holds with $V, W\colon \mathsf{X} \to [1,\infty)$ continuous and coercive.*

Proof. Proposition A.2 asserts that there exists a solution to the inequality $\mathcal{H}(V) \leq -\frac{1}{2}W + b\mathbb{I}_C$ with (V,W) continuous and coercive, C compact, and $b < \infty$. Under the assumptions of the proposition, compact sets are small (combine Proposition 6.2.8 with Theorem 5.5.7 of [34]). We may conclude that C is small, and hence that (DV3) holds. $\square$

2.4 Perron-Frobenius Theory

As in [32] we find strong connections between the theory developed in this paper, and the Perron-Frobenius theory of positive semigroups, as developed in [39].

Suppose that $\{\widehat{P}^n : n \in \mathbb{Z}_+\}$ is a semigroup of positive operators. We assume that $\{\widehat{P}^n\}$ has finite spectral radius $\hat{\xi}$ in L_∞^v. Then, the resolvent kernel defined by $\widehat{R}_\lambda := [I\lambda - \widehat{P}]^{-1}$ is a bounded linear operator on L_∞^v for each $\lambda > \hat{\xi}$. We assume moreover that the semigroup is ψ-irreducible, that is, whenever $A \in \mathcal{B}$ satisfies $\psi(A) > 0$, then $\sum_{k=0}^\infty \widehat{P}^k(x, A) > 0$, for all $x \in \mathsf{X}$. If $\mathbf{\Phi}$ is a ψ-irreducible Markov chain, then for *any* measurable function $F\colon \mathsf{X} \to \mathbb{R}$, the kernel $\widehat{P} = P_f$ generates a ψ-irreducible semigroup. In general, under ψ-irreducibility of the semigroup, one may find many solutions to the minorization condition,

$$\widehat{R}_\lambda(x, A) = \sum_{k=0}^\infty \lambda^{-k-1}\widehat{P}^k \geq s(x)\nu(A), \qquad x \in \mathsf{X},\ A \in \mathcal{B}, \tag{28}$$

with $\lambda > 0$, $s \in \mathcal{B}^+$, and $\nu \in \mathcal{M}^+$, that is, $s\colon \mathsf{X} \to \mathbb{R}_+$ is measurable with $\psi(s) > 0$, and ν is a positive measure on $(\mathsf{X}, \mathcal{B})$ satisfying $\nu(\mathsf{X}) > 0$. The pair (s, ν) is then called small, just as in the probabilistic setting.

Theorem 3.2 of [39] states that there exists a constant $\hat{\lambda} \in (0, \infty]$, the *generalized principal eigenvalue*, or *g.p.e.*, such that, for any small function $s \in \mathcal{B}^+$,

$$\sum_{k=0}^\infty \lambda^{-k-1}\widehat{P}^k s(x) \quad \begin{cases} = \infty & \text{for all } x \in \mathsf{X}, \qquad \lambda < \hat{\lambda} \\[1em] < \infty & \text{for a.e. } x \in \mathsf{X}\ [\psi], \quad \lambda > \hat{\lambda}. \end{cases} \tag{29}$$

The semigroup is said to be $\hat{\lambda}$-*transient* if for one, and then all small pairs (s, ν), satisfying $s \in \mathcal{B}^+$, $\nu \in \mathcal{M}^+$, we have $\sum_{k=0}^\infty \hat{\lambda}^{-k-1}\nu\widehat{P}^k s < \infty$; otherwise it is called $\hat{\lambda}$-*recurrent*.

Proposition 2.8 shows that the generalized principal eigenvalue coincides with the spectral radius when considering positive semigroups that admit a spectral gap. Related results may be found in Theorem 4.4 and Proposition 4.5 of [32].

Proposition 2.8 *Suppose that $\{\widehat{P}^n : n \in \mathbb{Z}_+\}$ is a ψ-irreducible, positive semigroup. Suppose moreover that the semigroup admits a spectral gap in L_∞^v, with finite spectral radius $\hat{\xi}$. Then:*

(i) $\hat{\xi} = \hat{\lambda}$.

(ii) *The semigroup is $\hat{\lambda}$-recurrent.*

(iii) *$\widehat{P}$ is v-uniform.*

(iv) *For any $\lambda > \hat{\xi}$, and any (s, ν) that solve (28) with $s \in \mathcal{B}^+$, $\nu \in \mathcal{M}^+$, the function $h := [I\hat{\gamma} - (\widehat{R}_\lambda - s \otimes \nu)]^{-1}s, \in L_\infty^v$ is an eigenfunction.*

Proof. Suppose that either (i) or (ii) is false. In either case, for all small pairs (s, ν),

$$\lim_{\lambda \downarrow \hat{\xi}} \nu\widehat{R}_\lambda s = \sum_{k=0}^\infty \hat{\xi}^{-k-1}\nu\widehat{P}^k s < \infty.$$

It then follows that the projection operator $\widehat{Q}$ defined in (25) satisfies $\nu\widehat{Q}s = 0$ for all small $s \in L_\infty^v$, $\nu \in \mathcal{M}_1^v$. This is only possible if $\widehat{Q} = 0$, which is impossible under our assumption that the semigroup admits a spectral gap.

To complete the proof, observe that the semigroup generated by the kernel $\widehat{R}_\lambda$ also admits a spectral gap, with spectral radius $\hat{\gamma} = (\lambda - \hat{\xi})^{-1}$. It follows that there is a closed ball $D \subset \mathbb{C}$ containing $\hat{\gamma}$ such that the two kernels below are bounded linear operators on L_∞^v for each $\gamma \in D \setminus \{\hat{\gamma}\}$,

$$X_\gamma = [I\gamma - \widehat{R}_\lambda]^{-1}, \qquad Y_\gamma = [I\gamma - (\widehat{R}_\lambda - s \otimes \nu)]^{-1}.$$

From (i) and (ii) we know that $\widehat{R}_\lambda$ is $\hat{\gamma}$-recurrent, which implies that $\nu Y_{\hat{\gamma}} s = 1$, and that $\widehat{P}h = \hat{\xi}h$ (see [39, Theorem 5.1]). Moreover, again from (i), (ii), since $\nu Y_{\hat{\gamma}} s < \infty$ it follows that the spectral radius of $(\widehat{R}_\lambda - s \otimes \nu)$ is strictly less than $\hat{\gamma}$, which implies (iii). Finally, since $\|\!|Y_{\hat{\gamma}}|\!\|_v < \infty$ we may conclude that $h \in L_\infty^v$, and this establishes (iv). ☐

On specializing to the kernels $\{P_f : F \in L_\infty^{W_0}\}$ we obtain the following corollary. Define for any measurable function $F \colon \mathsf{X} \to (-\infty, \infty]$:

$$\begin{aligned} &\text{(i)} \quad \Lambda(F) = \log(\lambda(F)) = \text{ the logarithm of the g.p.e. for } P_f. \\ &\text{(ii)} \quad \Xi(F) = \log(\xi(F)) = \text{the logarithm of the spectral radius of } P_f. \end{aligned} \tag{30}$$

Lemma 2.9 *Consider a ψ-irreducible Markov chain, and a measurable function $G \colon \mathsf{X} \to \mathbb{R}_+$. If $\Xi(G) < \infty$ then $G \in L_\infty^V$.*

Proof. We have $\|\!|P_g^n|\!\|_v < \infty$ for some $n \geq 1$ when $\Xi(G) < \infty$. Consequently, since G and V are assumed positive, we have $g(x) \leq P_g^n v\,(x) \leq \|\!|P_g^n|\!\|_v v(x)$, for all $x \in \mathsf{X}$. ☐

Proposition 2.10 *Under* (DV3+) *the functional Ξ is finite-valued and convex on $L_\infty^{W_0}$, and may be identified as the logarithm of the generalized principal eigenvalue:*

$$\Xi(F) = \Lambda(F), \qquad F \in L_\infty^{W_0}.$$

Proof. Theorem 2.4 implies that P_f is v-separable, and Proposition 2.8 then gives the desired equivalence. Convexity is established in Lemma C.1. ☐

The spectral radius of the twisted kernel given in (21) also has a simple representation, when the function h is chosen as a solution to the multiplicative Poisson equation:

Proposition 2.11 *Assume that the Markov chain $\boldsymbol{\Phi}$ satisfies condition* (DV3+) *with W unbounded. For real-valued $F \in L_\infty^{W_0}$, the twisted kernel $\check{P}_{\check{f}}$ satisfies* (DV3+) *with Lyapunov function $\check{V} := V - \check{F} + c$ for $c \geq 0$ sufficiently large. Consequently, the semigroup generated by the twisted kernel has a discrete spectrum in $L_\infty^{\check{v}}$, and its log-spectral radius has the representation,*

$$\check{\Xi}(G) = \Xi(F + G), \qquad G \in L_\infty^{W_0}.$$

Proof. The kernels P_f and $\check{P}_{\check{f}}$ are related by a scaling and a similarity transformation,

$$\check{P}_{\check{f}} = \lambda(f)^{-1} I_{\check{f}}^{-1} P_f I_{\check{f}}.$$

It follows that (DV3+) (i) is satisfied with the Lyapunov function $\check{V}$, and we have $\check{V} \geq 1$ for sufficiently large c since $\check{f} \in L^v_\infty$. The representation of $\check{\Xi}$ also follows from the above relationship between $\check{P}_{\check{f}}$ and P_f.

The density condition (DV3+) (ii) follows similarly. Letting $b_r = \|\lambda(f)^{-1} f \mathbb{I}_{C_W(r)}\|_\infty$, we have, under the transition law $\check{P}_{\check{f}}$,

$$\check{\mathsf{P}}_x\{\Phi(T_0) \in A,\ \tau_{C_W(r)} > T_0\} \leq \check{f}^{-1}(x) b_r^{T_0} \check{\beta}_r(A), \quad A \in \mathcal{B},\ x \in C_W(r),$$

where $\check{\beta}_r(dx) = \beta_r(dx)\check{f}(x)$. To establish (DV3+) (ii) it remains to show that $\check{f}^{-1}$ is bounded on $C_W(r)$.

Since the set $C_W(r)$ is small for the semigroup $\{P_f^t : t \geq 0\}$, there exists $\epsilon > 0$, $T_1 < \infty$, and a probability distribution ν such that

$$P_f^{T_1}(x, dy) \geq \epsilon \nu(dy), \qquad x \in C_W(r),\ y \in \mathsf{X}.$$

It follows that

$$\lambda(f)^{-T_1} \check{f} = P_f^{T_1} \check{f} \geq \epsilon \nu(\check{f}), \qquad x \in C_W(r).$$

Consequently, $\check{f}^{-1}$ is bounded on $C_W(r)$. $\square$

2.5 Doeblin and Uniform Conditions

The uniform upper bound in condition (DV3+) (ii) is easily verified in many models. Consider first the special case of a discrete time chain $\mathbf{\Phi}$ with a countable state space X, and with W such that $C_W(r)$ is finite for all $r < \|W\|_\infty$. In this case we may take $T_0 = 1$ in (DV3+) (ii), and set

$$\beta_r(A) = \sum_{x \in C_W(r)} P(x, A), \qquad A \in \mathcal{B}\,.$$

This is the starting point for the bounds obtained in [2].

A common assumption for general state space models is the following:

Condition (U) There exist $1 \leq T_1 \leq T_2$ and a constant $b_0 \geq 1$, such that

$$P^{T_1}(x, A) \leq b_0 \frac{1}{T_2} \sum_{t=1}^{T_2} P^t(y, A)\,, \qquad x, y \in \mathsf{X},\ A \in \mathcal{B}. \tag{31}$$

See [13, 12], as well as [43, 27, 29]. It is obvious that (31) implies the validity of the upper bound in our assumption (DV3+) (ii). Somewhat surprisingly, Condition (U) also implies a corresponding lower bound, and moreover we may take the bounding measure equal to the invariant measure π:

Proposition 2.12 *Suppose that $\mathbf{\Phi}$ is an aperiodic, ψ-irreducible chain. Then, condition* (U) *holds if and only if there is a probability measure π on $(\mathsf{X}, \mathcal{B})$, a constant $N_0 \geq 1$, and a sequence of non-negative numbers $\{\delta_n : n \geq N_0\}$, satisfying,*

$$\begin{aligned} |P^n(x, A) - \pi(A)| &\leq \delta_n \pi(A)\,, \qquad A \in \mathcal{B},\ x \in \mathsf{X},\ n \geq N_0; \\ \lim_{n\to\infty} \delta_n &= 0. \end{aligned} \tag{32}$$

Proof. It is enough to show that condition (U) implies the sequence of bounds given in (32).

Condition (U) implies the following minorization,

$$\sum_{t=1}^{T_2} P^t(y, A) \geq \epsilon\nu(A), \qquad A \in \mathcal{B}, y \in \mathsf{X},$$

where $\epsilon = T_2 b_0^{-1}$, and $\nu(A) = P^{T_1}(x_0, A)$, $A \in \mathcal{B}$, with $x_0 \in \mathsf{X}$ arbitrary. Since the chain is assumed aperiodic and ψ-irreducible, it follows that the chain is *uniformly ergodic*, a property somewhat stronger than Doeblin's condition [34, Theorem 16.2.2]. Consequently, there exists an invariant probability measure π, and constants $B_0 < \infty, b_0 > 0$ such that,

$$|\!|\!| P^n - \mathbf{1} \otimes \pi |\!|\!|_1 \leq e^{-b_0 n + B_0}, \qquad n \in \mathbb{T}. \tag{33}$$

Condition (U) then gives the following upper bound: On multiplying (31) by $\pi(dy)$, and integrating over $y \in \mathsf{X}$, we obtain,

$$P^{T_1}(x, A) \leq b_0 \pi(A), \qquad x \in \mathsf{X},\ A \in \mathcal{B}.$$

Let Γ denote the bivariate measure given by, $\Gamma(dx, dy) = \pi(dx) P^{T_1}(x, dy)$, for $x, y \in \mathsf{X}$. The previous bound implies that Γ has a density $p(x, y; T_1)$ with respect to $\pi \times \pi$, where $p(\,\cdot\,, \,\cdot\,; T_1)$ is jointly measurable, and may be chosen so that it satisfies the strict upper bound, $p(x, y; T_1) \leq b_0$, for $x, y \in \mathsf{X}$. The probability measure Γ has common one-dimensional marginals (equal to π). Consequently, we must have $\int p(x, y; T_1)\pi(dx) = 1$ a.e. $y \in \mathsf{X}$ $[\pi]$.

For $n \geq 2T_1$ we *define* the density $p(x, y; n)$ via,

$$p(x, y; n) := \int P^{n-T_1}(x, dz) p(z, y; T_1), \qquad x, y \in \mathsf{X}.$$

We have the upper bound $\sup_{x,y} p(x, y; n) \leq b_0$ for all $n \geq T_1$ since P^k is an L_∞-contraction for any $k \geq 0$. Combining this bound with (33) gives the strict bound,

$$\begin{aligned}
|p(x, y; n) - 1| &= \left| \int P^{n-T_1}(x, dz)(p(z, y; T_1) - 1) \right| \\
&= \left| \int P^{n-T_1}(x, dz) p(z, y; T_1) - \int \pi(dz) p(z, y; T_1) \right| \\
&\leq b_0 |\!|\!| P^{n-T_1} - \pi |\!|\!|_1 \leq b_0 e^{B_0 - b_0(n - T_1)}, \qquad n \geq T_1, x, y \in \mathsf{X}.
\end{aligned}$$

This easily implies the result. ☐

Note that, for the special case of reflected Brownian motion on a compact domain, a similar result is established in [3].

We have already noted in the above proof that the lower bound in (32) implies the Doeblin condition, which is known to be equivalent to (V4) with V bounded for a ψ-irreducible chain [34, Theorem 16.2.2]. Consequently, condition (U) frequently holds for models on compact state spaces but it rarely holds for models on $\mathbb{R}^n$. We summarize this and related correspondences with drift criteria here.

Proposition 2.13 *Suppose that $\mathbf{\Phi}$ is an aperiodic, ψ-irreducible chain.*

(i) *If $\mathbf{\Phi}$ satisfies Doeblin's condition, then* (DV4) *holds with respect to the Lyapunov function $V \equiv 1$.*

(ii) *If $\mathbf{\Phi}$ satisfies condition* (U) *and $V_0 \colon \mathsf{X} \to [1,\infty)$ is given with $|||P|||_{v_0} < \infty$, then* (DV4) *holds for a function $V \colon \mathsf{X} \to [1,\infty)$ that is equivalent to V_0. And, trivially, part* (ii) *of condition* (DV3+) *also holds.*

Proof. Result (i) is a consequence of [34, Theorems 16.2.3 and 16.2.3] which state that the state space X is small under these assumptions, and hence (DV4) holds with $V \equiv 1$.

To prove (ii) we define,

$$V(x) := 1 + \log\Bigl(\mathsf{E}_x\Bigl[\exp\Bigl(\epsilon \sum_{i=0}^{T_1-1} r^i V_0(\Phi(i))\Bigr)\Bigr]\Bigr), \qquad x \in \mathsf{X},$$

where $r > 1$ is arbitrary, and $\epsilon > 0$ is to be determined. The functions V and V_0 are equivalent when $\epsilon \le {T_1}^{-1} r^{-T_1+1}$ since then by Hölder's inequality,

$$V(x) \le 1 + \frac{1}{T_1} \sum_{i=0}^{T_1-1} \log\bigl(\mathsf{E}_x[\exp(T_1 \epsilon r^i V_0(\Phi(i)))]\bigr), \qquad x \in \mathsf{X},$$

and the right hand side is in $L_\infty^{V_0}$ since $|||P^i|||_{v_0} < \infty$ for $i \ge 0$ under the assumptions of (ii). Moreover, we have $V \ge \epsilon V_0$ by considering only the first term in the definition of V. Hence $V \in L_\infty^{V_0}$ and $V_0 \in L_\infty^{V}$, which shows that V and V_0 are equivalent. We assume henceforth that this bound holds on ϵ.

Hölder's inequality also gives the bound,

$$\begin{aligned} Pe^V &= \mathsf{E}_x\Bigl[\exp\Bigl(\epsilon \textstyle\sum_{i=0}^{n-1} r^i V_0(\Phi(i+1))\Bigr)\Bigr] \\ &\le \mathsf{E}_x\Bigl[\exp\Bigl(pr^{-1}\epsilon \textstyle\sum_{i=1}^{n-1} r^i V_0(\Phi(i))\Bigr)\Bigr]^{1/p} \mathsf{E}_x\Bigl[\exp\Bigl(qr^{T_1-1}\epsilon V_0(\Phi(T_1))\Bigr)\Bigr]^{1/q}, \end{aligned}$$

where we set $p = r > 1$ and $q = r(r-1)^{-1} > 1$. Under Condition (U) we have $\|P^{T_1} e^{V_0}\|_\infty < \infty$. Consequently, provided $\epsilon > 0$ is chosen so that $qr^{T_1-1}\epsilon < 1$ we then have, for some constant b_1,

$$\mathcal{H}(V) := \log(Pe^V) - V \le -(1 - r^{-1})V + b_1.$$

This implies the result since the state space is small. $\square$

2.6 Donsker-Varadhan Theory

In Donsker and Varadhan's classic papers [14, 15, 16] there are two distinct sets of assumptions that are imposed for ensuring the existence of a large deviations principle, roughly corresponding to parts (i) and (ii) of our condition (DV3+).

Lyapunov criteria. The Lyapunov function criterion of [16, 43] is essentially equivalent to (DV3), with the additional constraint that the function W has compact sublevel sets; see conditions (1)–(5) on [43, p. 34]. In the general case (when X is not compact) this implies that (DV3) holds with an unbounded W.

It is worth noting that the nonlinear generator is implicitly already present in the Donsker-Varadhan work, visible both in the form of the rate function, and in the assumptions imposed in [15, 16, 43].

Continuity and density assumptions. In [43] two additional conditions are imposed on $\mathbf{\Phi}$. It is assumed that the chain satisfies a strong version of the Feller property, and that for each x, $P(x, dy)$ has a continuous density $p_x(y)$ with respect to some reference measure $\alpha(dy)$ which is independent of x.

These rather strong assumptions are easily seen to imply condition (DV3+) (ii) when W is coercive, so that the sets $C_W(r)$ are pre-compact.

3 Multiplicative Ergodic Theory

3.1 Multiplicative Mean Ergodic Theorems

The main results of this section are summarized in the following two theorems. In particular, the multiplicative mean ergodic theorem given in (35) will play a central role in the proofs of the large deviations limit theorems in Section 5. For all these results we will assume that $\mathbf{\Phi}$ satisfies (DV3) with an unbounded function W. As above, we let $\mathcal{B}^+$ denote the set of functions $h\colon \mathsf{X} \to [0,\infty]$ with $\psi(h) > 0$; for $A \in \mathcal{B}$ we write $A \in \mathcal{B}^+$ if $\psi(A) > 0$; and let $\mathcal{M}^+$ denote the set of positive measures on $\mathcal{B}$ satisfying $\mu(\mathsf{X}) > 0$.

As in (6) in the Introduction, we choose an arbitrary measurable function $W_0 : \mathsf{X} \to [1, \infty)$ in L_∞^W, whose growth at infinity is strictly slower than W. This may be expressed in terms of the weighted L_∞ norm via,

$$\lim_{r\to\infty} \|W_0 \mathbb{I}_{C_W(r)^c}\|_W = 0\,, \tag{34}$$

where $\{C_W(r)\}$ are the sublevel sets of W defined in (5). The function W_0 is fixed throughout this section.

Given $F \in L_\infty^{W_0}$ and an arbitrary $\alpha \in \mathbb{C}$, we recall from [32] the notation $\widehat{P}_\alpha := e^{\alpha F} P$, and

$$\mathcal{S}_\alpha := \mathcal{S}(\widehat{P}_\alpha) := \text{spectrum of } \widehat{P}_\alpha \text{ in } L_\infty^v\,,$$

where $v := e^V$ and V is the Lyapunov function in (DV3+).

Next, we collect the main results of this section in the following theorem. Recall the definition of the empirical measures $\{L_n\}$ from (14).

