# Equivalences of Geometric Ergodicity of Markov Chains

by (in alphabetical order)

Marco A. Gallegos-Herrada, David Ledvinka, and Jeffrey S. Rosenthal

*Departments of Statistics and Mathematics, University of Toronto*

(March 2022; last revised April 9, 2022.)

**Abstract**

This paper gathers together different conditions which are all equivalent to geometric ergodicity of time-homogeneous Markov chains on general state spaces. A total of 33 different conditions are presented (26 for general chains plus 7 just for reversible chains), some old and some new, in terms of such notions as convergence bounds, drift conditions, spectral properties, etc., with different assumptions about the distance metric used, finiteness of function moments, initial distribution, uniformity of bounds, and more. Proofs of the connections between the different conditions are provided, mostly self-contained but using some results from the literature where appropriate.

## 1 Introduction

The increasing importance of Markov chain Monte Carlo (MCMC) algorithms (see e.g. [2] and the many references therein) has focused attention on the rate of convergence of (time-homogeneous) Markov chains to their stationary distribution. While it is most useful to have explicit quantitative bounds on the distance to stationarity (see e.g. [26, 12] and the references therein), qualitative convergence bounds are often more feasible to obtain. The most commonly-used qualitative convergence property is geometric ergodicity, i.e. exponentially fast convergence to stationarity, which has been widely studied (e.g. [28, 17, 22]).

In addition to fast convergence, geometric ergodicity also guarantees a Markov chain Central Limit Theorem (CLT), i.e. the convergence of scaled sums of functional values to a fixed normal distribution, for all functionals with finite $2+\delta$ moments [8, Theorem 18.5.3] (see also [7]), or even just $2^{\text{nd}}$ moments assuming reversibility [21]. Such CLTs are helpful for understanding the errors which arise from Monte Carlo estimation (see e.g. [28, 24, 11]). However, geometric ergodicity and CLTs do not hold for all Markov chains nor all MCMC algorithms (see e.g. [20] and [22, Theorem 22]).

For certain types of MCMC algorithms, geometric ergodicity is fairly well understood. For example, it is known that an Independence Sampler is geometrically ergodic if and only if its proposal density is bounded below by a constant multiple of the target density [15], and that the popular Random-Walk Metropolis algorithm is geometrically ergodic essentially if and only if its target distribution has exponentially light tails [16, 24]. However, for many other complicated Markov chains and MCMC algorithms, geometric ergodicity is not clear.

One promising way of establishing geometric ergodicity is to show that some other properties of Markov chains imply it, or are even equivalent to it. This has been shown, by [28, 17, 21, 25] and others, for properties such as drift conditions, spectral bounds, and more. However, such relationships are scattered throughout the literature, are not always stated in full generality, and are often presented as just one-way implications. In the current work, we present a total of 33 different conditions which are equivalent to geometric ergodicity for Markov chains on general state spaces (26 for general chains plus 7 just for reversible chains; some previously known and some new). We then provide proofs of all of the equivalences (mostly self-contained, though using known results where needed); see Figure 1.

To illustrate the flavour of the various equivalences, consider the following:

- The usual definitions of geometric ergodicity state that the Markov chain's *distance to stationarity* after $n$ iterations is bounded by a constant times $\rho^n$ for some $\rho < 1$. But what "distance" should be used: total variation, or $V$-norm, or $L^2(\pi)$? And, how does the "constant" depend on the starting state $X_0 = x$? Must those constants have finite expected value with respect to $\pi$? What about finite $j^{\text{th}}$ moments?
- If the initial state $X_0$ is itself chosen from a non-degenerate *initial distribution* probability measure $\mu$, then will the convergence to stationarity still be geometric, at least if $\mu$ is, say, in $L^p(\pi)$?
- Geometric ergodicity is well-known to be implied by *drift conditions* of the form $PV(x) \leq \lambda V(x) + b\,\mathbb{1}_S(x)$ for some function $V : \mathcal{X} \to [1, \infty]$ and $\lambda < 1$ and $b < \infty$ and small set $S$. But are such drift conditions actually *equivalent* to geometric ergodicity? And, can the drift function $V$ be taken to have finite stationary mean? finite $j^{\text{th}}$ moment?
- Geometric ergodicity is also related to the Markov operator $P$ having a *spectral gap*. But as an operator on what space: $L^\infty_V$? for what function $V$? having which finite moments? And should the "gap" be identified by removing the eigenvalue 1 directly, or by subtracting off $\Pi$, or by restricting to the zero-mean space $L^\infty_{V,0}$?
- Geometric ergodicity is implied by the Markov operator *norm* being less than 1. But for which operator: $P$, or $P^m$ for some $m \in \mathbb{N}$? Regarded as an operator on $L^\infty_V$ or $L^\infty_{V,0}$? For

what choice of $V$? Having which finite moments?