Theorem 3.1 (Multiplicative Mean Ergodic Theorem) *Assume that the Markov chain $\mathbf{\Phi}$ satisfies condition* (DV3+) *with an unbounded W. For any $m > 0, M > 0$ there exist $\overline{a} > m, \overline{\omega} > 0$ such that for any real-valued $F \in L_\infty^{W_0}$ with $\|F\|_{W_0} \le M$, and any α in the compact set*

$$\Omega = \Omega(\overline{a}, \overline{\omega}) := \{\alpha = a + i\omega \in \mathbb{C} \,:\, |a| \le \overline{a},\ \textit{and}\ |\omega| \le \overline{\omega}\}\,,$$

we have:

(i) *There is a maximal, isolated eigenvalue* $\lambda(\alpha F) \in \mathcal{S}_\alpha$ *satisfying* $|\lambda(\alpha F)| = \xi(\alpha F)$. *Furthermore,* $\Lambda(\alpha F) := \log(\lambda(\alpha F))$ *is analytic as a function of* $\alpha \in \Omega$*, and for real* α *it coincides with the log-generalized principal eigenvalue of Section 2.4.*

(ii) *Corresponding to each eigenvalue* $\lambda(\alpha F)$*, there is an eigenfunction* $\check{f}_\alpha \in L_\infty^v$ *and an eigenmeasure* $\check{\mu}_\alpha \in \mathcal{M}_1^v$*, where* $v := e^V$*, normalized so that* $\check{\mu}_\alpha(\check{f}_\alpha) = \check{\mu}_\alpha(\mathsf{X}) = 1$*. The function* $\check{f}_\alpha$ *solves the* multiplicative Poisson equation*,*

$$\widehat{P}_\alpha \check{f}_\alpha = \lambda(\alpha F)\check{f}_\alpha \,,$$

and the measure $\check{\mu}_\alpha$ *is a corresponding eigenmeasure:* $\check{\mu}_\alpha \widehat{P}_\alpha = \lambda(\alpha F)\check{\mu}_\alpha$.

(iii) *There exist constants* $b_0 > 0$, $B_0 < \infty$*, independent of* α*, such that for all* $x \in \mathsf{X}$, $\alpha \in \Omega$, $n \geq 1$,

$$\left| \mathsf{E}_x \left[\exp\Big(n[\alpha\langle L_n, F\rangle - \Lambda(\alpha F)] \Big) \right] - \check{f}_\alpha(x) \right| \leq |\alpha| v(x) e^{B_0 - b_0 n} \,. \tag{35}$$

Proof. Lemma B.3 in the Appendix shows that $(P_{f_0})^{2T_0+2}$ is v_η-separable for any $F_0 \in L_\infty^{W_0}$, and Theorem 3.5 then implies that the spectrum of P_{f_0} is discrete. It follows that solutions to the eigenvalue problem for P_{f_0} exist with $\check{f}_0 \in L_\infty^{v_\eta}$, $\check{\mu}_0 \in \mathcal{M}_1^{v_\eta}$. The eigenvalue satisfies $|\lambda(F_0)| = \xi(F_0) < \infty$. Smoothness of Λ is established in Proposition 4.3.

Theorem 3.4 establishes the limit (iii) for $\alpha \in \mathbb{C}$ in a neighborhood of the origin.

Consider then the twisted kernel $\check{P} = \check{P}_{\check{f}_a}$, where a is real. Proposition 2.11 states that this satisfies (DV3+) with Lyapunov function $\check{V} := V/\check{f}_a$. An application of Theorem 3.4 to this kernel then implies a uniform bound of the form (iii) for α in a neighborhood of a. For any given $\overline{a} > 0$ we may appeal to compactness of the line-segment $\{a \in \mathbb{R} : |a| \leq \overline{a}\}$ to construct $\overline{\omega} > 0$ such that (35) holds for $\alpha \in \Omega$. $\square$

We note that this result has many immediate extensions. In particular, if condition (DV3+) is satisfied, then this condition also holds with (V, W) replaced by $(1 - \eta + \eta V, W)$ for any $0 < \eta < 1$. Consequently, $\check{f} \in L_\infty^{v_\eta}$ for *any* $0 < \eta \leq 1$ when $F \in L_\infty^{W_0}$.

Part (iii) of the theorem is at the heart of the proof of all the large deviations properties we establish in Section 5. For example, from (35) we easily obtain that, for any $F \in L_\infty^{W_0}$, the log-moment generating functions of the partial sums

$$S_n = \sum_{i=0}^{n-1} F(\Phi(i)) = n\langle L_n, F\rangle$$

converge uniformly and exponentially fast:

$$\frac{1}{n} \log \mathsf{E}_x \big[\exp(\alpha n\langle L_n, F\rangle)\big] \to \Lambda(\alpha F), \quad n \to \infty. \tag{36}$$

We therefore think of $\Lambda(\alpha F)$ as the limiting log-moment generating function of the partial sums $\{S_n\}$ corresponding to the function F, and much of our effort in the following two section will be devoted to examining the regularity properties of Λ and its convex dual Λ^*.

Following [32], next we give a weaker multiplicative mean ergodic theorem for α in a neighborhood of the imaginary axis. Recall the following terminology: The *asymptotic variance*

$\sigma^2(F)$ of a function $F : \mathsf{X} \to \mathbb{R}$ is defined to be variance obtained in the corresponding Central Limit Theorem for the partial sums of $F(\Phi(n))$, assuming it exists. For a V-uniformly ergodic (or, equivalently, a geometrically ergodic) chain, the asymptotic variance is finite for any function F satisfying $F^2 \in L_\infty^V$, and [34, Theorem17.0.1] gives the representation,

$$\sigma^2(F) = \lim_{n\to\infty} n\mathsf{E}_\pi[(\langle L_n, F\rangle - \pi(F))^2] \,. \tag{37}$$

A function $F : \mathsf{X} \to \mathbb{R}$ is called *lattice* if there are $h > 0$ and $0 \le d < h$, such that $[F(x) - d]/h$ is an integer for all $x \in \mathsf{X}$. The minimal h for which this holds is called the *span* of F. If the function F can be written as a sum, $F = F_0 + F_\ell$, where F_ℓ is lattice with span h and F_0 has zero asymptotic variance then F is called *almost-lattice* (and h is its span). Otherwise, F is called *strongly non-lattice.* The lattice condition is discussed in more detail in [32]. The proof of the following result follows from Theorem 3.1 and the arguments used in the proof of [32, Theorem 4.2].

Theorem 3.2 (Bounds Around the $i\omega$-Axis) *Assume that the Markov chain $\mathbf{\Phi}$ satisfies condition* (DV3+) *with an unbounded W, and that $F \in L_\infty^{W_0}$ is real-valued.*

(NL) *If F is strongly non-lattice, then for any $m > 0$ and $0 < \omega_0 < \omega_1 < \infty$, there exist $\overline{a} > m$, $b_0 > 0$, $B_0 < \infty$ (possibly different than in Theorem 3.1), such that*

$$\Big|\mathsf{E}_x\Big[\exp\Big(n[\alpha\langle L_n, F\rangle - \Lambda(aF)]\Big)\Big]\Big| \le v(x)e^{B_0 - b_0 n}\,, \qquad x \in \mathsf{X},\ n \ge 1, \tag{38}$$

for all $\alpha = a + i\omega$ with $|a| \le \overline{a}$ and $\omega_0 \le |\omega| \le \omega_1$, where $v := e^V$.

(L) *If F is almost-lattice with span $h > 0$, then for any $m > 0$ and $\epsilon > 0$, there exist $\overline{a} > m$, $b_0 > 0$, and $B_0 < \infty$ (possibly different than above and in Theorem 3.1), such that (38) holds for all $\alpha = a + i\omega$ with $|a| \le \overline{a}$ and $\epsilon \le |\omega| \le 2\pi/h - \epsilon$.*

3.2 Spectral Theory of v-Separable Operators

The following continuity result allows perturbation analysis to establish a spectral gap under (DV3). Recall that we set $v_\eta := e^{\eta V}$; for any real-valued $F \in L_\infty^W$ we define $f := e^F$; and we let P_f denote the kernel $P_f(x, dy) := f(x)P(x, dy)$.

Lemma 3.3 *Suppose that $\mathbf{\Phi}$ is ψ-irreducible and aperiodic, and that condition* (DV3) *is satisfied. Then, for $0 < \eta \le 1$, $n \ge 1$, there exists $b_{\eta,n} < \infty$, such that for any $F, G \in L_\infty^{W_0}$,*

$$|\!|\!|P_f - P_g|\!|\!|_{v_\eta} \le b_{\eta,n}\|F - G\|_{W_0},$$

whenever $\|F\|_{W_0} \le n$, and $\|G\|_{W_0} \le n$. Moreover, for any $h \in L_\infty^{v_\eta}$ the map $F \mapsto P_f h$ is Frechet differentiable as a function from $L_\infty^{W_0}$ to $L_\infty^{v_\eta}$.

Proof. We have from the definition of the induced operator norm,

$$\begin{aligned} |\!|\!|P_f - P_g|\!|\!|_{v_\eta} &= \sup_{x\in\mathsf{X}}\Big(|f(x) - g(x)|\frac{Pv_\eta(x)}{v_\eta(x)}\Big) \\ &\le \sup_{x\in\mathsf{X}} |f(x) - g(x)| \exp\big(-\eta\delta W(x) + \eta b\big)\,. \end{aligned}$$

Also, we have the elementary bounds, for all $x \in \mathsf{X}$,

$$\begin{aligned}|f(x)-g(x)| = |e^{F(x)}-e^{G(x)}| &\leq |F(x)-G(x)|e^{|F(x)|+|G(x)|}\\ &\leq \|F-G\|_{W_0}W_0(x)\exp\bigl((\|F\|_{W_0}+\|G\|_{W_0})W_0(x)\bigr)\\ &\leq \|F-G\|_{W_0}\exp\bigl((1+\|F\|_{W_0}+\|G\|_{W_0})W_0(x)\bigr)\,.\end{aligned}$$

Combining these bounds gives,

$$|\!|\!|P_f - P_g|\!|\!|_{v_\eta} \leq \|F-G\|_{W_0}\sup_{x\in\mathsf{X}}\Bigl(\exp\bigl((1+\|F\|_{W_0}+\|G\|_{W_0})W_0(x)-\eta\delta W(x)+\eta b\bigr)\Bigr). \tag{39}$$

The supremum is bounded under the assumptions of the proposition, which establishes the desired bound.

We now show that, for any given $h \in L_\infty^{v_\eta}$, $F \in L_\infty^{W_0}$, the map $G \mapsto I_{G-F}P_f h$ represents the Frechet derivative of $P_f h$. We begin with the mean value theorem,

$$P_f h - P_g h - I_{G-F}P_f h = (G-F)[P_{f_\theta}h - P_f h]$$

where $F_\theta = \theta F + (1-\theta)G$ for some $\theta\colon \mathsf{X} \to (0,1)$. The bounds leading up to (39) then lead to the following bound, for all $x \in \mathsf{X}$,

$$\begin{aligned}&\bigl|[P_f h - P_g h - I_{G-F}P_f h]\,(x)\bigr|\\ &\leq \Bigl(\|G-F\|_{W_0}W_0(x)\Bigr)\Bigl(\|F-G\|_{W_0}\exp\bigl((1+\|F\|_{W_0}+\|F_\theta\|_{W_0})W_0(x)-\eta\delta W(x)+\eta b\bigr)\Bigr).\end{aligned}$$

It follows that there exists $b_1 < \infty$ such that

$$|\!|\!|[P_f h - P_g h - I_{G-F}P_f h]|\!|\!|_{v_\eta} \leq b_1\|F-G\|_{W_0}^2 \qquad G \in L_\infty^{W_0},\ \|F-G\|_{W_0} \leq 1\,,$$

which establishes Frechet differentiability. □

Next we present a *local* result, in the sense that it holds for all F with sufficiently small L_∞^W-norm, where the precise bound on $\|F\|_W$ is not explicit. Although a value can be computed as in [32], it is not of a very attractive form. Note that Theorem 3.4 does not require the density condition used in (DV3+).

The definition of the empirical measures $\{L_n\}$ is given in (14).

Theorem 3.4 (Local Multiplicative Mean Ergodic Theorem) *Suppose that* $\mathbf{\Phi}$ *is* ψ*-irreducible and aperiodic, and that condition* (DV3) *is satisfied. Then there exists* $\epsilon_0 > 0$, $0 < \eta_0 \leq 1$, *such that for any complex-valued* $F \in L_\infty^W$ *satisfying* $\|F\|_W \leq \epsilon_0$, *and any* $0 < \eta \leq \eta_0$:

(i) *There exist solutions* λ, $\check{f}$ *and* $\check{\mu}$ *to the eigenvalue problems*

$$P_f\check{f} = \lambda\check{f}, \quad \check{\mu}P_f = \lambda\check{\mu}\,. \tag{40}$$

These solutions satisfy $\check{f} \in L_\infty^{v_\eta}$, $\check{\mu} \in \mathcal{M}_1^{v_\eta}$, $\check{\mu}(\mathsf{X}) = \check{\mu}(\check{f}) = 1$, *and the eigenvalue* $\lambda = \lambda(F) \in \mathbb{C}$ *satisfies* $|\lambda| = \xi(\{P_f^t\})$. *Moreover, the solutions are uniformly continuous on this domain: For some* $b_\eta < \infty$,

$$|\Lambda(F) - \Lambda(G)| \leq b_\eta\|F-G\|_W, \qquad |\check{f} - \check{g}|_{v_\eta} \leq b_\eta\|F-G\|_W\,,$$

whenever $F, G \in L_\infty^W$ *satisfy* $\|F\|_W \leq \epsilon_0$, $\|G\|_W \leq \epsilon_0$.

(ii) *There exist positive constants B_0 and b_0 such that, for all $g \in L_\infty^{v_\eta}$, $x \in \mathsf{X}$, $n \geq 1$, we have*

$$\begin{aligned}\Big|\mathsf{E}_x\big[\exp(n\langle L_n, F\rangle - n\Lambda(F))g(\Phi(n))\big] - \check f(x)\check\mu(g)\Big| &\leq \|g\|_{v_\eta} e^{\eta V(x)+B_0-b_0 n} \\ \Big|\mathsf{E}_x\big[\exp(n\langle L_n, F\rangle - n\Lambda(F))\big] - \check f(x)\Big| &\leq \|F\|_W e^{\eta V(x)+B_0-b_0 n}\end{aligned} \tag{41}$$

with $\check f, \check\mu, \lambda(F)$ given as in (i).

(iii) *If V is bounded on the set C used in* (DV3) *then we may take $\eta_0 = 1$.*

Proof. Assumption (DV3) combined with Theorem 2.2 implies that P is v_η-uniform for all $\eta > 0$ sufficiently small (when V is bounded on C then (DV3) implies v-uniformity, so we may take $\eta = 1$).

It follows that the inverse $[I - P + \mathbf{1} \otimes \pi]^{-1}$ exists as a bounded linear operator on $L_\infty^{v_\eta}$ [34, Theorem 16.0.1]. An application of Lemma 3.3 implies that the kernels P_f converge to P in norm

$$|\!|\!| P - P_f |\!|\!|_{v_\eta} \to 0, \qquad \text{as } \|F\|_W \to 0\,, \qquad 0 < \eta \leq 1\,.$$

Consequently, there exists $\epsilon_1 > 0$ such that $[Iz - P_f + \mathbf{1} \otimes \pi]^{-1}$ is bounded for all $z \in \mathbb{C}$ satisfying $|z - 1| < \epsilon_1$, and all $F \in L_\infty^W$ satisfying $\|F\|_W \leq \epsilon_1$.

We have the explicit representation, writing $\Delta := [(z-1)I + I_{1-f}P]$, $H := [I - P + \mathbf{1} \otimes \pi]$,

$$\begin{aligned}[Iz - P_f + \mathbf{1} \otimes \pi]^{-1} &= [H + \Delta]^{-1} \\ &= [I + H^{-1}\Delta]^{-1} H^{-1}\,.\end{aligned}$$

The first term on the right hand side exists as a power series in $H^{-1}\Delta$, provided

$$|\!|\!|\Delta|\!|\!|_{v_\eta} < (|\!|\!|H^{-1}|\!|\!|_{v_\eta})^{-1}\,. \tag{42}$$

Moreover, in this case we obtain the bound,

$$|\!|\!|[Iz - P_f + \mathbf{1} \otimes \pi]^{-1}|\!|\!|_{v_\eta} \leq \frac{|\!|\!|H^{-1}|\!|\!|_{v_\eta}}{1 - |\!|\!|\Delta|\!|\!|_{v_\eta} |\!|\!|H^{-1}|\!|\!|_{v_\eta}} < \infty. \tag{43}$$

For any $F \in L_\infty^W$ we have the upper bound, $|F| \leq [|\!|\!|F|\!|\!|_W \delta^{-1}]\delta W$, where $\delta > 0$ is given in (DV3). Recalling the definition of the log-generalized principal eigenvalue functional Λ from Section 2.4, and assuming that $\theta := |\!|\!|F|\!|\!|_W \delta^{-1} < 1$, we may apply the convexity of Λ (see Lemma C.1) to obtain the upper bound,

$$|\Lambda(F)| \leq \Lambda(\theta\delta W) \leq \theta\Lambda(\delta W) \leq \theta b = |\!|\!|F|\!|\!|_W \delta^{-1} b \tag{44}$$

where b is given in (DV3).

From (44) we conclude that there is a constant $\epsilon_0 > 0$ such that $\epsilon_0 < \frac{1}{2}\epsilon_1$, and (42) together with the bound $|\lambda(F) - 1| < \frac{1}{2}\epsilon_1$ hold whenever $|\!|\!|F|\!|\!|_W < \epsilon_0$. For such F, it follows that (43) holds, and hence P_f is v_η-uniform. Setting $\check H := [I\lambda(F) - P_f + \mathbf{1} \otimes \pi]$ we may express the eigenfunction and eigenmeasure explicitly as:

$$\begin{aligned}\check f &:= c_1 \check H^{-1}\mathbf{1}\,, \qquad c_1 := \Big(\frac{\pi \check H^{-1}\mathbf{1}}{\pi \check H^{-2}\mathbf{1}}\Big) \\ \check\mu &:= c_2 \pi \check H^{-1}\,, \qquad c_2 := \Big(\frac{1}{\pi \check H^{-1}\mathbf{1}}\Big)\,.\end{aligned}$$

The remaining results follow as in [32, Theorem 4.1]. ☐

In order to extend Theorem 3.4 to a non-local result we invoke the density condition in (DV3+) (ii). In fact, any such extension seems to require some sort of a density assumption.

Recall that, in the notation of Section 2.2 and Section 2.4, we say that the spectrum $\mathcal{S}$ in L_∞^v of a linear operator $\widehat{P} : L_\infty^v \to L_\infty^v$ is *discrete*, if for any compact set $K \subset \mathbb{C} \setminus \{0\}$, $\mathcal{S} \cap K$ is finite and contains only poles of finite multiplicity. We saw earlier that condition (DV3+) implies that P^{2T_0+2} is v-separable. Next we show in turn that any v-separable linear operator $\widehat{P}$ has a discrete spectrum in L_∞^v.

Theorem 3.5 (v-Separability ⇒ Discrete Spectrum) *If the linear operator $\widehat{P}\colon L_\infty^v \to L_\infty^v$ is bounded and $\widehat{P}^{T_0}\colon L_\infty^v \to L_\infty^v$ is v-separable for some $T_0 \ge 1$, then $\widehat{P}$ has a discrete spectrum in L_∞^v.*

Proof. Assume first that $T_0 = 1$. For a given $\epsilon > 0$, set $\widehat{P} = K + \Delta$ with $|||\Delta|||_v < \epsilon$, and with K a finite-rank operator. Write $K = \sum_{i=1}^n s_i \otimes \nu_i$, and for each $z \in \mathbb{C}$ define the complex numbers $\{m_{ij}(z)\}$ via

$$m_{ij}(z) = \langle \nu_i, [Iz - \Delta]^{-1} s_j \rangle, \qquad 1 \le i, j \le n\,.$$

Let $M(z)$ denote the corresponding $n \times n$ matrix, and set $\gamma(z) = \det(I - M(z))$. The function γ is analytic on $\{|z| > |||\Delta|||_v\}$ because on this domain we have

$$[Iz - \Delta]^{-1} = \sum z^{-n-1}\Delta^n, \qquad |||[Iz - \Delta]^{-1}|||_v \le (|z| - |||\Delta|||_v)^{-1} < \infty.$$

Moreover, this function satisfies $\gamma(z) \to 1$ as $|z| \to \infty$, from which we may conclude that the equation $\gamma(z) = 0$ has at most a finite number of solutions in any compact subset of $\{|z| > |||\Delta|||_v\}$.

As argued in the proof of Theorem 3.4, if $\gamma(z) \neq 0$, then we have,

$$\begin{aligned}[Iz - \widehat{P}]^{-1} &= [(Iz - \Delta) - K]^{-1} \\ &= [Iz - \Delta]^{-1}[I - K[Iz - \Delta]^{-1}]^{-1}\,.\end{aligned}$$

Conversely, this inverse does not exist when $\gamma(z) = 0$. Recalling that $\epsilon \ge |||\Delta|||_v$, we conclude that $\mathcal{S}(\widehat{P}) \cap \{z : |z| > \epsilon\} = \{z : \gamma(z) = 0\}$. The right hand side denotes a finite set, and $\epsilon > 0$ is arbitrary. Consequently, it follows that the spectrum of $\widehat{P}$ is discrete.

If $T_0 > 1$ then from the foregoing we may conclude that the spectrum of $\widehat{P}^{T_0}$ is discrete. The conclusion then follows from the identity

$$\big[Iz - \widehat{P}\big]^{-1} = \sum_{k=0}^{T_0-1} z^{-k+T_0-1}\Big(\widehat{P}^k\big[Iz^{T_0} - \widehat{P}^{T_0}\big]^{-1}\Big), \qquad z \in \mathbb{C}. \qquad \square$$

For each $n \ge 1$, we define the nonlinear operators Λ_n and $\mathcal{G}_n$ the space of real-valued functions $F \in L_\infty^{W_0}$, via,

$$\begin{aligned}\Lambda_n(F) &:= \tfrac{1}{n}\log \mathsf{E}_x\big[\exp(n\langle L_n, F\rangle)\big] \\ \mathcal{G}_n(F) &:= \log \mathsf{E}_x\Big[\exp\Big(n[\langle L_n, F\rangle - \Lambda(F)]\Big)\Big]\,, \qquad F \in L_\infty^{W_0},\ x \in \mathsf{X}\,.\end{aligned}$$

The following result implies that both sequences of operators $\{\mathcal{G}_n\}$ and $\{\Lambda_n\}$ are convergent. Smoothness properties of the limiting nonlinear operators are established in Propositions 4.3 and 4.5.