- If the Markov chain is assumed to be *reversible*, so that the operator $P$ is self-adjoint on $L^2(\pi)$, then in most of the above conditions can the operator norm be taken to be $L^2(\pi)$?

We shall see that the answer to these questions is, essentially, "all of the above". That is, we shall state many different conditions, which cover essentially all of the above possibilities, and shall prove that they are all equivalent. In our desire to be thorough, we might have gone a bit overboard listing so many different conditions, including some which are just minor variations of each other. However, we believe that additional equivalent conditions can only help: the equivalences with weaker assumptions are easier to establish, while the equivalences with stronger assumptions are most useful for drawing conclusions or analysing further. We know from bitter experience that it can be very frustrating to discover a statement about geometric ergodicity which is almost, but not quite, exactly what is needed to finish a particular proof, and this has led us to adopt a "the more the merrier" attitude. The reader can, of course, choose to ignore all conditions which are not germaine to their work.

Basic definitions necessary to understand the conditions, such as total variance distance, $L_V^\infty$ norms, $L^p(\pi)$ spaces, reversibility, etc., are presented in Section 2. Then, in Section 3, all of the equivalent conditions are introduced (Theorem 1). Sections 4 through 8 are then devoted to proving all of the equivalences; see Figure 1 for a visual guide showing which implications are proved by which of our results. Our proofs are largely self-contained, but we do use known results in the literature (especially [17]) where needed. Finally, we close in Section 9 with some future directions and open problems (Q9.1 through Q9.8).

## 2 Definitions and Background

Throughout this paper, $\Phi = \{X_n\}_{n=0}^{\infty}$ is a discrete-time, time-homogeneous *Markov chain* on a general state space $\mathcal{X}$ equipped with a $\sigma$-algebra $\mathcal{F}$. And, $P$ is the corresponding Markov kernel, so that $P(x, A) = \mathbf{P}[X_n \in A \mid X_{n-1} = x]$ for all $x \in \mathcal{X}$ and $A \in \mathcal{F}$ and $n \in \mathbb{N}$. The *kernel* $P$ acts to the left on (possibly signed) measures, and to the right on functions, by:

$$(\mu P)(A) = \int P(x, A)\, \mu(dx), \qquad (Pf)(x) = \int f(y)\, P(x, dy).$$

The *higher-order transitions* are then defined inductively by:

$$P^n(x, A) \;=\; \int_{\mathcal{X}} P(x, dy)\, P^{n-1}(y, A), \qquad x \in \mathcal{X},\; A \in \mathcal{F},\; n \in \mathbb{N}.$$

We shall assume throughout $P$ has a *stationary distribution*, i.e. a probability distribution

$\pi$ on $(\mathcal{X}, \mathcal{F})$ which is preserved by $P$ in the sense that $\pi P = \pi$. We define $\Pi := 1_{\mathcal{X}} \otimes \pi$ by

$$\Pi(x, A) \;:=\; (1_{\mathcal{X}} \otimes \pi)(x, A) \;=\; \pi(A), \qquad x \in \mathcal{X},\; A \in \mathcal{F},$$

so that

$$(\mu\,\Pi)(A) \;:=\; \big(\mu(1_{\mathcal{X}} \otimes \pi)\big)(A) \;=\; \mu(\mathcal{X})\,\pi(A)\,.$$

If $\mu$ is a probability measure, then $(\mu\,\Pi)(A) = \pi(A)$, and $\mu(P^n - \Pi) = \mu P^n - \pi$. Also, by stationarity of $\pi$, we have $(P - \Pi)^n = P^n - \Pi$ for each $n \in \mathbb{N}$.

We shall assume that our Markov chain is *$\phi$-irreducible*, i.e. there exists a non-zero $\sigma$-finite measure $\phi$ on $(\mathcal{X}, \mathcal{F})$ such that for all $x \in \mathcal{X}$ and $A \subseteq \mathcal{X}$ with $\phi(A) > 0$, there is $n \in \mathbb{N}$ with $P^n(x, A) > 0$. We shall also assume that it is *aperiodic*, i.e. there do not exist $d \geq 2$ and disjoint $\mathcal{X}_1, \ldots, X_d \subseteq \mathcal{X}$ of positive $\pi$ measure, such that $P(x, \mathcal{X}_{i+1}) = 1$ for all $x \in \mathcal{X}_i$ $(i = 1, \ldots, d-1)$ and $P(x, \mathcal{X}_1) = 1$ for all $x \in \mathcal{X}_d$. It is well-known (e.g. [17, 22]) that these conditions guarantee that $P^n(x, A) \to \pi(A)$ as $n \to \infty$ (see also Q9.1 and Q9.4 below). Geometric ergodicity then corresponds to the property, which may or may not hold, that this convergence occurs exponentially quickly.

We shall also assume that the state space $(\mathcal{X}, \mathcal{F})$ is *countably generated*, i.e. that there exists $A_1, A_2, \ldots \in \mathcal{F}$ such that $\mathcal{F} = \sigma(A_1, A_2, \ldots)$, i.e. $\mathcal{F}$ is the smallest $\sigma$-algebra containing all of the $A_i$. This technical property ensures the existence of small sets [4, 9, 19] and the measurability of certain functions [21, Appendix] (see also Q9.3 below).

A subset $S \in \mathcal{F}$ is called *small* if $\pi(S) > 0$ and there is $m > 0$ and a non-zero measure $\nu$ on $(\mathcal{X}, \mathcal{F})$ such that $P^m(x, A) \;\geq\; \nu(A)$ for all $x \in S$ and $A \in \mathcal{F}$, i.e. if all of the $m$-step transition probabilities from within $S$ all have some "overlap". This property is very useful for coupling constructions and for ensuring convergence to stationarity (see e.g. [17, 22]).

The *total variation distance* between two probability measures $\mu_1$ and $\mu_2$ is defined by:

$$\|\mu_1 - \mu_2\|_{\mathrm{TV}} \;=\; \sup_{A \in \mathcal{F}} |\mu_1(A) - \mu_2(A)| \;\equiv\; \frac{1}{2} \sup_{|f| \leq 1} \Big| \int f d\mu_1 - \int f d\mu_2 \Big|$$

(see e.g. [22, Proposition 3(b)]). Given a positive function $V : \mathcal{X} \to \mathbb{R}$, we define [17, p. 390] the *$V$-norm* $|f|_V = \sup_{x \in \mathcal{X}} \frac{|f(x)|}{V(x)}$. We let $L_V^\infty$ be the vector space of all functions $f : \mathcal{X} \to \mathbb{R}$ such that $|f|_V < \infty$, and let $L_{V,0}^\infty = \{f \in L_V^\infty : \pi(f) = 0\}$. Then, we define the *$V$-norm* of a Markov kernel $P$ as

$$\|P\|_{L_V^\infty} = \sup_{\substack{f \in L_V^\infty \\ |f|_V = 1}} |Pf|_V; \qquad \|P\|_{L_{V,0}^\infty} = \sup_{\substack{f \in L_{V,0}^\infty \\ |f|_V = 1}} |Pf|_V.$$

For a (possibly signed) measure $\mu$, we define $\|\mu\|_{L^p(\pi)}$ for $1 \le p < \infty$ by

$$\|\mu\|_{L^p(\pi)}^p = \begin{cases} \mu^+(\mathcal{X}) + \mu^-(\mathcal{X}), & \text{if } p = 1 \\ \int_{\mathcal{X}} \left|\frac{d\mu}{d\pi}\right|^p d\pi, & \text{if } \mu \ll \pi \\ \infty, & \text{otherwise.} \end{cases}$$

(If $p = 1$ and $\mu \ll \pi$, then the two definitions coincide.) We let $L^p(\pi)$ be the collection of all signed measures $\mu$ on $(\mathcal{X}, \mathcal{F})$ with $\|\mu\|_{L^p(\pi)} < \infty$, and define the $L^p(\pi)$-norm of a transition kernel $P$ acting on the set $L^p(\pi)$ by:

$$\|P\|_{L^p(\pi)} = \sup_{\|\mu\|_{L^p(\pi)}=1} \|\mu P(\cdot)\|_{L^p(\pi)}.$$

(Note in particular that the $L^p(\pi)$ are collections of signed *measures*, while $L_V^\infty$ and $L_{V,0}^\infty$ are collections of *functions*.)

The transition kernel $P$ is *reversible* with respect to $\pi$ if $\pi(dx)\, P(x, dy) = \pi(dy)\, P(y, dx)$ for all $x, y \in \mathcal{X}$. This is equivalent to $P$ being a self-adjoint operator on the Hilbert space $L^2(\pi)$, with inner product given by

$$\langle \mu, \nu \rangle = \int_{\mathcal{X}} \frac{d\mu}{d\pi} \frac{d\nu}{d\pi}\, d\pi.$$

In particular, $\langle \mu, \pi \rangle = \int_{\mathcal{X}} \frac{d\mu}{d\pi}\, 1\, d\pi = \mu(\mathcal{X})$. We also let $\pi^\perp := \{\mu \in L^2(\pi) : \mu(\mathcal{X}) = 0\}$ be the set of signed measures in $L^2(\pi)$ which are "perpendicular" to $\pi$, i.e. for which $\langle \mu, \pi \rangle \equiv \mu(\mathcal{X}) = 0$. Our conditions (*xxvii*) through (*xxxiii*) are only proven to be equivalent for reversible chains (though see Q9.5 below).