Proposition 3.6 *Suppose that* (DV3+) *holds with an unbounded function W. Then there exists a nonlinear operator $\mathcal{G}\colon L_\infty^{W_0} \to L_\infty^V$ such that $\check{f} = e^{\mathcal{G}(F)}$ is a solution to the multiplicative Poisson equation for each $F \in L_\infty^{W_0}$. Moreover, for each $F_0 \in L_\infty^{W_0}$ and $\delta_0 > 0$ we have,*

$$\sup_{\|F-F_0\|_{W_0}\le\delta_0} \|\mathcal{G}_n(F) - \mathcal{G}(F)\|_V \quad\to\quad 0,$$

$$\sup_{\|F-F_0\|_{W_0}\le\delta_0} \|\Lambda_n(F) - \Lambda(F)\|_V \quad\to\quad 0\,, \qquad n\to\infty.$$

Proof. Note that the second bound follows from the first. So, let $\delta_0 > 0$ and $F_0 \in L_\infty^{W_0}$ be given, and consider an arbitrary $F \in L_\infty^{W_0}$ satisfying $\|F - F_0\|_{W_0} \le \delta_0$. We define $\check{F}_n := \mathcal{G}_n(F)$ for $n \ge 0$, and $\check{F} = \mathcal{G}(F) := \log(\check{f})$, with $\check{f}$ given in Theorem 3.1. We show below that for any $\eta > 0$, there exists $b(\eta) < \infty$ such that for all such F,

$$\begin{aligned} |\check{F}(x)| &\le \eta V(x) + b(\eta)\,, \qquad x\in\mathsf{X}\,;\\ |\check{F}_n(x)| &\le \eta V(x) + b(\eta)\,, \qquad x\in\mathsf{X}\,,\ n\ge 1. \end{aligned} \tag{45}$$

Taking this for granted for the moment, observe that we then have, for any $r \ge 1$, $n \ge 1$,

$$\sup_{\|F-F_0\|_{W_0}\le\delta_0} \||\check{F}_n - \check{F}|\mathbb{I}_{C_V(r)^c}\|_V \le 2[\eta + b(\eta)r^{-1}]\,.$$

Moreover, Theorem 3.1 implies that for any $r \ge 1$,

$$\sup_{\|F-F_0\|_{W_0}\le\delta_0} \||\check{F}_n - \check{F}|\mathbb{I}_{C_V(r)}\|_V \to 0\,, \quad \textit{exponentially fast as } n\to\infty,$$

provided we have the uniform bound (45). Putting these two conclusions together, and letting $r \to \infty$ then gives,

$$\limsup_{n\to\infty} \sup_{\|F-F_0\|_{W_0}\le\delta_0} \|\check{F}_n - \check{F}\|_V \le 2\eta.$$

This then proves the desired uniform convergence, since $\eta > 0$ is arbitrary.

We now prove the uniform bound (45). We begin with consideration of the functions $\{\check{F} : \|F - F_0\|_{W_0} \le \delta_0\}$, since the corresponding bounds on $\{\check{F}_n\}$ then follow relatively easily.

We know that $\check{f} \in L_\infty^{v_\eta}$ from Theorem 3.1. (If (DV3+) holds, then it also holds with V replaced by $(1-\eta) + \eta V$ for any $0 < \eta < 1$.) This implies that $\check{F}(x) \le \eta V(x) + \log\|\check{f}\|_{v_\eta}$, for $x \in \mathsf{X}$. Hence it remains to obtain a lower bound.

Let $\tau = \min\{k \ge 1 : |\check{F}(\Phi(k))| \le r\}$, with $r \ge 1$ chosen so that $\{x : |\check{F}(x)| \le r\} \in \mathcal{B}^+$. The stochastic process below is a positive local martingale,

$$m(t) = \exp\Big(t\langle L_t, (F - \Lambda(F))\rangle\Big)\check{f}(\Phi(t))\,, \qquad t\in\mathbb{Z}_+.$$

The local martingale property combined with Fatou's Lemma then gives the bound,

$$\check{f}(x) \ge \mathsf{E}_x\Big[\exp\Big(\tau\langle L_\tau, (F - \Lambda(F))\rangle\Big)\check{f}(\Phi(\tau))\Big]\,, \quad x\in\mathsf{X},$$

and then by Jensen's inequality and the definition of τ,

$$\begin{aligned}\check{F}(x) &\geq \mathsf{E}_x\Big[\check{F}(\Phi(\tau)) + \tau\langle L_\tau, (F - \Lambda(F))\rangle\Big] \\ &\geq -r - \mathsf{E}_x\Big[\tau\langle L_\tau, |F + \Lambda(F)|\rangle\Big], \qquad x \in \mathsf{X}.\end{aligned} \tag{46}$$

The right hand side is bounded below by $-k_0(V+1)$ for some finite k_0 by (V3) and [34, Theorem 14.0.1]. However, this bound can be improved. Since $F \in L_\infty^{W_0}$, and since $W \in L_\infty^V$ with (W_0, W) satisfying (6), we can find, for any $\eta_0 > 0$, a constant $b_0(\eta_0)$ and a small set S_{η_0} satisfying

$$|F + \Lambda(F)| \leq b_0(\eta_0)\mathbb{I}_{S_{\eta_0}} + \eta_0 V. \tag{47}$$

Small sets are *special* (see [39]), which implies that

$$\sup_{x\in\mathsf{X}} \mathsf{E}_x\big[\tau\langle L_\tau, S_{\eta_0}\rangle\big] < \infty. \tag{48}$$

Moreover, it follows from [34, Theorem 14.0.1] that for some $b_0 < \infty$,

$$\mathsf{E}_x\big[\tau\langle L_\tau, V\rangle\big] \leq b_0 V(x), \qquad x \in \mathsf{X}. \tag{49}$$

Combining the bounds (46–49) establishes (45) for $\check{F}$.

From (35) in Theorem 3.1 we have, for any $\eta > 0$, constants $B_\eta < \infty, b_\eta > 0$ such that, whenever $\|F - F_0\|_{W_0} \leq 1$,

$$\check{F}_n(x) \leq \check{F}(x) + \log\big(1 + \exp(\eta V(x) - \check{F}(x) + B_\eta - b_\eta n)\big), \qquad n \geq 1.$$

From the forgoing we see that the right hand side is bounded by $2\eta V + b(2\eta)$ for some $b(2\eta) < \infty$ and all n.

To complete the proof, we show that a corresponding lower bound holds: By definition of $\check{f}_n$ and an application of Jensen's inequality we have for all $n \geq 0$,

$$\check{f}_n(x)\check{f}^{-1}(x) = \check{\mathsf{E}}_x[\check{f}^{-1}(\Phi(n))] \geq (\check{\mathsf{E}}_x[\check{f}(\Phi(n))])^{-1}$$

where the expectation is with respect to the process with transition kernel $\check{P}_{\check{f}}$. On taking logarithms, and appealing to the mean ergodic limit for the twisted process, for constants $B_\eta < \infty, b_\eta > 0$,

$$\check{F}_n(x) - \check{F}(x) \geq -\log(\check{\mathsf{E}}_x[\check{f}(\Phi(n))]) \geq -\log\big(\check{\pi}(\check{f}) + \exp(\eta V(x) + B_\eta - n b_\eta)\big), \qquad n \geq 1.$$

This together with the bounds obtained on $\check{F}$ shows that (45) does hold. $\square$

4 Entropy, Duality and Convexity

In this section we consider structural properties of the operators $\mathcal{G}$, $\mathcal{H}$ and the functional Λ. As above, we assume throughout that $\mathbf{\Phi}$ satisfies (DV3+) with an unbounded function W, and we choose and fix an arbitrary function $W_0 \in L_\infty^W$ as in (34). Also, throughout this section we restrict attention to real-valued functions in $L_\infty^{W_0}$ and real-valued measures in $\mathcal{M}_1^{W_0}$ since one of our goals is to establish convexity and present Taylor series expansions of $\mathcal{G}$, $\mathcal{H}$, and Λ acting on $L_\infty^{W_0}$. Recall from Proposition 2.8 that the log-generalized principle eigenvalue Λ coincides with the log-spectral radius Ξ on this domain.

The convex dual of the functional $\Lambda\colon L_\infty^{W_0} \to \mathbb{R}$ is defined for $\mu \in \mathcal{M}_1^{W_0}$ via,

$$\Lambda^*(\mu) := \sup\{\langle \mu, F\rangle - \Lambda(F) : F \in L_\infty^{W_0}\}\,. \tag{50}$$

A probability measure $\mu \in \mathcal{M}_1^{W_0}$ and a function $F \in L_\infty^{W_0}$ form a *dual pair* if the above supremum is attained, so that $\Lambda(F) + \Lambda^*(\mu) = \langle \mu, F\rangle$.

The main result of this section is a proof that Λ^* can be expressed in terms of relative entropy (recall (17)) provided that we extend the definition to include bivariate measures on $(\mathsf{X}\times\mathsf{X}, \mathcal{B}\times\mathcal{B})$. Throughout this section we let M denote a generic function on $\mathsf{X}\times\mathsf{X}$, and Γ a generic measure on $(\mathsf{X}\times\mathsf{X}, \mathcal{B}\times\mathcal{B})$. The definitions of L_∞^W and $\mathcal{M}_1^W$ are extended as follows:

$$L_{\infty,2}^W \;:=\; \Big\{M : \|M\|_W := \sup_{(x,y)\in\mathsf{X}\times\mathsf{X}} \Big(\frac{|M(x,y)|}{[W(x)+W(y)]}\Big) < \infty\Big\} \tag{51}$$

$$\mathcal{M}_{1,2}^W \;:=\; \Big\{\Gamma : \|\Gamma\|_W := \int_{\mathsf{X}\times\mathsf{X}} [W(x)+W(y)]\, |\Gamma(dx,dy)| < \infty\Big\}\,. \tag{52}$$

The following proposition shows that consideration of the bivariate chain $\boldsymbol{\Psi}$,

$$\Psi(k) = \begin{pmatrix}\Phi(k+1)\\ \Phi(k)\end{pmatrix}, \qquad k\ge 0,\ \Psi(0)\in \mathsf{X}\times\mathsf{X}\,, \tag{53}$$

allows us to extend the domain of Λ to include bivariate functions, and then Λ^* is defined on bivariate measures via

$$\Lambda^*(\Gamma) := \sup_{M\in L_{\infty,2}^{W_0}} (\langle \Gamma, M\rangle - \Lambda(M)), \qquad \Gamma \in \mathcal{M}_{1,2}^{W_0}. \tag{54}$$

For any univariate measure μ and transition kernel $\check{P}$, we write $\mu\odot\check{P}$ for the bivariate measure $\mu\odot\check{P}(dx,dy) := \mu(dx)\check{P}(x,dy)$. In particular, Proposition 4.1 shows that if $\boldsymbol{\Phi}$ satisfies (DV3+) with an unbounded W, then so does $\boldsymbol{\Psi}$.

Proposition 4.1 *The following implications hold for any Markov chain $\boldsymbol{\Phi}$, with corresponding bivariate chain $\boldsymbol{\Psi}$:*

(i) *If $\boldsymbol{\Phi}$ is ψ-irreducible, then $\boldsymbol{\Psi}$ is ψ_2-irreducible, with $\psi_2 := \psi\odot P$;*

(ii) *If C is a small set for $\boldsymbol{\Phi}$, then $\mathsf{X}\times C$ is small for $\boldsymbol{\Psi}$;*

(iii) *If $C\in\mathcal{B}$, μ, and $T_0\ge 1$ satisfy $P^{T_0}(y,A)\le \mu(A)$ for $y\in C$, $A\in\mathcal{B}$, then on setting $C_2 = \mathsf{X}\times C$ and $\mu_2 = \mu\odot P$ we have,*

$$P_2^{T_0+1}((x,y),A_2) \le \mu_2(A_2), \qquad (x,y)\in C_2,\ A_2\in\mathcal{B}\times\mathcal{B},$$

where P_2 denotes the transition kernel for $\boldsymbol{\Psi}$;

(iv) *If $\nu\in\mathcal{M}^+$ is small for $\boldsymbol{\Phi}$ then $\nu_2 := \nu\odot P$ is small for $\boldsymbol{\Psi}$;*

(v) *Suppose that* $\mathbf{\Phi}$ *satisfies the drift condition* (DV3). *Then* $\mathbf{\Psi}$ *also satisfies the following version of* (DV3),

$$\mathcal{H}_2(V_2) \leq -\delta W_2 + b\mathbb{I}_{C_2}\,, \qquad \textit{on } S_{V_2}\,,$$

where $\mathcal{H}_2$ *is the nonlinear generator for* $\mathbf{\Psi}$, $C_2 = \mathsf{X} \times C$, *and*

$$V_2(x,y) = V(y) + \tfrac{1}{2}\delta W(x),\ W_2(x,y) = \tfrac{1}{2}(W(x) + W(y)),\ x, y \in \mathsf{X}.$$

Proof. To prove (i) consider any set $A_2 \in \mathcal{B} \times \mathcal{B}$ with $\psi_2(A_2) > 0$. Define

$$g(x) = \int_{y\in\mathsf{X}} P(x,dy)\mathbb{I}_{A_2}(x,y)\,, \quad x \in \mathsf{X}.$$

Then we have $\psi(g) > 0$, and hence by ψ-irreducibility of $\mathbf{\Phi}$, $\sum_{k=0}^{\infty} P^k g\,(x) > 0$, for all $x \in \mathsf{X}$. It follows immediately that $\sum_{k=0}^{\infty} P_2^k \mathbb{I}_{A_2}\,(x,y) > 0$, for all $x, y \in \mathsf{X}$, from which we deduce that $\mathbf{\Psi}$ is ψ_2-irreducible. This proves (i), and (ii)-(iv) are similar.

To see (v), observe that under (DV3),

$$\begin{aligned}
\log P_2 e^{V_2}\,(x,y) &= \log \int P(y,dz) e^{\frac{1}{2}\delta W(y) + V(z)} \\
&\leq \tfrac{1}{2}\delta W(y) + [V(y) - \delta W(y) + b\mathbb{I}_C(y)] \\
&= V_2(x,y) - \delta W_2(x,y) + b\mathbb{I}_{C_2}(x,y), \qquad x \in \mathsf{X},\ y \in S_V. \qquad \square
\end{aligned}$$

We show in Theorem 4.2 that the convex dual may be expressed as relative entropy when Γ is a probability measure in $\mathcal{M}_{1,2}^{W_0}$,

$$\Lambda^*(\Gamma) = H(\Gamma \| \check{\pi} \odot P) = \int_{\mathsf{X}\times\mathsf{X}} \log\Big(\frac{d\Gamma}{d[\check{\pi}\odot P]}(x,y)\Big)\,\Gamma(dx,dy), \tag{55}$$

where $\check{\pi}$ is the first marginal of Γ and $\check{\pi} \odot P$ denotes the bivariate measure $[\check{\pi} \odot P](dx,dy) = \check{\pi}(dx)P(x,dy)$. When $\Lambda^*(\Gamma) < \infty$, we show in Lemma 4.11 that the two marginals agree. Consequently, Γ may be expressed as, $\Gamma(dx,dy) = [\check{\pi} \odot \check{P}](dx,dy) = \check{\pi}(dx)\check{P}(x,dy)$, where $\check{P}$ is a transition kernel and $\check{\pi}$ is an invariant measure for $\check{P}$.

Theorem 4.2 (Identification of Λ^* as Relative Entropy) *Suppose that* (DV3+) *holds with an unbounded function* W. *Then:*

(i) *For any probability measure* $\Gamma \in \mathcal{M}_{1,2}^{W_0}$, *if* $\Lambda^*(\Gamma) < \infty$ *then the one-dimensional marginals* $\{\Gamma_1, \Gamma_2\}$ *agree. Consequently, letting* $\check{\pi} = \Gamma_1$ *denote the first marginal of* Γ *we can write, for some transition kernel* $\check{P}$,

$$\Gamma(dx,dy) = \check{\pi}(dx)\check{P}(x,dy)\,,$$

where $\check{\pi}$ *is an invariant measure for the transition kernel* $\check{P}$.

(ii) *If* $\Lambda^*(\Gamma) < \infty$ *for some probability measure* $\Gamma \in \mathcal{M}_{1,2}^{W_0}$, *then*

$$\Lambda^*(\Gamma) = H(\Gamma \,\|\, \check{\pi} \odot P) := \int_{\mathsf{X}\times\mathsf{X}} \log\Big(\frac{d\Gamma}{d[\check{\pi}\odot P]}(x,y)\Big)\Gamma(dx,dy)\,, \tag{56}$$

where $[\check{\pi} \odot P](dx,dy) := \check{\pi}(dx)P(x,dy)$ *and* $\check{\pi}$ *is the first marginal of* Γ.

(iii) *For any $c > 0$, the set $\{\Gamma \in \mathcal{M}_{1,2}^{W_0} : \Lambda^*(\Gamma) \le c\}$ is a bounded subset of $\mathcal{M}_{1,2}^{W_0}$.*

Proof. Any probability measure Γ on $(\mathsf{X} \times \mathsf{X}, \mathcal{B} \times \mathcal{B})$ can be decomposed as $\Gamma(dx, dy) = \check{\pi}(dx)\check{P}(x, dy)$, where $\check{\pi}$ is the first marginal for Γ. We show in Lemma 4.11 that the marginals of Γ must agree when $\Lambda^*(\Gamma) < \infty$, and this establishes (i).

Finiteness of $\Lambda^*(\Gamma)$ also implies that Γ is absolutely continuous with respect to $\pi \odot P$. This follows from Proposition 4.6 (iv) below, applied to the bivariate chain $\boldsymbol{\Psi}$. Consequently, the transition kernel can be expressed, $\check{P}(x, dy) = m(x, y)P(x, dy)$, for $x, y \in \mathsf{X}$, for some measurable function $m : \mathsf{X} \times \mathsf{X} \to [0, \infty]$.

With $M = \log m$, Proposition C.10 gives the upper bound,

$$\Lambda^*(\Gamma) \le \langle \Gamma, M \rangle = H(\Gamma \| \check{\pi} \odot P)\,.$$

We apply Proposition C.4 to obtain a corresponding lower bound: There is a sequence $\{M_k : k \ge 1\} \subset L_\infty$ such that $M_k \to M$ point-wise, $|M_k| \le |M|$ for all $k \ge 1$, and $\Lambda(M_k) \to \Lambda(M)$, as $k \to \infty$. Moreover, we have $\Lambda(M) = 0$ since $\check{P}(x, dy) = m(x, y)P(x, dy)$ is transition kernel for a positive recurrent Markov chain, and hence '1-recurrent [39]. Consequently,

$$\Lambda^*(\Gamma) \ge \langle \Gamma, M_k - \Lambda(M_k) \rangle \to \langle \Gamma, M \rangle, \qquad k \to \infty\,.$$

We thus obtain the identity $\Lambda^*(\Gamma) = \langle \Gamma, M \rangle$, which is precisely (ii).

Finally, part (iii) follows from Proposition 4.6 (iii) combined with Proposition 4.10. ☐

4.1 Convexity and Taylor Expansions

We now return to consideration of the univariate chain $\boldsymbol{\Phi}$, and establish some regularity and smoothness properties for the (univariate) functional Λ and the nonlinear operators $\mathcal{H}$ and $\mathcal{G}$.

We recall the definition of the twisted kernel $\check{P}_h$ from (21), and for any $h : \mathsf{X} \to (0, \infty)$ we define the bilinear and quadratic forms,

$$\begin{aligned} \langle\!\langle F, G \rangle\!\rangle_h &:= \big[\check{P}_h(FG) - (\check{P}_h F)(\check{P}_h G)\big] \\ \mathcal{Q}_h(F) &:= \langle\!\langle F, F \rangle\!\rangle_h \end{aligned} \qquad F, G \in L_\infty^{W_0}\,. \tag{57}$$

When $h \equiv 1$ we remove the subscript so that $\langle\!\langle F, G \rangle\!\rangle := P(FG) - (PF)(PG)$, and $\mathcal{Q}(F) := P(F^2) - (PF)^2$. It is well-known that $\sigma^2(F) := \pi(\mathcal{Q}(ZF))$ is equal to the asymptotic variance given in (37) [34, Theorem 17.5.3], where one version of the fundamental kernel $Z : L_\infty^v \to L_\infty^v$ is given by $Z = [I - P + \mathbf{1} \otimes \pi]^{-1}$; see [34, 32] for details.

The fundamental kernels $\{Z_h\}$ for $\{\check{P}_h\}$ and the quadratic forms $\{\mathcal{Q}_h\}$ determine the second-order Taylor series expansions for Λ, $\mathcal{G}$ and $\mathcal{H}$. We begin with an examination of Λ.

Proposition 4.3 *Suppose that* (DV3+) *holds with an unbounded function W. Then the functional Λ is finite-valued on $L_\infty^{W_0}$, and has the following properties:*

(i) Λ is strongly continuous: *For each $F_0 \in L_\infty^{W_0}$ there exists $B < \infty$, such that for all $F \in L_\infty^{W_0}$ satisfying $\|F\|_{W_0} < 1$,*

$$|\Lambda(F_0 + F) - \Lambda(F_0)| \le B \|F\|_{W_0};$$

(ii) $\Lambda\colon L_\infty^{W_0} \to L_\infty^V$ is smooth: *For each $F, F_0 \in L_\infty^{W_0}$, the function $\Lambda(F_0 + aF)$ is analytic as a function of a. Moreover, we have the second-order Taylor expansion,*

$$\Lambda(F_0 + aF) = \Lambda(F_0) + a\pi_g(F) + \tfrac{1}{2}a^2\pi_g(\mathcal{Q}_g(Z_g F)) + O(a^3), \qquad a \in \mathbb{R},$$

where $g = \check{f}_0 := e^{\mathcal{G}(F_0)}$, and π_g is the invariant probability measure of $\check{P}_g$.

Proof. Part (i) follows from Proposition 2.10 combined with Lemma 3.3.

To establish (ii) we note that $\Lambda_n(F_0 + aF)$ is an analytic function of a for each initial x, and $F_0, F \in L_\infty^{W_0}$. Proposition 3.6 states that this converges to $\Lambda(F_0 + aF)$, which is convex and hence also continuous on $\mathbb{R}$, and the convergence is uniform for a in compact subsets of $\mathbb{R}$. This implies that the limit is an analytic function of a.

The second-order Taylor series expansion follows as in the proof of property P4 in the Appendix of [32]. $\square$

We now consider $\mathcal{H}$, viewed as a nonlinear operator from $L_\infty^{W_0}$ to L_∞^V. Proposition 4.4 establishes smoothness and pointwise convexity of $\mathcal{H}$, and Proposition 4.5 gives analogous results for $\mathcal{G}$. See [6, Chapter 3] for related results for finite-dimensional positive matrices, and various applications to optimization.

Proposition 4.4 *Suppose that* (DV3+) *holds with an unbounded function W.*

(i) $\mathcal{H}\colon L_\infty^{W_0} \to L_\infty^V$ is pointwise convex: *For any $F_1, F_2 \in L_\infty^{W_0}$, and for any $\theta \in (0,1)$ we have,*

$$\mathcal{H}(\theta F_1 + (1-\theta)F_2) \le \theta\mathcal{H}(F_1) + (1-\theta)\mathcal{H}(F_2)\,,$$

where inequalities between functions are interpreted pointwise.

(ii) $\mathcal{H}$ is smooth: *We have the second-order Taylor expansions, for any $F, F_0 \in L_\infty^{W_0}$,*

$$\mathcal{H}(F_0 + aF) = \mathcal{H}(F_0) + a\mathcal{A}_g F + \tfrac{1}{2}a^2\mathcal{Q}_g(F) + O(a^3), \qquad a \in \mathbb{R},$$

where $g = \check{f}_0 := e^{\mathcal{G}(F_0)}$ and $\mathcal{A}_g$ is the generator of $\check{P}_g$.

Proof. We first show that $\mathcal{H}\colon L_\infty^{W_0} \to L_\infty^V$. To see this, take any $F \in L_\infty^{W_0}$. Since $W \in L_\infty^V$ and $W_0 \in L_\infty^W$ satisfies (6), we can find $b(F) < \infty$ such that $|F| \le V + b(F)$. It follows from (DV3) that,

$$\begin{aligned}
\log(Pe^F) &\ge \log(Pe^{-V}) - b(F) \ge -(V - \delta W + b + b(F)) \\
\log(Pe^F) &\le \log(Pe^V) + b(F) \le V - \delta W + b + b(F),
\end{aligned}$$

which shows that $\mathcal{H}(F) \in L_\infty^V$. Given these bounds, the smoothness result (ii) is a consequence of elementary calculus.

To establish convexity, we let $H_i = \mathcal{H}(F_i)$ and $f_i = e^{F_i}$, so that $Pf_i = e^{H_i} f_i$, $i = 1, 2$. An application of Hölder's inequality gives the bound,

$$P(f_1^\theta f_2^{(1-\theta)}) \le (Pf_1)^\theta (Pf_2)^{(1-\theta)} \exp(\theta H_1 + (1-\theta)H_2) f_1^\theta f_2^{(1-\theta)}.$$

With $F := \theta F_1 + (1-\theta)F_2 = \log(f_1^\theta f_2^{(1-\theta)})$ we then have

$$\mathcal{H}(F) = \log(Pf/f) \le \theta\mathcal{H}(F_1) + (1-\theta)\mathcal{H}(F_2)\,. \qquad \square$$

We can also obtain a Taylor-series approximation for $\mathcal{G}$, but it is convenient to consider a re-normalization to avoid additive constants. Define,

$$\mathcal{G}_0(F) = \mathcal{G}(F) - \pi(\mathcal{G}(F)), \qquad F \in L_\infty^{W_0}.$$

Proposition 4.5 *Suppose that* (DV3+) *holds with an unbounded function* W*. For each* $F_0 \in L_\infty^{W_0}$*,* $0 < \eta \le 1$*, there is* $\epsilon_0 > 0$*,* $b_0 < \infty$*, such that*

$$\|e^{\mathcal{G}_0(F_0+F)} - e^{\mathcal{G}_0(F_0)}\|_{v_\eta} \le b_0\|F\|_W,$$

whenever $\|F\|_{W_0} < \epsilon_0$*. We have the Taylor series expansion,*

$$\mathcal{G}_0(F_0 + aF) = \mathcal{G}_0(F_0) + aZ_{\check f_0}F + \tfrac{1}{2}a^2 Z_{\check f_0}\mathcal{Q}_{\check f_0}(Z_{\check f_0}F) + O(a^3), \qquad a \in \mathbb{R},$$

where $Z_{\check f_0}$ *is the fundamental kernel for* $\check P_{\check f_0}$*, normalized so that* $\pi Z_{\check f_0}F = 0$*,* $F \in L_\infty^{W_0}$*.*

Proof. The strong continuity follows from strong continuity of P_g given in Lemma 3.3.

The Taylor-series expansion is established first with $F_0 = 0$. Given $F \in L_\infty^{W_0}$, $a \in \mathbb{R}$, we let $f_a = \exp(aF)$, and let $\check f_a$ be the solution to the eigenfunction equation given by

$$\check f_a = [I\lambda_a - P_{f_a} + \mathbf{1}\otimes\pi]^{-1}\mathbf{1}\,.$$

Under assumption (DV3) alone we have seen in Theorem 3.4 that this is an eigenfunction in L_∞^v for small $|a|$. We also have $\check F_a = \log(\check f_a) = \mathcal{G}_0(F_a) + k(a)$, with $k(a) = \pi(\check F_a)$. In the analysis that follow, our consideration will focus on $\check F_a$ rather than $\mathcal{G}_0(F_a)$ since constant terms will be eliminated through our normalization.

We note that the first derivative may be written explicitly as,

$$\frac{d}{da}\check g_a = [I\lambda_a - P_{f_a} + \mathbf{1}\otimes\pi]^{-1}(\frac{d}{da}\lambda_a I - I_F P_{f_a})[I\lambda_a - P_{f_a} + \mathbf{1}\otimes\pi]^{-1}\mathbf{1}\,.$$

Observe that the derivative is in L_∞^v since both $I_F P_{f_a}$ and $[I - P_{f_a} + \mathbf{1}\otimes\pi]^{-1}$ are bounded linear operators on L_∞^v. Similar conclusions hold for all higher-order derivatives.

We define the twisted kernel as above,

$$\check P_a(x,A) := \check P_{\check f_a}(x,A) = \frac{\int_A P(x,dy)\check f_a(y)}{P\check f_a(x)} \qquad x \in \mathsf{X},\ A \in \mathcal{B}.$$

As in [32] we may verify that the function $\widehat F_a = \frac{d}{da}\check F_a$ is a solution to Poisson's equation,

$$\check P_a \widehat F_a = \widehat F_a - F + \pi_a(F), \qquad \pi_a(F) = \frac{d}{da}\Lambda(aF)\,,$$

where π_a is invariant for $\check P_a$. Setting $a = 0$ gives the first term in the Taylor series expansion for $\mathcal{G}_0$.

To obtain an expression for the second term we differentiate Poisson's equation:

$$\frac{d}{da}\bigl(\widehat{F}_a - F + \frac{d}{da}\Lambda(aF)\bigr) = \frac{d}{da}\bigl(\check{P}_a\widehat{F}_a\bigr) = \bigl(\frac{d}{da}\check{P}_a\bigr)\widehat{F}_a + \check{P}_a\widehat{F}_a^{(2)}. \tag{58}$$

We wish to compute the second derivative, $\widehat{F}_a^{(2)} = \frac{d^2}{da^2}\log(\check{f}_a)$, which requires a formula for the derivative of $\check{P}_a$: For any $G \in L_\infty^V$,

$$\begin{aligned} \tfrac{d}{da}\Bigl(\check{P}_a G\Bigr) &= \frac{P(\check{f}_a' G)P\check{f}_a - P(\check{f}_a G)(P\check{f}_a')}{(P\check{f}_a)^2} \\ &= \check{P}_a(\widehat{F}_a G) - \bigl(\check{P}_a G\bigr)\bigl(\check{P}_a\widehat{F}_a\bigr) = \langle\!\langle \widehat{F}_a, G\rangle\!\rangle_{\check{f}_a}. \end{aligned} \tag{59}$$

Letting $H_a = \langle\!\langle \widehat{F}_a, \widehat{F}_a\rangle\!\rangle_{\check{f}_a}$, the identities (58) and (59) then give,

$$\check{P}_a\widehat{F}_a^{(2)} = \widehat{F}_a^{(2)} - H_a + \Lambda''(aF). \tag{60}$$

Letting Z_a denote the fundamental kernel for $\check{P}_a$ we conclude that

$$\widehat{F}_a^{(2)} - \pi(\widehat{F}_a^{(2)}) = Z_a H_a = Z_a\langle\!\langle \widehat{F}_a, \widehat{F}_a\rangle\!\rangle_{\check{f}_a}.$$

Evaluating all derivatives at the origin provides the quadratic approximation for $\mathcal{G}_0$,

$$\mathcal{G}_0(aF) = aZF + \tfrac{1}{2}a^2 Z[\langle\!\langle \widehat{F}, \widehat{F}\rangle\!\rangle] + O(a^3)$$

where Z is the fundamental kernel for P, normalized so that $\pi Z = 0$, and $\widehat{F} = ZF$.

To establish the Taylor-series expansion at arbitrary $a_0 \in \mathbb{R}$ we repeat the above arguments, applied to the Markov chain with transition kernel $\check{P}_{a_0}$. This satisfies (DV3+) with $\check{V} = c + V - \check{F}_{a_0}$ for sufficiently large $c > 0$, by Proposition 2.11. $\square$

4.2 Representations of the Univariate Convex Dual

The following result provides bounds on the (univariate) convex dual functional Λ^*, and gives some alternative representations:

Proposition 4.6 *Suppose that* (DV3+) *holds with an unbounded function W. Then, for any probability measure $\mu \in \mathcal{M}_1^{W_0}$:*

(i) $\Lambda^*(\mu) = \sup\{\langle \mu, F\rangle - \Lambda(F) : F \in L_\infty$ *and* $\check{F} \in L_\infty\}$.

(ii) $\Lambda^*(\mu) = \sup\Bigl\{\langle \mu, -\mathcal{H}(H)\rangle : H \in L_\infty\Bigr\}$.

(iii) *There exists $\epsilon_0 > 0$, independent of $\mu \in \mathcal{M}_1^{W_0}$, such that*

$$\Lambda^*(\mu) \geq \epsilon_0\Bigl(\frac{\|\mu - \pi\|_{W_0}^2}{1 + \|\mu - \pi\|_{W_0}}\Bigr), \qquad \mu \in \mathcal{M}_1^{W_0}.$$

(iv) *If μ is not absolutely continuous with respect to π, then $\Lambda^*(\mu) = \infty$.*

The proof is provided after the following bound.

Lemma 4.7 *Suppose that* (DV3+) *holds with an unbounded function W. Then, $\check F \in L_\infty$ provided the following conditions hold: $F \in L_\infty^V$; $\Lambda(F)=0$; and $F = F\mathbb{I}_{C_V(r)}$ for some $r \ge 1$.*

PROOF. From the local martingale property we have,

$$\begin{aligned}\check F(x) &= \log \mathsf{E}_x\Big[\exp\Big(\textstyle\sum_{i=0}^{\tau_{C_V(r)}-1} F(\Phi(i))\Big)\check f(\Phi(\tau_{C_V(r)}))\Big]\\ &= F(x) + \log \mathsf{E}_x\big[\exp\big(\check f(\Phi(\tau_{C_V(r)}))\big)\big].\end{aligned}$$

This then gives the bound, $\|\check F\|_\infty \le \|F\|_\infty + \|\check F \mathbb{I}_{C_V(r)}\|_\infty < \infty$. □

PROOF OF PROPOSITION 4.6. For any $F \in L_\infty^{W_0}$, and any $r \ge 1$ we write, $F_r = \mathbb{I}_{C_V(r)}[F-\gamma_r]$, where $\gamma_r \in \mathbb{R}$ is chosen so that $\Lambda(F_r)=0$. Its existence follows from Proposition 4.3.

From Proposition C.5 we can show that $\gamma_r \to \Lambda(F)$, and then also that $\Lambda(F_r) \to 0$ as $r \to \infty$. Consequently, $\Lambda^*(\mu) = \sup\{\langle \mu, F_r\rangle - \Lambda(F_r) : F \in L_\infty^{W_0},\ r \ge 1\}$, and Lemma 4.7 implies that $\check F_r \in L_\infty$ for each r, which completes the proof of (i).

Part (ii) is essentially a reinterpretation of (i): From the equation $\mathcal{H}(\check F) = -F + \Lambda(F)$ and part (i) we obtain the upper bound,

$$\begin{aligned}\Lambda^*(\mu) &= \sup\{\langle \mu, F\rangle - \Lambda(F) : \check F \in L_\infty\}\\ &\le \sup\{\langle \mu, -\mathcal{H}(\check F)\rangle : \check F \in L_\infty\}\\ &\le \sup\{\langle \mu, -\mathcal{H}(G)\rangle : G \in L_\infty\}.\end{aligned}$$

Conversely, for any function $G \in L_\infty$, the function $F := -\mathcal{H}(G)$ satisfies $\Lambda(F)=0$, $F \in L_\infty$. This gives the desired lower bound, $\Lambda^*(\mu) \ge \langle \mu, F\rangle = \langle \mu, -\mathcal{H}(G)\rangle$, for $G \in L_\infty$.

Result (iii) is obtained from the mean value theorem, justified by Proposition 4.3: For any $F \in L_\infty^{W_0}$, $\epsilon \ge 0$, there is $0 \le \tilde\epsilon \le \epsilon$ such that $\Lambda(\epsilon F) = \epsilon\pi(F) + \frac12\epsilon^2\Lambda''(\tilde\epsilon F)$. Let $B_0 = \sup\{\Lambda''(\epsilon G) : \|G\|_{W_0} \le 1,\ 0 \le \epsilon \le 1\}$. Note that $B_0 < \infty$ by the Lemma following the proof. Then, whenever $\|F\|_{W_0} \le 1$, $\varepsilon \le 1$, we have $\Lambda(\epsilon F) \le \epsilon\pi(F) + \frac12 B_0\epsilon^2$. The definition of the convex dual then gives,

$$\begin{aligned}\epsilon\mu(F) = \langle \mu, \epsilon F\rangle &\le \Lambda^*(\mu) + \Lambda(\epsilon F)\\ &\le \Lambda^*(\mu) + \epsilon\pi(F) + \tfrac12 B_0\epsilon^2\,,\end{aligned}$$

and since this holds for any $\|F\|_{W_0} \le 1$, we have the absolute bound,

$$|\mu(F) - \pi(F)| \ \le\ \tfrac{1}{\epsilon}\Lambda^*(\mu) + \tfrac12 B_0\epsilon^2\,, \qquad \|F\|_{W_0} \le 1\,.$$

Letting $\epsilon = \sqrt{\Lambda^*(\mu)}$ we obtain

$$\|\mu - \pi\|_{W_0} = \sup_{\|F\|_{W_0} \le 1} |\mu(F) - \pi(F)| \le \sqrt{\Lambda^*(\mu)} + \tfrac12 B_0\Lambda^*(\mu)\,, \qquad |\Lambda^*(\mu)| < 1,$$

which implies the desired lower bound on Λ^*.

To prove (iv), write $\mu = p\mu_0 + (1-p)\mu_1$ where μ_0, μ_1 are probability measures on $(\mathsf{X}, \mathcal{B})$ such that $\mu_1 \prec \pi$ is absolutely continuous and μ_0 is singular with respect to π. Let S denote the support of μ_0. We have $\Lambda(F) = 0$ whenever $F \in L_\infty$ is supported on S, and hence

$$\begin{aligned}\Lambda^*(\mu) &\geq \sup\{\langle \mu, F\rangle - \Lambda(F) : F \in L_\infty^{W_0},\ F = \mathbb{I}_S F\} \\ &= p \sup\{\langle \mu_0, F\rangle : F \in L_\infty^{W_0},\ F = \mathbb{I}_S F\},\end{aligned}$$

which is infinite, as claimed. ☐

Lemma 4.8 $B_0 = \sup\{\frac{d^2}{da^2}\Lambda(aG) : \|G\|_{W_0} \leq 1,\ 0 \leq a \leq 1\} < \infty$.

Proof. (sketch) Let $\check{P}_a = \check{P}_{\check{g}_a}$ and let π_a denote the invariant distribution for given $\|G\|_{W_0} \leq 1$, and $a \in [0,1]$. We let Z_a the fundamental kernel for $\check{P}_a$, normalized so that $\pi_a(Z_a G) = \pi_a(G)$, and we let $\widehat{G}_a = Z_a G$. Proposition 4.3 then gives the representation,

$$\Lambda''(aG) = \pi_a(\mathcal{Q}_a(Z_a G)) = \pi_a\big(P_a(\widehat{G}_a^2) - (P_a \widehat{G}_a)^2\big).$$

The proof is completed on showing that

$$\sup \|\pi_a\|_v < \infty, \quad \sup \|\widehat{G}_a\|_V < \infty,$$

where the supremum is over all a and G in this class. This follows from the arguments above – see in particular (45) and the surrounding arguments. ☐

In the following proposition we give another characterization of dual pairs (μ, G) for Λ^*.

Proposition 4.9 *Suppose that* (DV3+) *holds with an unbounded function* W. *We then have:*

(i) *For any* $H \in L_\infty^{W_0}$, $\pi\big(\mathcal{H}(H)\big) \geq 0$, *with equality if and only if* $\mathcal{H}(H) = 0$, *in which case* H *is constant a.e.* $[\pi]$.

(ii) *If* $\mu \in \mathcal{M}_1^{W_0}$ *is not invariant under* P *then there is* $H \in L_\infty$ *satisfying* $\mu\big(\mathcal{H}(H)\big) < 0$.

(iii) *Suppose that* $\mu \in \mathcal{M}_1^{W_0}$, *and that there exists* $G \in L_\infty^{W_0}$ *satisfying,*

$$\Lambda^*(\mu) = \langle \mu, -\mathcal{H}(G)\rangle = \sup\Big\{\langle \mu, -\mathcal{H}(H)\rangle : H \in L_\infty^{W_0}\Big\}.$$

Then μ *is invariant under the twisted kernel* $\check{P}_g$.

Proof. The first result is simply Jensen's inequality:

$$\begin{aligned}\pi\big(\mathcal{H}(H)\big) &= \int \log\Big(\mathsf{E}_x\big[\exp\big(H(\Phi(1)) - H(\Phi(0))\big)\big]\Big)\pi(dx) \\ &\geq \mathsf{E}_\pi\big[H(\Phi(1)) - H(\Phi(0))\big] = 0.\end{aligned}$$

If equality holds, it then follows that e^H is constant a.e. $[\pi]$.

To prove (ii) let $F(x) = \epsilon[\mathbb{I}_A - \gamma_\epsilon \mathbb{I}_B]$, with $\epsilon > 0$, $A, B \in \mathcal{B}^+$ small sets such that $\sup_{A\cup B} V(x)$ is finite, and $\gamma_\epsilon > 0$ is chosen so that $\Lambda(F) = 0$. The function $H := \check{F} = \mathcal{G}(F)$ is then bounded, by Lemma 4.7. Moreover,

$$\mu\big(\mathcal{H}(H)\big) = -\mu(F) = -\epsilon[\mu(A) - \gamma_\epsilon \mu(B)].$$

Under (DV3+) we may apply Proposition 4.3 to justify the Taylor series expansion,

$$0 = \Lambda(F) = \epsilon[\pi(A) - \gamma_\epsilon \pi(B)] + O(\epsilon^2),$$

which gives $\gamma_\epsilon = \pi(A)/\pi(B) + O(\epsilon)$. Choosing A, B so that $\mu(A)/\mu(B) > \pi(A)/\pi(B)$ we see that this function H satisfies the desired bound for $\epsilon > 0$ sufficiently small.

We now prove (iii). Applying Proposition 4.6 (ii), the convex dual $\check{\Lambda}^*$ for the kernel $\check{P}_g$ may be expressed as

$$\check{\Lambda}^*(\mu) := -\Big(\inf_{H\in L_\infty^{W_0}} \Big\langle \mu, \log\Big(\frac{\check{P}_g h}{h}\Big)\Big\rangle\Big).$$

For any $H \in L_\infty^{W_0}$ set $H' = H + G$ so that,

$$\begin{aligned}
\check{\Lambda}^*(\mu) &= -\Big(\inf_{H'\in L_\infty^{W_0}} \Big\langle \mu, \log\Big(\tfrac{g}{h'}\tfrac{Ph'}{Pg}\Big)\Big\rangle\Big) \\
&= -\Big(\inf_{H'\in L_\infty^{W_0}} \Big\langle \mu, \log\Big(\tfrac{Ph'}{h'}\Big)\Big\rangle\Big) + \Big\langle \mu, \log\Big(\tfrac{Pg}{g}\Big)\Big\rangle = 0.
\end{aligned}$$

Thus μ is invariant for $\check{P}_g$, by (i). $\square$

4.3 Characterization of the Bivariate Convex Dual

We now turn to the case of bivariate functions and measures.

Given any function of two variables $M\colon \mathsf{X}\times\mathsf{X} \to \mathbb{R}$, we let $m = e^M$ and extend the definition of the scaled kernel in (20) via,

$$P_m(x, dy) := m(x, y)P(x, dy)\,, \qquad x, y \in \mathsf{X}\,.$$

The following result shows that the spectral radius of this kernel coincides with that defined for the bivariate chain $\mathbf{\Psi}$. The proof is routine.

Proposition 4.10 *Suppose that P_m has finite spectral radius λ_m in v_η-norm for all sufficiently small $\eta > 0$. Let P_2 denote the transition kernel for the bivariate chain $\mathbf{\Psi}$.*

(i) *$I_m P_2$ has the same spectral radius in $v_{\eta 2}$-norm for sufficiently small $\eta > 0$, with $v_{\eta 2}(x, y) = \exp(\eta[V(y) + \frac{1}{2}\delta W(x)])$.*

(ii) *If P_m has an eigenfunction $\check{f}$, then $I_m P_2$ also possesses an eigenfunction given by,*

$$\check{f}_2(x_1, x_2) = m(x_1, x_2)\check{f}(x_2).$$

For a Markov process with transition kernel P satisfying (DV3+), we say that M and $\widetilde{M}$ are *similar* if there exists $H \in L_\infty^V$ such that

$$\widetilde{M}(x,y) = M(x,y) + H(x) - H(y) \qquad a.e.\ (x,y) \in \mathsf{X} \times \mathsf{X}\ [\pi \odot P]\,.$$

The function M is called *degenerate* if it is similar to $\widetilde{M} \equiv 0$. The log-generalized principal eigenvalues agree ($\Lambda(M) = \Lambda(\widetilde{M})$) whenever $M, \widetilde{M}$ are similar. This is the basis of the following two lemmas.

Lemma 4.11 *Suppose that* (DV3+) *holds with an unbounded function W. If $\Gamma \in \mathcal{M}_{1,2}^{W_0}$ is a probability measure with $\Lambda^*(\Gamma) < \infty$, then $\Gamma \prec \pi \odot P$, and the one-dimensional marginals of Γ agree.*

Proof. The conclusion that $\Gamma \prec \pi \odot P$ follows from Proposition 4.6 (iv).

For any $M \in L_\infty(\mathsf{X} \times \mathsf{X})$, $H \in L_\infty$, we have $\Lambda(M) = \Lambda(\widetilde{M})$, where $\widetilde{M}(x,y) := M(x,y) + H(x) - H(y)$. Hence, for all such M, H,

$$\Lambda^*(\Gamma) \geq \langle \Gamma, \widetilde{M} \rangle - \Lambda(\widetilde{M}) = \langle \Gamma, M \rangle - \Lambda(M) + \langle \Gamma_1, H \rangle - \langle \Gamma_2, H \rangle$$

where Γ_1 and Γ_2 denote the two marginals. If $\Gamma_1 \neq \Gamma_2$ it is obvious that the right hand side cannot be bounded in H. $\square$

Lemma 4.12 *Suppose that* (DV3+) *holds with an unbounded function W. Suppose moreover that $M \in L_{\infty,2}^V$, and that the asymptotic variance of the partial sums $\sum_{k=0}^{n-1} M(\Phi(k), \Phi(k+1))$, $n \geq 1$, is equal to zero. Then the function M is degenerate.*

Proof. Applying [32, Proposition 2.4] to the bivariate chain $\mathbf{\Psi}$ with transition kernel P_2, we can find $\widehat{M}$ such that

$$\widehat{M}(\Phi(k), \Phi(k+1)) - \widehat{M}(\Phi(k-1), \Phi(k)) = -M(\Phi(k-1), \Phi(k)) + \pi_2(M) \qquad a.s.\ [\mathsf{P}_\pi],\ k \geq 1,$$

where $\pi_2 = \pi \odot P$ is the invariant probability measure for P_2. Since $\Phi(k+1)$ is conditionally independent of $\Phi(k-1)$ given $\Phi(k)$, it follows that $\widehat{M}$ does not depend on its first variable. Thus we can find $\widehat{F} \in L_\infty^V$ satisfying

$$\widehat{F}(\Phi(k+1)) - \widehat{F}(\Phi(k)) = -M(\Phi(k-1), \Phi(k)) + \pi_2(M) \qquad a.s.\ [\mathsf{P}_\pi],\ k \geq 1\,,$$

therefore, M is similar to the constant function $\pi_2(M)$:

$$M(x,y) = \pi_2(M) + G(x) - G(y), \qquad a.e.\ \pi \odot P\,,$$

with $G(x) = P\widehat{F}\,(x)$. $\square$

Theorem 4.13 (Identification of Dual Pairs) *Suppose that* (DV3+) *holds with an unbounded function* W.

(i) *Assume that* $M \in L^{W_0}_{\infty,2}$ *and* $\Gamma \in \mathcal{M}^{W_0}_{1,2}$ *are given, such that* $\Lambda^*(\Gamma) < \infty$ *and* (M, Γ) *is a dual pair, i.e.,* $\langle \Gamma, M \rangle = \Lambda(M) + \Lambda^*(\Gamma)$. *Define* M_0 *as the Radon-Nikodym derivative,*

$$M_0(x,y) = \log\Big(\frac{d\Gamma}{d[\check{\pi} \odot P]}(x,y)\Big) \qquad x, y \in \mathsf{X},$$

where $\check{\pi}$ *is a marginal of* Γ *(see Lemma 4.11). Then, the function* M_0 *is similar to* $M - \Lambda(M)$,

$$M_0(x,y) = M(x,y) - \Lambda(M) - \check{F}(x) + \check{F}(y),$$

where $\check{F} = \log(\check{f})$, *with* $\check{f}$ *equal to an eigenfunction for* P_m, *with eigenvalue* $\lambda(M)$.

(ii) *Conversely, suppose that* $\Gamma \in \mathcal{M}^{W_0}_{1,2}$ *is given, satisfying* $\Gamma \prec [\pi \odot P]$, *and suppose that its one-dimensional marginals agree. Consider the decomposition,* $\Gamma(dx,dy) = [\check{\pi} \odot \check{P}](dx,dy)$, *where* $\check{\pi} := \Gamma_1 = \Gamma_2$ *is the (common) first marginal of* Γ *on* $(\mathsf{X}, \mathcal{B})$, *and* $\check{P}$ *is a transition kernel. Let*

$$M(x,y) = \log\Big(\frac{d\Gamma}{d[\check{\pi} \odot P]}(x,y)\Big) \qquad x, y \in \mathsf{X}.$$

If $M \in L^{W_0}_{\infty,2}$, *then* $\Lambda^*(\Gamma)$ *is finite and* (Γ, M) *is a dual pair.*

Proof. Part (i) is a bivariate version of Proposition 4.9: We know that Γ is an invariant measure for a bivariate process, whose one-dimensional transition kernel is of the form,

$$\check{P}_m(x,dy) = e^{M(x,y) - \Lambda(M) - \check{F}(x) + \check{F}(y)} P(x,dy).$$

Invariance may be expressed as follows:

$$\Gamma(dy,dz) = \int_{x \in \mathsf{X}} \Gamma(dx,dy) \check{P}_m(y,dz), \qquad y, z \in \mathsf{X}.$$

Since Γ has equal marginals, denoted $\check{\pi}$, this identity may be expressed,

$$\check{\pi}(dy)\check{P}(y,dz) = \check{\pi}(dy)\check{P}_m(y,dz), \qquad y, z \in \mathsf{X},$$

which is the desired identity in (i).

To prove (ii), let $\check{\Lambda}(\,\cdot\,)$ denote the functional defining the log-generalized principal eigenvalue for the transition kernel $\check{P} = P_m$. Proposition 2.11 gives, $\check{\Lambda}(N) = \Lambda(N+M) - \Lambda(M)$, for any $N \in L^{W_0}_{\infty,2}$. We can then write,

$$\begin{aligned}
\Lambda^*(\Gamma) &= \sup\nolimits_{N \in L_\infty} \Big(\langle \Gamma, N \rangle - \Lambda(N)\Big) \\
&= \sup\nolimits_{N \in L_\infty} \Big(\langle \Gamma, N + M \rangle - \Lambda(N+M)\Big) \\
&= \sup\nolimits_{N \in L_\infty} \Big(\langle \Gamma, N \rangle + \langle \Gamma, M \rangle - \check{\Lambda}(N) - \Lambda(M)\Big) \\
&= \check{\Lambda}^*(\Gamma) + \langle \Gamma, M \rangle - \Lambda(M).
\end{aligned}$$

We have $\check{\Lambda}^*(\Gamma) = 0$ by Proposition 4.9, and consequently $\langle \Gamma, M \rangle = \Lambda(M) + \Lambda^*(\Gamma)$. This shows that (M, Γ) is a dual pair. ☐

5 Large Deviations Asymptotics

In this section we use the multiplicative mean ergodic theorems of Section 3 and the structural results of Section 4 to study the large deviations properties of the empirical measures $\{L_n\}$ induced by the Markov chain $\mathbf{\Phi}$ on $(\mathsf{X}, \mathcal{B})$; recall the definition of $\{L_n\}$ in (14).

As in the previous section, we also assume throughout this section that the Markov chain $\mathbf{\Phi}$ satisfies (DV3+) with an unbounded function W, and we choose and fix a function $W_0 : \mathsf{X} \to [1, \infty)$ in L_∞^W as in (34). Our first result, the large deviations principle (LDP) for the sequence of measures $\{L_n\}$, will be established in a topology finer (and hence stronger) than either the topology of weak convergence, or the τ-topology. As described in the Introduction, we consider the τ^{W_0}-topology on the space $\mathcal{M}_1$ of probability measures on $(\mathsf{X}, \mathcal{B})$, defined by the system of neighborhoods (16).

Since the map $(x_1, \ldots, x_n) \mapsto \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$ from X^n to $\mathcal{M}_1$ may not be measurable with respect to the natural Borel σ-field induced by the τ^{W_0}-topology on $\mathcal{M}_1$, we will instead consider the (smaller) σ-field $\mathcal{F}$, defined as the smallest σ-field that makes all the maps below measurable:

$$\nu \mapsto \int F\, d\nu, \qquad \text{for real-valued } F \in L_\infty^{W_0}. \tag{61}$$

Theorem 5.1 (LDP for Empirical Measures) *Suppose that $\mathbf{\Phi}$ satisfies* (DV3+) *with an unbounded function W. Then, for any initial condition $\Phi(0) = x$, the sequence of empirical measures $\{L_n\}$ satisfies the LDP in the space $(\mathcal{M}_1, \mathcal{F})$ equipped with the τ^{W_0}-topology, with the good, convex rate function*

$$I(\nu) := \inf \check{P} H(\nu \odot \check{P} \| \nu \odot P) \tag{62}$$

where the infimum is over all transition kernels $\check{P}$ for which ν is an invariant measure, and $\nu \odot \check{P}$ denotes the bivariate measure $[\nu \odot \check{P}](dx, dy) := \nu(dx)\check{P}(x, dy)$ on $(\mathsf{X}\times\mathsf{X}, \mathcal{B}\times\mathcal{B})$: Writing $\mu_{n,x}$ for the law of the empirical measure L_n under the initial condition $\Phi(0) = x$, then for any $E \in \mathcal{F}$,

$$\begin{aligned} -\inf_{\nu\in E^o} I(\nu) \;&\le\; \liminf_{n\to\infty} \tfrac{1}{n}\log \mu_{n,x}(E) \\ &\le\; \limsup_{n\to\infty} \tfrac{1}{n}\log \mu_{n,x}(E) \;\le\; -\inf_{\nu\in\bar{E}} I(\nu)\,, \end{aligned}$$

where E^o and $\bar{E}$ denote the interior and the closure of E in the τ^{W_0} topology, respectively.

The proof is based on an application of the Dawson-Gärtner projective limit theorem along the same lines as the proof of Theorem 6.2.10 in [12]. The main two technical ingredients are provided by, first, the multiplicative mean ergodic theorem Theorem 3.1 (iii) which, as noted in (36), shows that the log-moment generating functions converge to Λ. And second, by the regularity properties of Λ and the identification of Λ^* in terms of relative entropy, established in Section 4 and Section C of the Appendix.

As in Section 4, in order to identify the rate function for the LDP we find it easier to consider the bivariate chain $\mathbf{\Psi}$. Recall the bivariate extensions of our earlier definitions from equations (51), (52), (53) and (54).

Proof of Theorem 5.1. We begin by establishing an LDP for $\mathbf{\Phi}$ with rate function given by Λ^*. Recall that Proposition 3.6 gives

$$\Lambda_n(F) := \tfrac{1}{n} \log \mathsf{E}_x\Big[\exp\Big(n\langle L_n, F\rangle\Big)\Big] \to \Lambda(F)\,, \qquad n \to \infty. \tag{63}$$

In order to apply the projective limit theorem we need to extend the domain of the convex dual functional Λ^* as follows. For probability measures $\nu \in \mathcal{M}_1^{W_0}$, $\Lambda^*(\nu)$ is defined in (50), and the same definition applies when ν is a probability measure not necessarily in $\mathcal{M}_1^{W_0}$. More generally, let L' denote the algebraic dual of the space $L = L_\infty^{W_0}$, consisting of all linear functionals $\Theta : L \to \mathbb{R}$, and equipped with the weakest topology that makes the functional

$$\Theta \mapsto \Theta(F) = (\Theta, F) : L' \to \mathbb{R}$$

continuous, for each in $F \in L_\infty^{W_0}$. Note that each probability measure ν on $(\mathsf{X}, \mathcal{B})$ induces a linear functional $\Theta_\nu : L \to \mathbb{R}$ via

$$(\Theta_\nu, F) = \langle \nu, F\rangle = \int F d\nu.$$

Therefore, we can identify the space of probability measures $\mathcal{M}_1$ with the corresponding subset of L', and observe that the induced topology on $\mathcal{M}_1$ is simply the τ^{W_0}-topology.

Next, extend the definition of Λ^* to all $\Theta \in L'$ via

$$\Lambda^*(\Theta) = \sup\{(\Theta, F) - \Lambda(F) \;:\; F \in L_\infty^{W_0}\}\,, \tag{64}$$

and observe that [12, Assumption 4.6.8] is satisfied by construction (with $\mathcal{W} = L = L_\infty^{W_0}$, $\mathcal{X} = L'$ and $\mathcal{B} = \mathcal{F}$), and that by Proposition 4.3 the function $\Lambda(F_0 + \alpha F)$ is Gateaux differentiable. Therefore, we can apply the Dawson-Gärtner projective limit theorem [12, Corollary 4.6.11 (a)] to obtain that the sequence of empirical measures $\{L_n\}$ satisfy the LDP in the space L' with respect to the convex, good rate function Λ^*. Moreover, since by Proposition C.9 we know that $\Lambda^*(\Theta) = \infty$ for $\Theta \notin \mathcal{M}_1$, we obtain the same LDP in the space $(\mathcal{M}_1, \mathcal{F})$, with respect to the induced topology, namely, the τ^{W_0}-topology; see, e.g., [12, Lemma 4.1.5].

Next note that, in view of Proposition 4.1, the bivariate chain $\mathbf{\Psi}$ also satisfies the same LDP. But in this case, we claim that can express $\Lambda^*(\Gamma)$ for any bivariate probability measure Γ as follows:

$$\Lambda^*(\Gamma) = \begin{cases} H(\Gamma \| \Gamma_1 \odot P)\,, & \text{if the two marginals } \Gamma_1 \text{ and } \Gamma_2 \text{ of } \Gamma \text{ agree;} \\ \infty\,, & \text{otherwise.} \end{cases}$$

To see this, first consider the case when $\Gamma_1 \neq \Gamma_2$; then Theorem 4.2 (ii) and Proposition C.10 imply that $\Lambda^*(\Gamma) = \infty$. Suppose now that $\Gamma_1 = \Gamma_2$. Then Proposition C.10 shows that $\Lambda^*(\Gamma)$ must equal $H(\Gamma \| \Gamma_1 \odot P)$ whenever $\Lambda^*(\Gamma) = \infty$. And if the marginals agree and $\Lambda^*(\Gamma)$ is finite, then the identification follows form Theorem 4.2 (iii).

Finally, an application of the contraction principle [12, Theorem 4.2.1] implies that the univariate convex dual $\Lambda^*(\nu)$ coincides with $I(\nu)$ in (62). Simply note that the τ^{W_0}-topology on the space of probability measures is Hausdorff, and that the map $\Gamma \mapsto \Gamma_1$ is continuous in that topology. $\square$

Theorem 5.1 strengthens the "local" large deviations of [32] to a full LDP. The assumptions under which this LDP is proved are more restrictive that those in [32], but apparently they cannot be significantly relaxed. In particular, the density assumption of (DV3+) (ii) cannot be removed, as illustrated by the counter-example given in [18]. This example is of an irreducible, aperiodic Markov chain with state space $\mathsf{X} = [0,1]$, satisfying Doeblin's condition. It can be easily seen that this Markov chain satisfies condition (DV3) with Lyapunov function $V(x) = -\frac{1}{2}\log x$, $x \in [0,1]$, and with W given by

$$W(x) := \begin{cases} 2 - \log\left\{\frac{4\sqrt{x}}{1-2x}\left[\sqrt{\frac{3}{4}-\frac{x}{2}} - \sqrt{\frac{1}{4}+\frac{x}{2}}\right]\right\} & \text{for } x \in [0,1/2); \\ 2 - \log(2\sqrt{x}) & \text{for } x \in [1/2,1]. \end{cases}$$

Taking $\delta = 1$, $C = [0,1]$ and $b = 2$ yields a solution to (DV3), with the Lyapunov function V and the *unbounded* function W as above. But for this Markov chain the density assumption in (DV3+) (ii) is *not* satisfied, and as shown in [18], it satisfies the LDP with a rate function different from the one in Theorem 5.1.

The LDP of Theorem 5.1 can easily be extended to the sequence of empirical measures of k-tuples $L_{n,k}$, defined for each $k \geq 2$ by

$$L_{n,k} := \frac{1}{n}\sum_{t=0}^{n-1} \delta_{(\Phi(t),\Phi(t+1),\ldots,\Phi(t+k-1))}, \qquad n \geq 1\,. \tag{65}$$

We write $\mathcal{M}_{1,k}$ for the space of all probability measures on $(\mathsf{X}^k, \mathcal{B}^k)$, and we let $\mathcal{F}_k$ denote the σ-field of subsets of $\mathcal{M}_{1,k}$ defined analogously to $\mathcal{F}$ in (61), with X^k in place of X, and with real-valued functions F in the space

$$L_{\infty,k}^{W_0} := \left\{F : \mathsf{X}^k \to \mathbb{C} \,:\, \|F\|_{W_0} := \sup_{(x_1,\ldots,x_k)\in\mathsf{X}^k} \left(\frac{|F(x_1,\ldots,x_k)|}{W_0(x_1)+\cdots+W_0(x_k)}\right) < \infty\right\}$$

instead of $L_\infty^{W_0}$. Similarly, the $\tau_k^{W_0}$-topology on $\mathcal{M}_{1,k}$ is defined by the system of neighborhoods

$$N_F^k(c,\delta) := \left\{\nu \in \mathcal{M}_{1,k} : |\nu(F) - c| < \delta\right\}, \quad \text{for real-valued } F \in L_{\infty,k}^{W_0},\ c \in \mathbb{R},\ \delta > 0\,.$$

A straightforward generalization of the argument in the above proof yields the following corollary. The proof is omitted.

Corollary 5.2 *Under the assumptions of Theorem 5.1, for any initial condition $\Phi(0) = x$, the sequence of empirical measures $\{L_{n,k}\}$ satisfies the LDP in the space $(\mathcal{M}_{1,k}, \mathcal{F}_k)$ equipped with the $\tau_k^{W_0}$-topology, with the good, convex rate function*

$$I_k(\nu_k) = \begin{cases} H(\nu_k \| \nu_{k-1} \odot P), & \textit{if } \nu \textit{ is shift-invariant} \\ \infty, & \textit{otherwise.} \end{cases}$$

where ν_{k-1} denotes the first $(k-1)$-dimensional marginal of ν_k.

Next we show that under the assumptions of Theorem 5.1 it is possible to obtain exact large deviations results for the partial sums S_n,

$$S_n := \sum_{t=0}^{n-1} F(\Phi(t)) = \langle L_n, F \rangle \,, \qquad n \geq 1, \tag{66}$$

of a real-valued functional $F \in L_\infty^{W_0}$. In the next two theorems we prove analogs of the corresponding expansions of Bahadur and Ranga Rao for the partial sums of independent random variables [1]. Our results generalize those obtained by Miller [36] for finite state Markov chains, and those in [32] proved for geometrically ergodic Markov processes but only in a neighborhood of the mean; see [32] for further bibliographical references.

First we note that, since for any $F \in L_\infty^{W_0}$ the map $\nu \mapsto \langle \nu, F \rangle$ from $\mathcal{M}_1$ to $\mathbb{R}$, is continuous under the τ^{W_0} topology, we can apply the contraction principle to obtain an LDP for the partial sums $\{S_n\}$ in (66): Their laws satisfy the LDP on $\mathbb{R}$ with respect to the good, convex rate function $J(c)$ as in (19),

$$\begin{aligned} J(c) &= \inf\big\{ I(\nu) : \nu \text{ is a probability measure on } (\mathsf{X}, \mathcal{B}) \text{ satisfying } \nu(F) \geq c \big\} \\ &= \inf\Big\{ H(\Gamma \| \Gamma_1 \odot P) : \Gamma \in \mathcal{M}_{1,2} \text{ with marginals } \Gamma_1 = \Gamma_2 \text{ such that } \Gamma_1(F) \geq c \Big\}. \end{aligned}$$

Alternatively, based on (the weak version of) the multiplicative mean ergodic theorem in (63), we can apply the Gärtner-Ellis theorem [12, Theorem 2.3.6] to conclude that the laws of the partials sums $\{S_n\}$ satisfy the LDP on $\mathbb{R}$ with respect to the good rate function $J^*(c)$,

$$J^*(c) := \sup_{a \in \mathbb{R}} \left[ac - \Lambda(aF) \right], \qquad c \in \mathbb{R}\,, \tag{67}$$

so that, in particular, $J(c) = J^*(c)$ for all c.

Now suppose for simplicity that the function F has zero mean $\pi(F) = 0$ and nontrivial central limit theorem variance $\sigma^2(F) > 0$; recall the definition of $\sigma^2(F)$ from Section 3.1. To evaluate the supremum in (67), we recall from Lemma 2.10 that $\Lambda(aF)$ is convex in $a \in \mathbb{R}$, and since by Theorem 3.1 it is also analytic, it is strictly convex. Therefore, if we define

$$F_{\max} := \lim_{a \to \infty} \frac{d}{da} \Lambda(aF) = \sup_{a \in \mathbb{R}} \frac{d}{da} \Lambda(aF),$$

then $J^*(c) = \infty$ for values of c larger than $F_{\max}$, and the probabilities of the large deviations events $\{S_n \geq nc\}$ decay to zero *super*-exponentially fast.

Therefore, from now on we concentrate on the interesting range of values $0 < c < F_{\max}$. Note that, although in the case of independent and identically distributed random variables it is easy to identify $F_{\max}$ as the right endpoint of the support of F, for Markov chains this need not be the case, as illustrated by the following example.

Example. Let $\boldsymbol{\Phi} = \{\Phi(n) : n \geq 0\}$ be a discrete-time version of the Ornstein-Uhlenbeck process in $\mathbb{R}^2$, with $\Phi(0) = x \in \mathbb{R}^2$ and,

$$\Phi(n+1) = \begin{pmatrix} \Phi_1(n+1) \\ \Phi_2(n+1) \end{pmatrix} = \begin{bmatrix} 0 & 1 \\ -a_2 & -a_1 \end{bmatrix} \Phi(n) + \begin{pmatrix} 0 \\ N(n+1) \end{pmatrix},$$

where $\{N(k)\}$ is a sequence of independent and identically distributed $N(0,1)$ random variables. Let A denote the above 2-by-2 matrix, and assume that the roots of the quadratic equation $z^2 + a_1 z + a_2 = 0$ lie within the open unit disk in $\mathbb{C}$.

Note that there exists $\gamma < 1$ and a positive definite matrix P satisfying, $A^{\scriptscriptstyle T} P A \le \gamma I$. One may take $P = \sum_0^\infty \gamma^{-k} (A^k)^{\scriptscriptstyle T} A^k$, where $\gamma < 1$ is chosen so that the sum is convergent.

Then $\mathbf{\Phi}$ satisfies (DV3+) (i) with Lyapunov function $V(x) = 1 + \epsilon x^{\scriptscriptstyle T} P x$, and $W = V$, for suitably small $\epsilon > 0$ (hence, the drift condition (DV4) also holds). Condition (DV3+) (ii) holds with $T_0 = 2$ since $P^2(x, \cdot\,)$ has a Gaussian distribution with full-rank covariance.

Consider the functions

$$F_+(x) = \mathbb{I}_{\{|x_1|<1\}}, \quad F_0(x) = x_2 - x_1, \quad F(x) = F_1(x) + F_0(x), \qquad x = (x_1, x_2)^{\scriptscriptstyle T} \in \mathbb{R}^2.$$

The asymptotic variance of F_0 is zero, and for any initial condition we have

$$\sum_{t=0}^{n-1} F(\Phi(t)) = \sum_{t=0}^{n-1} F_+(\Phi(t)) + [\Phi_2(n-1) - x_1]\,.$$

We conclude that $F_{\max} = (F_+)_{\max} = 1$, although $\pi\{F > c\} > 0$ for each $c \ge 0$ under the invariant distribution π.

Recall form Section 3.1 the definitions of lattice and non-lattice functionals.

Theorem 5.3 (Exact Large Deviations for Non-Lattice Functionals) *Suppose that $\mathbf{\Phi}$ satisfies* (DV3+) *with an unbounded function W, and that $F \in L_\infty^{W_0}$ is a real-valued, strongly-non-lattice functional, with $\pi(F) = 0$ and $\sigma^2(F) \neq 0$. Then, for any $0 < c < F_{\max}$ and all $x \in \mathsf{X}$,*

$$\mathsf{P}_x\{S_n \ge nc\} \;\sim\; \frac{\check{f}_a(x)}{a\sqrt{2\pi n \sigma_a^2}} e^{-nJ(c)}, \quad n \to \infty,$$

where $a > 0$ is the unique solution of the equation $\frac{d}{da}\Lambda(aF) = c$, $\sigma_a^2 := \frac{d^2}{da^2}\Lambda(aF) > 0$, $\check{f}_a(x)$ is the eigenfunction constructed in Theorem 3.1, and $J(c)$ is defined in (19). A corresponding result holds for the lower tail.

The proof of Theorem 5.3 is identical to that of the corresponding result in [32], based on the following simple properties of a Markov chain satisfying (DV3+). We omit properties P5 and P6 since they are not needed here.

Properties. Suppose $\mathbf{\Phi}$ satisfies (DV3+) with an unbounded function W, and choose and fix an arbitrary $x \in \mathsf{X}$ and a function $F \in L_\infty^{W_0}$ with zero asymptotic mean $\pi(F) = 0$ and nontrivial asymptotic variance $\sigma^2 = \sigma^2(F) \neq 0$. Let S_n denote the partial sums in (66) and write $m_n(\alpha)$ for the moment generating functions

$$m_n(\alpha) := \mathsf{E}_x[\exp(\alpha S_n)] = \mathsf{E}_x[\exp(\alpha\langle L_n, F\rangle)]\,, \qquad n \ge 1,\ \alpha \in \mathbb{C}. \tag{68}$$

The proofs of the following properties are exactly as those of the corresponding results in [32], and are based primarily on the multiplicative mean ergodic theorem Theorem 3.1, and the Taylor expansion of $\Lambda(F)$ given in Proposition 4.3. Observe that by Theorem 2.2 we have that the Lyapunov function V in (DV3+) satisfies $\pi(V^2) < \infty$.

P1. For any $m > 0$ there is $\overline{a} > m$, $\overline{\omega} > 0$ and a sequence $\{\epsilon_n\}$ such that

$$m_n(\alpha) = \exp(n\Lambda(\alpha F))[\check{f}_\alpha(x) + |\alpha|\epsilon_n], \quad n \geq 1,$$

and $|\epsilon_n| \to 0$ exponentially fast as $n \to \infty$, uniformly over all $\alpha \in \Omega(\overline{a}, \overline{\omega})$, with $\Omega(\overline{a}, \overline{\omega})$ as in Theorem 3.1.

P2. If F is strongly non-lattice, then for any $m > 0$ and any $0 < \omega_0 < \omega_1 < \infty$, there is $\overline{a} > m$ and a sequence $\{\epsilon'_n\}$ such that

$$m_n(\alpha) = \exp(n\Lambda(aF))\epsilon'_n, \quad n \geq 1,$$

and $|\epsilon'_n| \to 0$ exponentially fast as $n \to \infty$, uniformly over all $\alpha = a + i\omega$ with $|a| \leq \overline{a}$ and $\omega_0 \leq |\omega| \leq \omega_1$.

P3. If F is lattice (or almost lattice) with span $h > 0$, then for any $\epsilon > 0$, as $n \to \infty$,

$$\sup_{\epsilon \leq |\omega| \leq 2\pi/h - \epsilon} |m_n(i\omega)| \to 0 \qquad \text{exponentially fast.}$$

P4. For any $m > 0$ there exist $\overline{a} > m$ and $\overline{\omega} > 0$ such that the function $\Lambda(\alpha F)$ is analytic in $\alpha \in \Omega(\overline{a}, \overline{\omega})$, and for $\alpha = a \in \mathbb{R}$ we have $\Lambda(aF)|_{a=0} = \frac{d}{da}\Lambda'(af)|_{a=0} = 0$, and $\frac{d^2}{da^2}\Lambda''(aF)|_{a=0} = \sigma^2 > 0$. Moreover, $\sigma_a^2 := \frac{d^2}{da^2}\Lambda(aF)$ is strictly positive for real $a \in [-\overline{a}, \overline{a}]$.

P7. For each $m > 0$ there exist $\overline{a} > m$ and $\overline{\omega} > 0$ such that the eigenfunction $\check{f}_\alpha$ is analytic in $\alpha \in \Omega(\overline{a}, \overline{\omega})$, it satisfies $\check{f}_\alpha\big|_{\alpha=0} \equiv 1$, and it is strictly positive for real α. Moreover, there is some $\overline{\omega}_0 \in (0, \overline{\omega})$ such that

$$\delta(i\omega) := |\log \check{f}_{i\omega}(x) - i\omega\widehat{F}(x)| \leq (\text{Const})\omega^2,$$

for all $|\omega| \leq \overline{\omega}_0$, where $\widehat{F}$ is as in Theorem 1.1.

An analogous asymptotic expansion for lattice functionals is given in the next theorem; again, its proof is omitted as it is identical to that of the corresponding result in [32].

Theorem 5.4 (Exact Large Deviations for Lattice Functionals) *Suppose* $\mathbf{\Phi}$ *satisfies* (DV3+) *with an unbounded function* W*, and that* $F \in L_\infty^{W_0}$ *is a real-valued, lattice functional with span* $h > 0$*,* $\pi(F) = 0$ *and* $\sigma^2(F) \neq 0$*. Let* $\{c_n\}$ *be a sequence of real numbers in* (ϵ, ∞) *for some* $\epsilon > 0$*, and assume (without loss of generality) that, for each* n*,* c_n *is in the support of* S_n*. Then, for all* $x \in \mathsf{X}$*,*

$$\mathsf{P}_x\{S_n \geq nc_n\} \sim \frac{h}{(1 - e^{-ha_n})\sqrt{2\pi n\Lambda''_n(a_n)}} e^{-nJ_n(c_n)}, \quad n \to \infty, \tag{69}$$

where $\Lambda_n(a)$ *is the log-moment generating function of* S_n*,*

$$\Lambda_n(a) := \log \mathsf{E}_x\Big[e^{aS_n}\Big], \qquad n \geq 1,\, a \in \mathbb{R},$$

each $a_n > 0$ is the unique solution of the equation $\frac{d}{da}\Lambda_n(a) = c_n$, and $J_n(c)$ is the convex dual of $\Lambda_n(a)$,

$$J_n(c) := \Lambda_n^*(c) := \sup_{\lambda\in\mathbb{R}}[\lambda c - \Lambda_n(\lambda)]\,, \qquad n \geq 1,\, c \in \mathbb{R}\,.$$

A corresponding result holds for the lower tail.

Observe that the expansion (69) in the lattice case is slightly more general than the one in Theorem 5.3. If the sequence $\{c_n\}$ converges to some $c > \epsilon$ as $n \to \infty$, then, as in [32], the a_n also converge to some $a > 0$, and

$$\mathsf{P}_x\{S_n \geq nc_n\} \;\sim\; \frac{h\check{f}_a(x)}{(1-e^{-ha})\sqrt{2\pi n\sigma_a^2}}e^{-nJ(c)}, \quad n \to \infty,$$

where $\sigma_a^2 := \frac{d^2}{da^2}\Lambda(aF)$.

Acknowledgments

The authors would like to express their thanks to Amir Dembo and Jamal Najim for several interesting pointers in the literature, and also to Tom Kurtz, Jin Feng and Luc Rey-Bellet for sharing their unpublished work.

Appendix

A Drift Conditions and Multiplicative Regularity

Lemma A.1 allows us to bound the expansive term $b\mathbb{I}_C(x)$ in condition (DV3). We say that a set $S \in \mathcal{B}$ is *multiplicatively-special* (m.-special) if for every $A \in \mathcal{B}^+$ there exists $\eta > 0$ such that

$$\sup_{x\in\mathsf{X}} \mathsf{E}_x\Big[\exp\Big(\eta\tau_A L_{\tau_A}(S)\Big)\Big] < \infty\,.$$

Lemma A.1 *If $\mathbf{\Phi}$ is ψ-irreducible, then every small set is m.-special.*

Proof. Let S be a small set, and fix $A \in \mathcal{B}^+$. For a given fixed $T > 0$, define the stopping times $\{T_n : n \geq 0\}$ inductively via $T_0 = 0$, and

$$T_{n+1} = \inf\Big\{t \geq T_n + T : \Phi(t) \in S\Big\}, \quad n \geq 0\,.$$

We consider the sequence of functions,

$$g_n(x) = \mathsf{E}_x\Big[\exp\Big(\eta\int_{[0,\tau_A\wedge T_n)} \mathbb{I}_S(\Phi(t))\,dt\Big)\Big], \qquad n \geq 1\,,$$

and we let $B_n = B_n(\eta) = \sup_{x\in\mathsf{X}} g_n(x)$, $n \geq 1$. Since S is small, there exists $\epsilon > 0$, $T > 0$, such that $\mathsf{P}_x\{\tau_A > T_1\} \leq 1 - \epsilon < 1$ for all $x \in S$. From the strong Markov property we then have,

$$
\begin{aligned}
g_{n+1}(x) &= \mathsf{E}_x\Big[\exp\Big(\eta \textstyle\int_{[0,\tau_A\wedge T_{n+1})} \mathbb{I}_S(\Phi(t))\,dt\Big)\Big] \\
&\le e^{\eta T}\mathsf{E}_x\Big[\mathbb{I}(\tau_A \le T_1)\Big] \\
&\quad +\mathsf{E}_x\Big[\exp\Big(\eta \textstyle\int_{[0,T_1)} \mathbb{I}_S(\Phi(t))\,dt\Big)\mathsf{E}_{\Phi(T_1)}\Big[\exp\Big(\eta \textstyle\int_{[0,\tau_A\wedge T_n)} \mathbb{I}_S(\Phi(t))\,dt\Big)\Big]\mathbb{I}(\tau_A > T_1)\Big] \\
&\le e^{\eta T} + (e^{\eta T}\sqrt{B_1(2\eta)\mathsf{P}\{\tau_A > T_1\}})B_n \le e^{\eta T} + (e^{\eta T}\sqrt{B_1(2\eta)(1-\epsilon)})B_n\,,
\end{aligned}
$$

for all $x \in S$, where the last bound uses Cauchy-Schwartz.

This gives an upper bound for $x \in S$, and the same bound also holds for all x since $g_n(x) \le \sup_{y\in S} g_n(y)$. Choosing $\eta > 0$ so small that $\rho := \Big(e^{\eta 2T}B_1(2\eta)(1-\epsilon)\Big)^{1/2} < 1$, we see from induction that $\{B_n\}$ is a bounded sequence, and lim $\sup_{n\to\infty} B_n \le (1-\rho)^{-1}e^{\eta T}$. $\square$

Proof of Theorem 2.5.

Recall that, under (DV3), the stochastic process $(m(t), \mathcal{F}_t)$ given in (9) is a super-martingale. That is, for any stopping time τ,

$$
\mathsf{E}_x[m(\tau)] \le m(0) = v(x), \qquad x \in \mathsf{X}. \tag{70}
$$

Fix any set $A \in \mathcal{B}^+$. An application of Lemma A.1 implies that there exist constants $b_1, b_2 < \infty$, and $\eta_1 > 0$ such that for any stopping time τ,

$$
\mathsf{E}_x\Big[\exp\Big(\sum_{s=0}^{\tau-1}[\eta_1\mathbb{I}_C(\Phi(s)) - b_1\mathbb{I}_A(\Phi(s))]\Big)\Big] \le \exp(b_2)\,. \tag{71}
$$

From (70), Jensen's inequality, and Hölder's inequality, for all sufficiently small $\eta > 0$, and all finite $b_3 > 0$,

$$
\begin{aligned}
&\mathsf{E}_x\Big[\exp\Big(\eta V(\Phi(\tau)) + \eta\sum_{s=0}^{\tau-1}[W(\Phi(s)) - b_3\mathbb{I}_A(\Phi(s))]\Big)\Big] \\
&= \mathsf{E}_x\Big[\exp\Big(\eta V(\Phi(\tau)) + \eta\sum_{s=0}^{\tau-1}[W(\Phi(s)) - \tfrac{1}{2}b\mathbb{I}_C(\Phi(s))]\Big)\exp\Big(\eta\sum_{s=0}^{\tau-1}[\tfrac{1}{2}b\mathbb{I}_C(\Phi(s)) - b_3\mathbb{I}_A(\Phi(s))]\Big)\Big] \\
&\le \mathsf{E}_x\Big[\exp\Big(2\eta V(\Phi(\tau)) + 2\eta\sum_{s=0}^{\tau-1}[W(\Phi(s)) - b\mathbb{I}_C(\Phi(s))]\Big)\Big]^{\frac{1}{2}} \\
&\quad \times \mathsf{E}_x\Big[\exp\Big(2\eta\sum_{s=0}^{\tau-1}[\tfrac{1}{2}b\mathbb{I}_C(\Phi(s)) - b_3\mathbb{I}_A(\Phi(s))]\Big)\Big]^{\frac{1}{2}} \\
&\le e^{\eta V(x)}\mathsf{E}_x\Big[\exp\Big(2\eta\sum_{s=0}^{\tau-1}[b\mathbb{I}_C(\Phi(s)) - b_3\mathbb{I}_A(\Phi(s))]\Big)\Big]^{\frac{1}{2}}.
\end{aligned}
$$

Setting $b_3 = b_1\eta_1^{-1}$ we obtain from this and (71), for all $\eta < \eta_1(2b)^{-1}$,

$$\mathsf{E}_x\Big[\exp\Big(\eta V(\Phi(\tau)) + \eta\sum_{s=0}^{\tau-1}[W(\Phi(s)) - b_3\mathbb{I}_A(\Phi(s))]\Big)\Big] \le v_\eta(x)\exp(2\eta b b_2\eta_1^{-1}), \qquad x \in \mathsf{X}. \quad (72)$$

Setting $\tau = \tau_A \wedge m$ for $m \ge 1$, and then letting $m \to \infty$ completes the proof. $\square$

Proof of Theorem 2.2.

The construction of a Lyapunov function V_* follows from the bounds given above, beginning with (72) (note however that $W \equiv 1$ under (DV2)). Assume that the set $A \in \mathcal{B}^+$ is fixed, with V bounded on A. We assume moreover that A is small – this is without loss of generality by [34, Proposition 5.2.4 (ii)]. Fix $k \ge 0$, and define,

$$\sigma_A := \min\{i \ge 0 : \Phi(i) \in A\}, \qquad \tau := \sigma_A \wedge k.$$

Consideration of this stopping time in (72) gives the upper bound, for some $b_1 < \infty$,

$$\mathsf{E}_x\Big[\mathbb{I}(\sigma_A \ge k)\exp(\eta V(\Phi(k)) + \tfrac{1}{2}\eta k)\Big] \le b_1 v_\eta(x)e^{-\frac{1}{2}\eta k}, \qquad x \in \mathsf{X},\ k \ge 0,$$

and on summing both sides we obtain the pair of bounds,

$$v_\eta(x) \le V_*(x) := \mathsf{E}_x\Big[\sum_{k=0}^{\sigma_A}\exp(\eta V(\Phi(k)) + \tfrac{1}{2}\eta k)\Big] \le \Big(\frac{b_1}{1-e^{-\frac{1}{2}\eta}}\Big)v_\eta(x), \qquad x \in \mathsf{X}.$$

We now demonstrate that this function satisfies the desired drift condition: We have,

$$PV_*(x) = \mathsf{E}_x\Big[\sum_{k=1}^{\tau_A}\exp(\eta V(\Phi(k)) + \tfrac{1}{2}\eta k)\Big] \le e^{-\frac{1}{2}\eta}V_*(x) + b'\mathbb{I}_A(x),$$

with $b' = \big(\frac{b_1}{1-e^{-\frac{1}{2}\eta}}\big)\sup_{y\in A} v_\eta(y)$. This is indeed a version of (V4). $\square$

Proposition A.2 *Suppose that X is σ-compact and locally compact; that P has the Feller property; and that there exists a sequence of compact sets $\{K_n : n \ge 1\}$ satisfying (27): For any compact set $K \subset \mathsf{X}$,*

$$\sup_{x\in K}\mathsf{E}_x[e^{n\tau_{K_n}}] < \infty.$$

Then, there exists a solution to the inequality,

$$\mathcal{H}(V) \le -\tfrac{1}{2}W + b\mathbb{I}_C$$

such that $V, W\colon \mathsf{X} \to [1,\infty)$ are continuous, their sublevel sets are precompact, $C \in \mathcal{B}$ is compact, and $b < \infty$.

Proof. Let $\{O_n : n \geq 1\}$ denote a sequence of open, precompact sets satisfying $O_n \uparrow \mathsf{X}$, and $K_n \subset$ closure of $O_n \subset O_{n+1}$, $n \geq 1$. For each $n \geq 1$ we consider a continuous function $s_n \colon \mathsf{X} \to [0,1]$ satisfying $s_n(x) = 1$ for $x \in O_n$, and $s_n(x) = 0$ for $x \in O_{n+1}^c$. We then define a stopping time $\tau_n \geq 1$ through the conditional distributions,

$$\mathsf{P}\{\tau_n > n \mid \mathcal{F}_n\} = \prod_{i=1}^{n} \bigl(1 - s_n(\Phi(i))\bigr), \qquad n \geq 1.$$

From the conditions imposed on s_n we may conclude that $\tau_{K_n} \geq \tau_{O_n} \geq \tau_n \geq \tau_{O_{n+1}}$ for each $n \geq 1$.

For $n \geq 1$, $m \geq 1$ we define $V_{n,m} \colon \mathsf{X} \to \mathbb{R}_+$ by,

$$V_{n,m}(x) := \log \mathsf{E}_x \Bigl[\exp\Bigl(\sum_{i=0}^{\tau_n - 1} \bigl(n-1\bigr)\bigl(1 - s_m(\Phi(i))\bigr)\Bigr)\Bigr], \qquad x \in \mathsf{X}.$$

Continuity of this function is established as follows: First, observe that under the Feller property we can infer that $\mathsf{P}_x\{\tau_n = k\}$ is a continuous function of $x \in \mathsf{X}$ for any $k \geq 1$. The bound $\tau_n \leq \tau_{K_n}$, $n \geq 1$, combined with (27) then establishes a form of uniform integrability sufficient to infer the desired continuity.

Moreover, by the dominated convergence theorem we have $V_{n,m}(x) \downarrow 0$, $m \to \infty$, for each $x \in \mathsf{X}$. Continuity implies that this convergence is uniform on compacta. We choose $\{m_n : n \geq 1\}$ so that $V_{n,m_n}(x) \leq 1$ on O_{n+1}, and we define $V_n = V_{n,m_n}$. Letting $W_n = \bigl(n-1\bigr)\bigl(1 - s_m\bigr)$, we obtain the bound $\mathcal{H}(V_n) \leq -W_n + 1$. Let $\{p_n\} \subset \mathbb{R}_+$ satisfy $\sum_{n\geq 1} p_n = 1$, $\sum p_n n = \infty$, and define,

$$W := 1 + \sum_{n\geq 1} p_n W_n, \quad V := 1 + \sum_{n \geq 1} p_n V_n.$$

Convexity of $\mathcal{H}$ then gives, $\mathcal{H}(V) \leq V - W + 1$. The functions W and V are evidently coercive and continuous. Hence the desired inequality is obtained with $C = \{x \in \mathsf{X} : W(x) \leq \frac{1}{2}\}$. $\square$

B v-Separable Kernels

The following result is immediate from the definition (24).

Lemma B.1 *Suppose that $\{\widehat{P}^n : n \in \mathbb{Z}_+\}$ is a positive semigroup, with finite spectral radius $\hat{\xi} > 0$. Then the inverse $[Iz - \widehat{P}]^{-1}$ admits the power series representation,*

$$[Iz - \widehat{P}]^{-1} = \sum_{n=0}^{\infty} z^{-n-1} \widehat{P}^n, \qquad |z| > \hat{\xi},$$

where the sum converges in norm.

Lemma B.2 (i) is a simple corollary:

Lemma B.2 *Consider a positive semigroup $\{\widehat{P}^n : n \in \mathbb{Z}_+\}$ that is ψ-irreducible. Then:*

(i) *The spectral radius $\hat{\xi}$ in L_∞^v of $\{\widehat{P}^t\}$ satisfies $\hat{\xi} < b_0$ for a given $b_0 < \infty$ if and only if there is a $b < b_0$, and a function $v_1 \colon \mathsf{X} \to [1, \infty)$ such that v_1 equivalent to v, and $\widehat{P} v_1 \le b v_1$.*

(ii) *The generalized principal eigenvalue $\hat{\lambda}$ (see Section 2.4) satisfies $\hat{\lambda} \le b < \infty$ if and only if there is a measurable function $v_1 \colon \mathsf{X} \to (0, \infty)$ such that, $\widehat{P} v_1 \le b v_1$.*

Proof. Part (ii) is a consequence of [39, Theorem 5.1].

To see (i), suppose first that $b > \hat{\xi}$, and set $v_1 = b[Ib - \widehat{P}]^{-1} v = \sum_{n=0}^\infty b^{-n} \widehat{P}^n v$. Then $v_1 \in L_\infty^v$ by Lemma B.1, and $v \le v_1$ by construction. Moreover, it is easy to see that v_1 satisfies the desired inequality.

Conversely, if the inequality holds then for any $0 < \eta < 1$, $n \ge 1$,

$$(\eta^{-1} b)^{-n-1} \widehat{P}^n v_1 \le b^{-1} \eta^{n+1} n v_1,$$

which shows that $\|\!|[I\eta^{-1} b - \widehat{P}]^{-1}|\!\|_{v_1} \le (1-\eta)^{-1} \eta b^{-1}$. It follows that $\hat{\xi} \le \eta^{-1} b$ since v and v_1 are equivalent. Since $\eta < 1$ is arbitrary, this shows that $b \ge \hat{\xi}$, and completes the proof. $\square$

The following result will be used below to construct v-separable kernels.

Lemma B.3 *Suppose that $\widehat{P}$ is a positive kernel, and that there is a measure $\mu \in \mathcal{M}_1^v$ satisfying*

$$\widehat{P}(x, A) \le \mu(A)\,, \qquad x \in \mathsf{X},\ A \in \mathcal{B}\,.$$

Then $\widehat{P}^2$ is v-separable.

Proof. Consider the bivariate measure, $\Gamma(dx, dy) = \mu(dx) \widehat{P}(x, dy) v(y)$, for $x, y \in \mathsf{X}$. Under the assumptions of the proposition we have the upper bound, $\Gamma(dx, dy) \le v(y) \mu(dx) \mu(dy)$, and hence there exists a density r satisfying $r(x, y) \le v(y)$, $x, y \in \mathsf{X}$, and $\Gamma = r[\mu \times \mu]$. It follows that for any $g \in L_\infty^v$ we have

$$\widehat{P} g\,(x) = \int r(x, y) g(y) v^{-1}(y) \mu(dy)\,, \qquad a.e.\ x \in \mathsf{X}\ [\mu].$$

For a given $\epsilon > 0$ the function r can be approximated from below in $L_1(\mu \times \mu)$ by the simple functions,

$$r_\epsilon(x, y) = \sum_{i=1}^N \alpha_i \mathbb{I}_{A_i}(x) \mathbb{I}_{B_i}(y) \le r(x, y), \qquad x, y \in \mathsf{X},$$

and

$$\int\!\!\int |r(x, y) - r_\epsilon(x, y)| \mu(dx) \mu(dy) \le \epsilon\,.$$

We then define

$$\widehat{P}_\epsilon(x, dy) = r_\epsilon(x, y) v^{-1}(y) \mu(dy)\,, \quad x, y \in \mathsf{X}\,,$$

and $\widehat{P}_{\epsilon 2} := \widehat{P} \widehat{P}_\epsilon$. The latter kernel may be expressed $\widehat{P}_{\epsilon 2} = \sum s_i \otimes \nu_i$, with

$$s_i(x) := \alpha_i \widehat{P}(x, A_i)\,, \quad \nu_i(dy) = \mathbb{I}_{B_i}(y) v^{-1}(y) \mu(dy)\,, \quad x, y \in \mathsf{X}\,.$$

We have $s_i \in L_\infty^v$ and $\nu_i \in \mathcal{M}_1^v$ for each i.

For any $g \in L_\infty^v$, $x \in \mathsf{X}$, we then have,

$$\begin{aligned}
|\widehat{P}_{\epsilon 2} g\,(x) - \widehat{P}^2 g\,(x)| &= |\widehat{P}[\widehat{P}_\epsilon g - \widehat{P} g]\,(x)| \\
&= \Big|\int \widehat{P}(x,dy)\Big\{\int [r_\epsilon(y,z) - r(y,z)] g(z) v^{-1}(z)\mu(dz)\Big\}\Big| \\
&\le \int \mu(dy)\Big\{\int |r_\epsilon(y,z) - r(y,z)||g(z)| v^{-1}(z)\mu(dz)\Big\} \\
&\le \|g\|_v \int\int |r_\epsilon(y,z) - r(y,z)|\mu(dy)\mu(dz) \le \epsilon\|g\|_v\,. \qquad \square
\end{aligned}$$

Lemma B.4 *Suppose that* (DV3) *holds with W unbounded. Fix $0 < \eta \le 1$, and consider any measurable function F satisfying*

$$\begin{aligned}
F^+ := \max(F, 0) &\in L_\infty^W; \\
\lim_{r\to\infty} \|F^+ \mathbb{I}_{C_W(r)^c}\|_W &< \delta\eta.
\end{aligned} \tag{73}$$

We then have $|||I_{C_W(r)^c} P_f|||_{v_\eta} \to 0$, exponentially fast, as $r \to \infty$.

Proof. For simplicity we consider only $\eta = 1$. Choosing $r_0 \ge 1$ so that $\|F^+ \mathbb{I}_{C_W(r_0)^c}\|_W = \delta_0 < \delta$, we have,

$$P_f e^V \le e^{V - (\delta-\delta_0)W + b} \le e^{V - (\delta-\delta_0)r + b} \qquad \text{on } C_W(r_0)^c,$$

and hence $|||\mathbb{I}_{C_W(r)^c} P_f|||_v \le e^{-(\delta-\delta_0)r+b}$ for all $r \ge 1$. $\square$

Lemma B.5 *Suppose that* (DV3+) *holds with W unbounded. Fix $0 < \eta \le 1$, and consider any measurable function F satisfying (73).Then $(P_f)^{2T_0+2}$ is v_η-separable.*

Proof. For simplicity we present the proof only for $\eta = 1$. We define the truncation,

$$\widehat{P}_r := (\mathbb{I}_{C_W(r)} P_f)^{T_0+1}\,.$$

For each $r \ge 1$ we have

$$\widehat{P}_r(x, A) \le \beta_r'(A) := \int_{C_W(r)} \beta_r(dx) P_f(x, A) \qquad x \in \mathsf{X},\ A \in \mathcal{B}.$$

It then follows from Lemma B.3 that the kernel $\widehat{P}_r^2$ is v-separable.

Finally, applying Lemma B.4 we may conclude that $|||(P_f)^{2T_0+2} - \widehat{P}_r^2|||_v \to 0$, $r \to \infty$, which implies that $(P_f)^{2T_0+2}$ is also v-separable. $\square$

PROOF OF THEOREM 2.4.

(a) ⇒ (b). When (DV3) holds we can conclude from Lemma B.4 that $\|\!|P - I_{C_W(r)}P|\!\|_{v_0} \to 0$ as $r \to \infty$. It follows that $\|\!|P^T - I_{C_W(r)}P^T|\!\|_{v_0} \to 0$ as $r \to \infty$ for any $T \geq 1$. In particular, this holds for $T = T_0$. Under the separability assumption on $\{I_{C_W(r)}P^{T_0} : r \geq 1\}$ it then follows that P^{T_0} is v-separable.

(b) ⇒ (a). We first show that each of the sets $\{C_{v_0}(r) : r \geq 1\}$ is small. Under the assumptions of (b) we may find, for each $\epsilon > 0$, an integer $N \geq 1$, functions $\{s_i : 1 \leq i \leq N\} \subset L_\infty^{v_0}$, and probability measures $\{\nu_i : 1 \leq i \leq N\} \subset \mathcal{M}_1^{v_0}$ such that, with $K = \sum s_i \otimes \nu_i$,

$$\|\!|P^{T_0} - K|\!\|_{v_0} < \epsilon\,. \tag{74}$$

This gives for any $r \geq 1$,

$$|1 - \sum s_i(x)| = |P^{T_0}\mathbf{1}\,(x) - K\mathbf{1}\,(x)| \leq \epsilon v_0(x) \leq \epsilon r, \qquad x \in C_{v_0}(r).$$

Let $A \in \mathcal{B}$ be a small set with $\nu_i(A^c) < \epsilon$ for each i. From the bound above and using similar arguments,

$$\begin{aligned} P^{T_0}(x, A^c) &\leq K(x, A^c) + \epsilon v_0(x) \\ &\leq \textstyle\sum_i s_i(x)\nu_i(A^c) + \epsilon v_0(x) \\ &\leq (1 + \epsilon r)\epsilon + \epsilon r, \qquad x \in C_{v_0}(r)\,. \end{aligned}$$

It follows that for any $r \geq 1$, we may find a small set $A(r)$ such that $P^{T_0}(x, A(r)) \geq \frac{1}{2}$, for $x \in C_{v_0}(r)$. It then follows from [34, Proposition 5.2.4] that $C_{v_0}(r)$ is small.

We now construct a solution to the drift inequality in (DV3). Using finite approximations as in (74), we may construct, for each $n \geq 1$, an integer $r_n \geq n$ such that

$$\|\!|(PI_{C^c_{r_n}})^{T_0}|\!\|_{v_0} \leq \|\!|P^{T_0}I_{C^c_{r_n}}|\!\|_{v_0} \leq e^{-2nT_0}.$$

Since the norm is submultiplicative, this then gives the bound,

$$\|\!|(PI_{C^c_{r_n}})^k|\!\|_{v_0} \leq b_0 e^{-2nk}, \qquad k \geq 0\,,$$

where $b_0 := (\|\!|P|\!\|_{v_0})^{T_0}$.

We then define for each $n \geq 1$,

$$v_n = I_{C^c_{r_n}} \sum_{k=0}^{\infty} e^{kn}(PI_{C^c_{r_n}})^k v_0\,.$$

From the previous bound on $\|\!|(PI_{C^c_{r_n}})^k|\!\|_{v_0}$ we have the pair of bounds,

$$\|v_n\|_{v_0} \leq b_0 \frac{1}{1 - e^{-n}}, \quad \text{and} \qquad \|Pv_n\|_{v_0} \leq b_0 \frac{e^{-2n}}{1 - e^{-n}}\,. \tag{75}$$

Finally, we set

$$\begin{aligned} V &:= \log\Big(1 + \textstyle\sum_{n=1}^{\infty} v_n\Big) \\ W &:= b\mathbb{I}_C - \mathcal{H}(V), \end{aligned}$$

where $C = C_v(r)$ for some r, and the constants b and r are chosen so that $W(x) \geq 1$ for all $x \in \mathsf{X}$. The bounds (75) together with the lower bound $v_n \geq v_0 e^n \mathbb{I}_{C^c_{r_n}}$ imply that

$$\lim_{r\to\infty} \inf_{x \in C_v(r)^c} \exp(-\mathcal{H}(V)) = \lim_{r\to\infty} \inf_{x \in C_v(r)^c} \frac{e^{V(x)}}{(Pe^V)(x)} = \infty,$$

which implies the existence of r and b satisfying these requirements. ☐

In much of the remainder of the appendix we replace (DV3+) with the following more general condition:

(i) The Markov process $\mathbf{\Phi}$ is ψ-irreducible, aperiodic, and it satisfies condition (DV3) with some Lyapunov function $V\colon \mathsf{X} \to [1,\infty)$, and an unbounded function $W\colon \mathsf{X} \to [1,\infty)$. (76)

(ii) There exists $T_0 > 0$ such that $I_{C_W(r)} P^{T_0}$ is v-separable for for each $r < \infty$.

Theorem 2.4 states that this is roughly equivalent to (DV3+) with an unbounded function W. In fact, we do have an analogous upper bound for P^{T_0}:

Lemma B.6 *Suppose that the conditions of (76) hold. Then, for each $r \geq 1$, $\epsilon > 0$, there is a positive measure $\beta_{r,\epsilon} \in \mathcal{M}_1^v$ such that*

$$P^{T_0} h\,(x) \leq \beta_{r,\epsilon}(h) + \epsilon \|h\|_v\,, \qquad x \in C_W(r),\ h \in L_\infty^v.$$

Proof. We apply the approximation (74) used in the proof of Theorem 2.4, where $\{s_i : 1 \leq i \leq N\} \subset L_\infty^{v_0}$ are non-negative valued, and $\{\nu_i : 1 \leq i \leq N\} \subset \mathcal{M}_1^{v_0}$ are probability measures. We may assume that the $\{s_i\}$ satisfy the bound $1 = P^{T_0}(x, \mathsf{X}) \geq \sum s_i(x) - 1$, $x \in C_W(r)$, and it follows that we may take $\beta_{r,\epsilon} = 2\sum_{i=1}^N \nu_i$. ☐

The following result is proven exactly as Lemma B.5, using Lemma B.6.

Lemma B.7 *Suppose that the conditions of (76) hold. Fix $0 < \eta \leq 1$, and consider any $F \in L_\infty^W$ satisfying (73). Then $(P_f)^{2T_0}$ is v_η-separable.*

C Properties of Λ and Λ^*

In this section we obtain additional properties of Λ and Λ^*. One of the main goals is to establish approximations of $\Lambda(G)$ through bounded functions when G is possibly unbounded. Similar issues are treated in [13, Chapter 5] where a tightness condition is used to provide related approximations.

Lemma C.1 *For a ψ-irreducible Markov chain:*

(i) *The log-generalized principal eigenvalue Λ is convex on the space of measurable functions $F\colon \mathsf{X} \to (-\infty,\infty]$.*

(ii) *The log-spectral radius Ξ is convex on the space of measurable functions $F\colon \mathsf{X} \to (-\infty, \infty]$.*

Proof. The proofs of (i) and (ii) are similar, and both proofs are based on Lemma B.2. We provide a proof of (ii) only.

Fix $F_1, F_2 \in L_\infty^{W_0}$, $\eta, \theta \in (0,1)$, and let $b_i = \eta^{-1}\xi(F_i)$, $i = 1,2$. Lemma B.2 implies that there exists functions $\{v_1, v_2\}$ equivalent to v, and satisfying

$$\mathsf{E}_x\big[\exp\big(F_i(\Phi(0)) + V_i(\Phi(1))\big)\big] := P_{f_i} v_i\,(x) \le b_i v_i(x), \qquad i = 1,2,\ x \in \mathsf{X}.$$

We then define

$$F_\theta = \theta F_1 + (1-\theta) F_2, \quad V_\theta = \theta V_1 + (1-\theta) V_2,$$

so that by Hölder's inequality,

$$\begin{aligned} P_{f_\theta} v_\theta\,(x) &= \mathsf{E}_x\big[\exp\big(\theta[F_1(\Phi(0)) + V_1(\Phi(1))] + (1-\theta)[F_2(\Phi(0)) + V_2(\Phi(1))]\big)\big] \\ &\le \mathsf{E}_x\big[\exp\big(F_1(\Phi(0)) + V_1(\Phi(1))\big)\big]^\theta \mathsf{E}_x\big[\exp\big(F_2(\Phi(0)) + V_2(\Phi(1))\big)\big]^{1-\theta} \\ &\le b_1^\theta b_2^{1-\theta} v_\theta(x)\,, \qquad x \in \mathsf{X}. \end{aligned}$$

The function v_θ is equivalent to v. Consequently, we may apply Lemma B.2 once more to obtain that $\xi(F_\theta) \le b_1^\theta b_2^{1-\theta}$. Taking logarithms then gives,

$$\Xi(F_\theta) \le \theta \log(b_1) + (1-\theta)\log(b_2) = \theta \Xi(F_1) + (1-\theta)\Xi(F_2) - \log(\eta).$$

This completes the proof since $0 < \eta < 1$ is arbitrary. ☐

The following result establishes a form of upper semi-continuity for the functional Λ.

Lemma C.2 *Suppose that $\mathbf{\Phi}$ is ψ-irreducible, and consider a sequence $\{F_n\}$ of measurable, real-valued functions on X. Suppose there exists a measurable function $F\colon \mathsf{X} \to \mathbb{R}$ such that $F_n \uparrow F$, as $n \uparrow \infty$. Then the corresponding generalized principal eigenvalues converge: $\Lambda(F_n) \to \Lambda(F)$, as $n \uparrow \infty$.*

Proof. It is obvious that $\limsup_{n\to\infty} \Lambda(F_n) \le \Lambda(F)$. To complete the proof we establish a bound on the limit infimum.

Under the assumptions of the proposition we have $P_{f_n}^T \ge P_{f_1}^T$, for any $T \ge 1$, $n \ge 1$. It follows that we can find an integer $T_0 \ge 1$, a function $s\colon \mathsf{X} \to [0,1]$, and a probability ν on $\mathcal{B}$ satisfying $\psi(s) > 0$ and

$$P_{f_n}^{T_0} \ge s \otimes \nu\,, \qquad 1 \le n \le \infty.$$

Let $(\check{f}_n, \lambda_n)$ denote the Perron-Frobenius eigenfunction and generalized principal eigenvalue for P_{f_n}, normalized so that $\nu(h_n) = 1$ for each n. For each $n \ge 1$ we have the upper bound, $P_{f_n}\check{f}_n \le \lambda_n \check{f}_n$. This gives a lower bound on the $\{\check{f}_n\}$:

$$\check{f}_n \ge \lambda_n^{-T_0} P_{f_n}^{T_0} \check{f}_n \ge \lambda_n^{-T_0} \nu(\check{f}_n) s = \lambda_n^{-T_0} s.$$

Let $h = \liminf_{n\to\infty} \check{f}_n$, $\lambda = \liminf_{n\to\infty} \lambda_n$. Then, by Fatou's Lemma, $P_f h \le \lambda h$. We also have $\nu(h) \le 1$ by Fatou's Lemma, and the lower bound $h \ge \lambda^{-T_0} s$. It follows from Lemma B.2 that $\Lambda(F) \le \log(\lambda)$. ☐

In applying Lemma C.2 we typically assume that suitable regularity conditions hold so that $\Xi(F) = \Lambda(F)$. Under a finiteness assumption alone we obtain a complementary continuity result for certain classes of decreasing sequences of functions. One such result is given here:

Lemma C.3 *Suppose that $\|P\|_v < \infty$, and that $F\colon \mathsf{X} \to \mathbb{R}$ is measurable, with $\Xi(F_+) < \infty$. Then, with $F_n := \max(F, -n)$ we have, $\Xi(F_n) \downarrow \Xi(F)$, as $n \uparrow \infty$.*

Proof. This follows immediately from the approximation, $\|P_{f_n} - P_f\|_v \leq e^{-n}\|P\|_v$, $n \geq 1$. $\square$

To establish a tight approximation for $\Lambda(M)$, where $M = \log m$ is as in the proof Theorem 4.2, we will approximate M by bounded functions.

Proposition C.4 *Suppose that $\|P\|_v < \infty$, and that $F\colon \mathsf{X} \to \mathbb{R}$ is measurable, with $\Xi(F) < \infty$, and $\Lambda(F) = \Xi(F)$. Then, there exists a sequence $\{n_k : k \geq 1\}$ such that with $F_k := F\mathbb{I}\{-n_k \leq F \leq k\}$ we have:*

$$\Lambda(F_k) \to \Lambda(F) \quad \textit{and} \quad \Xi(F_k) \to \Xi(F) \quad \textit{as } k \to \infty\,.$$

Proof. Let $F_k^0 := F\mathbb{I}\{F \leq k\}$. From Lemma C.2 we have $\Lambda(F_k^0) \uparrow \Lambda(F)$, $k \to \infty$. It follows that we also have $\Xi(F_k^0) \uparrow \Xi(F)$, $k \to \infty$, since Ξ dominates Λ.

We now apply Lemma C.3: For each $k \geq 1$ we may find $n_k \geq 1$ such that with $F_k := F\mathbb{I}\{-n_k \leq F \leq k\}$,

$$\begin{aligned}
\Lambda(F_k^0) &\leq \Lambda(F_k) \leq \Lambda(F_k^0) + k^{-1}, \\
\Xi(F_k^0) &\leq \Xi(F_k) \leq \Xi(F_k^0) + k^{-1}, \qquad k \geq 1\,.
\end{aligned}$$

$\square$

The following proposition implies that Λ is tight in a strong sense under (DV3+):

Proposition C.5 *Suppose that the conditions of (76) hold. Then, for any increasing sequence of measurable sets $K_n \uparrow \mathsf{X}$, and any $G \in L_\infty^{W_0}$,*

$$\text{(i)} \qquad \lim_{n\to\infty} \Lambda(G\mathbb{I}_{K_n^c}) = 0$$

$$\text{(ii)} \qquad \lim_{n\to\infty} \Lambda(G\mathbb{I}_{K_n}) = \Lambda(G)$$

The proof is postponed until after the following lemma.

Lemma C.6 *Suppose that the conditions of (76) hold, and consider any increasing sequence of measurable sets $K_n \uparrow \mathsf{X}$, and any $G \in L_\infty^{W_0}$. Then, on letting $g_n = \exp(I_{K_n^c} G)$, $n \geq 1$, we have*

$$\|P^{T_0} P_{g_n} - P^{T_0+1}\|_v \to 0\,, \qquad n \to \infty.$$

Proof. We may assume without loss of generality that $G \geq 0$. As usual, we set $g = e^G$.

Under (76) we have $\|P_{g_n}\|_v \leq \|P_g\|_v < \infty$, $n \geq 1$. Consequently, given Lemma B.4, it is enough to show that for any $r \geq 1$,

$$\|I_{C_W(r)}[P^{T_0} P_{g_n} - P^{T_0+1}]\|_v \to 0\,, \qquad n \to \infty.$$

To see this, observe that for any $h \in L_\infty^v$, $x \in \mathsf{X}$,

$$\begin{aligned} \bigl|I_{C_W(r)}[P^{T_0} P_{g_n} - P^{T_0+1}]h\,(x)\bigr| &= \bigl|I_{C_W(r)}[P^{T_0} I_{K_n{}^c}[P_g - P]]h\,(x)\bigr| \\ &\le I_{C_W(r)} P^{T_0} I_{K_n{}^c}[P_g - P]|h|\,(x) \\ &\le \|h\|_v \|\!|P_g|\!\|_v (I_{C_W(r)} P^{T_0} I_{K_n^c}) v\,(x) \\ &\le \|h\|_v \|\!|P_g|\!\|_v [\beta_{r,\epsilon}(\mathbb{I}_{K_n^c} v) + \epsilon v]\,, \end{aligned}$$

where the measure $\beta_{r,\epsilon} \in \mathcal{M}_1^v$ is given in Lemma B.6. Consequently,

$$\limsup_{n\to\infty} \|\!|I_{C_W(r)}[P^{T_0} P_{g_n} - P^{T_0+1}]|\!\|_v \le \epsilon \|\!|P_g|\!\|_v .$$

This proves the result since $\epsilon > 0$ is arbitrary. □

Proof of Proposition C.5. To see (i), consider any $G \in L_\infty^{W_0}$, and any sequence of measurable sets $K_n \uparrow \mathsf{X}$. We assume without loss of generality that $G \ge 0$.

Fix any $b > 1$, and define for $n \ge 1$, $G_n = (T_0 + 1) b \mathbb{I}_{K_n^c} G$. In view of Lemma C.6, given any $\Lambda > 0$, we may find $n \ge 1$ such that the spectral radius of the semigroup generated by the kernel $\widehat{P}_n := P^{T_0} P_{g_n}$ satisfies $\xi_n < e^\Lambda$. With n, Λ fixed, we then have for some $b_n < \infty$, $\widehat{P}_n^k v \le b_n e^{k\Lambda} v$ for $k \ge 1$. This has the sample path representation,

$$\mathsf{E}_x\Bigl[\exp\Bigl(\sum_{i=1}^k G_n(\Phi((T_0+1)i - 1))\Bigr) v\bigl((T_0+1)k)\bigr)\Bigr] \le b_n e^{k\Lambda} v(x), \qquad x \in \mathsf{X}, k \ge 1.$$

Denote by $h_{0,k}(x)$ the expectation on the left hand side. We then have, for each $j \ge 1$,

$$h_{j,k}(x) := P^j h_{0,k}(x) \le b_n e^{k\Lambda} (\|\!|P|\!\|_v)^j v(x)\,, \qquad x \in \mathsf{X}.$$

Moreover, each of these functions has a sample path representation,

$$h_{j,k}(x) = \mathsf{E}_x\Bigl[\exp\Bigl(\sum_{i=1}^k G_n(\Phi(j-1+(T_0+1)i))\Bigr) v\bigl(\Phi(j+(T_0+1)k)\bigr)\Bigr]\,, \quad x \in \mathsf{X},\ j \ge 1,\ k \ge 1.$$

We then obtain the following bound using Hölder's inequality,

$$\begin{aligned} &\mathsf{E}_x\Bigl[\exp\Bigl(\sum_{i=T_0}^{(T_0+1)(k+1)-1} b \mathbb{I}_{K_n^c}(\Phi(i)) G(\Phi(i))\Bigr) v\bigl(\Phi((T_0+1)(k+1))\bigr)\Bigr] \\ &\le \Bigl(\textstyle\prod_{j=0}^{T_0} \mathsf{E}_x\Bigl[\exp\Bigl(\sum_{i=1}^k G_n(\Phi(j-1+(T_0+1)i))\Bigr) v\bigl(\Phi((T_0+1)(k+1))\bigr)\Bigr]\Bigr)^{(T_0+1)^{-1}} \\ &\le \|\!|P|\!\|_v \Bigl(\textstyle\prod_{j=0}^{T_0} \mathsf{E}_x\Bigl[\exp\Bigl(\sum_{i=1}^k G_n(\Phi(j-1+(T_0+1)i))\Bigr) v\bigl(\Phi(j+(T_0+1)k)\bigr)\Bigr]\Bigr)^{(T_0+1)^{-1}} \\ &= \|\!|P|\!\|_v \Bigl(\textstyle\prod_{j=0}^{T_0} h_{j,k}(x)\Bigr)^{(T_0+1)^{-1}} \\ &\le b_n (\|\!|P|\!\|_v)^{T_0+1} e^{k\Lambda} v(x), \qquad x \in \mathsf{X}, k \ge 1. \end{aligned}$$

We conclude that $\Lambda(\mathbb{I}_{K_n^c} G) \le \Lambda/(T_0+1)$. Since $\Lambda > 0$ is arbitrary, it follows that $\Lambda(\mathbb{I}_{K_n^c} G) \to 0$, $n \to \infty$.

To see (ii), fix $\theta \in (0,1)$, and obtain the following bound using convexity,

$$\begin{aligned}\Lambda(\theta G) &= \Lambda\big(\theta \mathbb{I}_{K_n} G + (1-\theta)\theta(1-\theta)^{-1}\mathbb{I}_{K_n^c} G\big) \\ &\leq \theta\Lambda\big(\mathbb{I}_{K_n} G\big) + (1-\theta)\Lambda\big(\theta(1-\theta)^{-1}\mathbb{I}_{K_n^c} G\big)\,.\end{aligned}$$

From (i) we conclude that

$$\Lambda(\theta G) \leq \theta \liminf_{n\to\infty} \Lambda\big(\mathbb{I}_{K_n} G\big), \quad 0<\theta<1,$$

which gives $\Lambda(G) \leq \liminf_{n\to\infty} \Lambda\big(\mathbb{I}_{K_n} G\big)$. To obtain the reverse inequality we argue similarly:

$$\Lambda\big(\mathbb{I}_{K_n} G\big) \leq \theta\Lambda\big(\theta^{-1} G\big) + (1-\theta)\Lambda\big(-(1-\theta)^{-1}\mathbb{I}_{K_n^c} G\big),$$

which shows that

$$\limsup_{n\to\infty} \Lambda\big(\mathbb{I}_{K_n} G\big) \leq \theta\Lambda\big(\theta^{-1} G\big), \quad 0<\theta<1.$$

This shows that $\Lambda\big(\mathbb{I}_{K_n} G\big) \to \Lambda\big(G\big)$ as claimed. $\square$

Proposition C.5 allows us to broaden the class of functions for which Ξ is finite-valued.

Proposition C.7 *Suppose that the conditions of (76) hold. Then, there exists $W_1\colon \mathsf{X} \to [1,\infty)$ satisfying the following:*

(i) $W_0 \in L_\infty^{W_1}$, *and* $W_1 \in L_\infty^V$;

(ii) $\sup\{V(x) : x \in C_{W_1}(r)\} < \infty$ *for each* $r \geq 1$;

(iii) $\Xi(W_1) < \infty$.

If the state space X is σ-compact, then we may assume that W_1 is also coercive.

Proof. Fix a sequence of measurable sets satisfying $K_n \uparrow \mathsf{X}$, with $\sup_{x\in K_n} V(x) < \infty$ for each n. Proposition C.5 implies that we may find, for each $k \geq 1$, an integer $n_k \geq 1$ such that $\Xi(2^{k+1}\mathbb{I}_{K_{n_k}^c} W_0) \leq 1$. We then define

$$W_1 = \Big(1 + \sum_{k=1}^{\infty} \mathbb{I}_{K_{n_k}^c}\Big) W_0.$$

The functional Ξ is convex by Lemma C.1, which gives the bound,

$$\Xi(W_1) \leq \tfrac{1}{2}\Xi(2W_0) + \sum_{k=1}^{\infty} 2^{-k-1}\Xi(2^{k+1}\mathbb{I}_{K_{n_k}^c} W_0) \leq \tfrac{1}{2}(1+\Xi(2W_0)) < \infty.$$

To see that $W_1 \in L_\infty^V$ we apply Lemma 2.9.

Finally, if X is σ-compact, then the $\{K_n\}$ may be taken to be compact sets, which then implies the coercive property for W_1. $\square$

We have the following useful corollary. The proof is routine, given Proposition C.7 and Proposition B.2 (i); see also [2, Theorem 2.4].

Lemma C.8 *Suppose that the conditions of (76) hold. Then, for any $N < \infty$, there exists $r_0 \geq 1$, $b_0 < \infty$, such that with $\tau = \tau_{C_V(r_0)}$,*

$$\mathsf{E}_x\big[\exp\big(N\tau\big)\big] < b_0 e^{V(x)}, \qquad x \in \mathsf{X}.$$

We now turn to properties of the dual functional Λ^* defined in (64). The continuity results stated in Proposition C.5 lead to the following representation.

Proposition C.9 *Suppose that the conditions of (76) hold. Let Θ be a linear functional on $L^{W_0}_{\infty,2}$ satisfying $\Lambda^*(\Theta) < \infty$. Then Θ may be represented as,*

$$\langle \Theta, G\rangle = \nu(G)\,, \qquad G \in L^{W_0}_{\infty},$$

where $\nu \in \mathcal{M}^{W_0}_1$ is a probability measure.

Proof. We proceed in several steps, making repeated use of the bound,

$$\langle \Theta, G\rangle \leq \Lambda^*(\Theta) + \Lambda(G) < \infty\,, \qquad G \in L^{W_0}_{\infty}. \tag{77}$$

First note that on considering constant functions in (77) we obtain,

$$\Lambda^*(\Theta) \geq \sup_c [\langle \Theta, c\rangle - \Lambda(c)] = \sup_{c\in\mathbb{R}} [\langle \Theta, 1\rangle - 1]c.$$

It is clear that finiteness of Λ^* implies that $\langle \Theta, 1\rangle = 1$. Next, consider any $G\colon \mathsf{X} \to \mathbb{R}_+$ with $G \in L^{W_0}_{\infty}$. Then, since $\Lambda(cG) \leq 0$ for $c \leq 0$,

$$\Lambda^*(\Theta) \geq \sup_c [\langle \Theta, cG\rangle - \Lambda(cG)] \geq \sup_{c<0} \langle \Theta, G\rangle c.$$

We conclude that $\langle \Theta, G\rangle \geq 0$ for $G \geq 0$.

Consider now a set $A \in \mathcal{B}$ of ψ-measure zero. Then $\Lambda(c\mathbb{I}_A) = 0$ for any $c \geq 0$, and we can argue as above using (77) that $\infty > \Lambda^*(\Theta) \geq \sup_{c>0}\langle \Theta, \mathbb{I}_A\rangle c$, which shows that $\langle \Theta, \mathbb{I}_A\rangle = 0$.

Finally, we demonstrate that Θ defines a countably additive set function on $\mathcal{B}$. Let $\{A_i\} \subset \mathcal{B}$ denote disjoint sets, and let $G_n = \sum_{i=n+1}^{\infty} \mathbb{I}_{A_i}$. Then $0 \leq G_n \leq 1$ everywhere, and $G_n \downarrow 0$. Proposition C.5 implies that $\Lambda(bG_n) \to 0$, $n \to \infty$, for any $b \in \mathbb{R}$. Consequently,

$$\begin{aligned} \Lambda^*(\Theta) &\geq \limsup_{n\to\infty} [\Theta(bG_n) - \Lambda(bG_n)] \\ &= b \limsup_{n\to\infty} \Theta(G_n)\,. \end{aligned}$$

It follows that $\limsup_{n\to\infty} \Theta(G_n) = 0$, which implies that Θ defines a countably additive set function, so that Θ is in fact a probability measure. $\square$

More generally, we define Λ^* for bivariate probability measures Γ not necessarily in $\mathcal{M}_{1,2}^{W_0}$ using the same definition as in (54). Recall from Lemma 4.11 that the two marginals of Γ agree whenever $\Lambda^*(\Gamma) < \infty$. Proposition C.10 provides further structure.

Proposition C.10 *For any probability measure* Γ *on* $(\mathsf{X} \times \mathsf{X}, \mathcal{B} \times \mathcal{B})$ *with first and second marginal equal to some* $\check{\pi}$*,*

$$\Lambda^*(\Gamma) \leq H(\Gamma \,\|\, \check{\pi} \odot P)\,, \tag{78}$$

and, moreover,

$$\Lambda^*(\Gamma) = H(\Gamma \,\|\, \check{\pi} \odot P) = \infty \quad \textit{for } \Gamma \notin \mathcal{M}_{1,2}^{W}. \tag{79}$$

Proof. If we view W as a function on $\mathsf{X} \times \mathsf{X}$ with $W(x,y) \equiv W(x)$, $x, y \in \mathsf{X}$, then we have the bound, for all $\epsilon > 0$, $n \geq 1$,

$$\epsilon\langle \Gamma, W \wedge n\rangle \leq \Lambda(\epsilon W \wedge n) + \Lambda^*(\Gamma) \leq \Lambda(\epsilon W) + \Lambda^*(\Gamma).$$

Lemma B.5 shows that $\Lambda(\epsilon W) < \infty$ for $\epsilon > 0$ sufficiently small, and this gives (79).

Define $\check{P}$ through the decomposition $\Gamma = \check{\pi} \odot \check{P}$, and let $\check{\mathsf{E}}$ denote the expectation for the Markov chain with transition kernel $\check{P}$. We assume that $\check{P}$ is of the form

$$\check{P}(x, dy) = m(x,y) P(x, dy), \qquad x, y \in \mathsf{X},$$

and set $M = \log(m)$, since otherwise the relative entropy is infinite and there is nothing to prove. We then have, for any $G \in L_{\infty,2}^{W_0}$,

$$\begin{aligned}
\Lambda(G) &= \lim_{T\to\infty} \tfrac{1}{T} \log\Big(\mathsf{E}_x\Big[\exp(T\langle L_T, G\rangle)\Big]\Big) \\
&= \lim_{T\to\infty} \tfrac{1}{T} \log\Big(\check{\mathsf{E}}_x\Big[\exp(T\langle L_T, G - M\rangle)\Big]\Big) \\
&\geq \limsup_{T\to\infty} \tfrac{1}{T} \check{\mathsf{E}}_x\Big[T\langle L_T, G - M\rangle\Big] && \text{(Jensen's inequality)} \\
&= \langle \Gamma, G - M\rangle \qquad a.e.\ x \in \mathsf{X}\ [\check{\pi}]\,, && \text{(mean ergodic theorem for } \check{P})
\end{aligned}$$

where the application of the mean ergodic theorem is justified by the f-norm ergodic theorem [34, Theorem 14.0.1]. The integrability conditions required in this result are obtained as follows. First, recall that $\Gamma(|G|) < \infty$ when $\Lambda^*(\Gamma)$ is finite and $G \in L_\infty^{W_0}$. Also, as in the proof that $H(\Gamma \,\|\, \check{\pi} \odot P) \geq 0$, one can show that $\Gamma(M_-) < \infty$, where $M_- := |M \wedge 0|$. Consequently, $(M - G)_-$ is Γ-integrable, which is what is required in the mean ergodic theorem.

The above bound may be interpreted as,

$$H(\Gamma \,\|\, \check{\pi} \odot P) = \langle \Gamma, M\rangle \geq \langle \Gamma, G\rangle - \Lambda(G).$$

Taking the supremum over all $G \in L_{\infty,2}^{W_0}$ gives (78). □

References

[1] R.R. Bahadur and R. Ranga Rao. On deviations of the sample mean. *Ann. Math. Statist.*, 31:1015–1027, 1960.

[2] S. Balaji and S.P. Meyn. Multiplicative ergodicity and large deviations for an irreducible Markov chain. *Stochastic Process. Appl.*, 90(1):123–144, 2000.

[3] R.F. Bass and P. Hsu. Some potential theory for reflecting Brownian motion in Hölder and Lipschitz domains. *Ann. Probab.*, 19(2):486–508, 1991.

[4] E. Bolthausen. Markov process large deviations in τ-topology. *Stochastic Process. Appl.*, 25(1):95–108, 1987.

[5] A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein. Metastability and small eigenvalues in Markov chains. *J. Phys. A*, 33(46):L447–L451, 2000.

[6] S. Boyd and L. Vandenberghe. *Convex Optimization.* In preparation, Department of Electrical Engineering, Stanford University, 2002.

[7] N.R. Chaganty and J. Sethuraman. Strong large deviation and local limit theorems. *Ann. Probab.*, 21(3):1671–1690, 1993.

[8] E.B. Davies. Pseudospectra of differential operators. *J. Operator Theory*, 43:243–262, 2000.

[9] D.A. Dawson and J. Gärtner. Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. *Stochastics*, 20(4):247–308, 1987.

[10] A. de Acosta. Large deviations for empirical measures of Markov chains. *J. Theoret. Probab.*, 3(3):395–431, 1990.

[11] Michael Dellnitz and Oliver Junge. On the approximation of complicated dynamical behavior. *SIAM J. on Numerical Analysis*, 36(2):491–515, 1999.

[12] A. Dembo and O. Zeitouni. *Large Deviations Techniques And Applications.* Springer-Verlag, New York, second edition, 1998.

[13] J.D. Deuschel and D.W. Stroock. *Large Deviations.* Academic Press, Boston, 1989.

[14] M.D. Donsker and S.R.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.

[15] M.D. Donsker and S.R.S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time. III. *Comm. Pure Appl. Math.*, 29(4):389–461, 1976.

[16] M.D. Donsker and S.R.S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time. IV. *Comm. Pure Appl. Math.*, 36(2):183–212, 1983.

[17] D. Down, S. P. Meyn, and R. L. Tweedie. Exponential and uniform ergodicity of Markov processes. *Ann. Probab.*, 23(4):1671–1691, 1995.

[18] P. Dupuis and O. Zeitouni. A nonstandard form of the rate function for the occupation measure of a markov chain. *Stochastic Process. Appl.*, 61:249–261, 1996.

[19] P. Eichelsbacher and U. Schmock. Large deviations for products of empirical measures of dependent sequences. *Markov Process. Related Fields*, 7(3):435–468, 2001.

[20] J. Feng. Martingale problems for large deviations of Markov processes. *Stochastic Process. Appl.*, 81:165–212, 1999.

[21] J. Feng and T.G. Kurtz. Large deviations for stochastic processes. *Preprint*, 2000.

[22] W.H. Fleming. Exit probabilities and optimal stochastic control. *App. Math. Optim.*, 4:329–346, 1978.

[23] M. Fukushima and D. Stroock. Reversibility of solutions to martingale problems. In *Probability, statistical mechanics, and number theory*, pages 107–123. Academic Press, Orlando, FL, 1986.

[24] P. Groeneboom, J. Oosterhoff, and F.H. Ruymgaart. Large deviation theorems for empirical probability measures. *Ann. Probab.*, 7(4):553–586, 1979.

[25] A. Hordijk and F. Spieksma. On ergodicity and recurrence properties of a Markov chain with an application to an open Jackson network. *Adv. in Appl. Probab.*, 24(2):343–376, 1992.

[26] W. Huisinga, S.P. Meyn, and C. Schuette. Phase transitions and metastability in Markovian and molecular systems. *Ann. Appl. Probab.*, 2001. to appear.

[27] I. Iscoe, P. Ney, and E. Nummelin. Large deviations of uniformly recurrent Markov additive processes. *Adv. Appl. Math.*, 6:373–412, 1985.

[28] N.C. Jain. Large deviation lower bounds for additive functionals of Markov processes. *Ann. Probab.*, 18(3):1071–1098, 1990.

[29] J.L. Jensen. Saddlepoint expansions for sums of Markov dependent variables on a continuous state space. *Probab. Theory Related Fields*, 89(2):181–199, 1991.

[30] N.V. Kartashov. Criteria for uniform ergodicity and strong stability of Markov chains with a common phase space. *Theor. Probability Appl.*, 30:71–89, 1985.

[31] N.V. Kartashov. Inequalities in theorems of ergodicity and stability for Markov chains with a common phase space. *Theor. Probability Appl.*, 30:247–259, 1985.

[32] I. Kontoyiannis and S.P. Meyn. Spectral theory and limit theorems for geometrically ergodic Markov processes. *Ann. Appl. Probab.*, 13:304–362, February 2003.

[33] H. Kunita. *Stochastic Flows and Stochastic Differential Equations*. Cambridge University Press, Cambridge, 1990.

[34] S.P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability.* Springer-Verlag, London, 1993.

[35] S.P. Meyn and R.L. Tweedie. Computable bounds for geometric convergence rates of Markov chains. *Ann. Appl. Probab.*, 4(4):981–1011, 1994.

[36] H.D. Miller. A convexivity property in the theory of random variables defined on a finite Markov chain. *Ann. Math. Statist.*, 32:1260–1270, 1961.

[37] P. Ney and E. Nummelin. Markov additive processes. I. Eigenvalue properties and limit theorems. *Ann. Probab.*, 15(2):561–592, 1987.

[38] P. Ney and E. Nummelin. Markov additive processes II. Large deviations. *Ann. Probab.*, 15(2):593–609, 1987.

[39] E. Nummelin. *General Irreducible Markov Chains and Nonnegative Operators.* Cambridge University Press, Cambridge, 1984.

[40] R.G. Pinsky. *Positive Harmonic Functions and Diffusion.* Cambridge University Press, Cambridge, 1995.

[41] L. Rey-Bellet and L. E. Thomas. Asymptotic behavior of thermal nonequilibrium steady states for a driven chain of anharmonic oscillators. *Comm. Math. Phys.*, 215:1–24, 2000.

[42] L. Rey-Bellet and L. E. Thomas. Fluctuations of the entropy production in anharmonic chains. *Ann. Henri Poincaré*, 3(3):483–502, 2002.

[43] S.R.S. Varadhan. *Large Deviations and Applications.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa., 1984.

[44] L. Wu. Large and moderate deviations and exponential convergence for stochastic damping Hamiltonian systems. *Stochastic Process. Appl.*, 91(2):205–238, 2001.