Finally, given an operator $P$ on a Banach space (i.e. a complete normed vector space) $\mathcal{V}$, e.g. $\mathcal{V} = L_V^\infty$ or $L^2(\pi)$, the *spectrum* of $P$, denoted by $\mathcal{S}(P)$ or $\mathcal{S}_{\mathcal{V}}(P)$, is the set of all complex numbers $\lambda$ such that $\lambda I - P$ is not invertible (see e.g. [27, p. 253]). And, the *spectral radius* of $P$ is the number $r(P) = r_{\mathcal{V}}(P) = \sup_{\lambda \in \mathcal{S}_{\mathcal{V}}(P)} |\lambda|$.

# 3 Main Result: Statement of Equivalences

We now provide a list of 26 conditions which are always equivalent to geometric ergodicity of Markov chains, and an additional 7 (for 33 total) which are also equivalent for reversible chains. Some of the conditions are very similar to each other, but are included to allow for maximum flexibility when establishing or using geometric ergodicity in both theoretical investigations and applications. For ease of comprehension, similar conditions are grouped together under common subheadings.

**Theorem 1.** *Let $P$ be the transition kernel of a $\phi$-irreducible, aperiodic Markov chain $\Phi = \{X_n\}$ with stationary probability distribution $\pi$ on a countably generated measurable state space $(\mathcal{X}, \mathcal{F})$. Then the following are equivalent (and all correspond to being "geometrically ergodic"):*

Geometric Convergence in TV:

*i)* *$\Phi$ is geometrically ergodic starting from $\pi$-a.e. $x \in \mathcal{X}$ with constant geometric rate. This means there is fixed $\rho < 1$ such that for $\pi$-a.e. $x \in \mathcal{X}$ there is $C_x < \infty$ with*

$$\|P^n(x,\cdot) - \pi(\cdot)\|_{\text{TV}} \;\le\; C_x\,\rho^n \qquad \text{for all } n \in \mathbb{N}.$$

*ii)* *There exists $A \in \mathcal{F}$ with $\pi(A) > 0$ such that $\Phi$ is geometrically ergodic starting from each $x \in A$. This means for each $x \in A$, there are $\rho_x < 1$ and $C_x < \infty$ with*

$$\|P^n(x,\cdot) - \pi(\cdot)\|_{\text{TV}} \;\le\; C_x\,\rho_x^n \qquad \text{for all } n \in \mathbb{N}.$$

*iii)* *There exists $p \in (1,\infty)$ such that $\Phi$ is geometrically ergodic starting from all probability measures in $L^p(\pi)$. This means there is some $p \in (1,\infty)$ such that for each probability measure $\mu \in L^p(\pi)$ there are constants $\rho_\mu < 1$ and $C_\mu < \infty$ with*

$$\|\mu P^n(\cdot) - \pi(\cdot)\|_{\text{TV}} \;\le\; C_\mu\,\rho_\mu^n \qquad \text{for all } n \in \mathbb{N}.$$

*iv)* *For all $p \in (1,\infty)$, $\Phi$ is geometrically ergodic starting from all probability measures in $L^p(\pi)$ with geometric rate depending only on $p$. This means for each $p \in (1,\infty)$, there is $\rho_p < 1$ such that for each probability measure $\mu \in L^p(\pi)$ there is $C_{p,\mu} < \infty$ with*

$$\|\mu P^n(\cdot) - \pi(\cdot)\|_{\text{TV}} \;\le\; C_{p,\mu}\,\rho_p^n \qquad \text{for all } n \in \mathbb{N}.$$

*v)* *There exists a small set $S \in \mathcal{F}$ such that $\Phi$ is geometrically ergodic starting from the stationary distribution restricted to $S$. This means there are constants $\rho_S < 1$ and $C_S < \infty$ with*

$$\|\pi_S P^n(\cdot) - \pi(\cdot)\|_{\text{TV}} \;\le\; C_S\,\rho_S^n \qquad \text{for all } n \in \mathbb{N},$$

*where $\pi_S$ is the probability measure defined by $\pi_S(A) = \pi(S \cap A) \,/\, \pi(S)$ for $A \in \mathcal{F}$.*

Geometric Return Time:

*vi) There exists a small set $S \in \mathcal{F}$ and constant $\kappa > 1$ such that*

$$\sup_{x \in S} \mathbf{E}_x[\kappa^{\tau_S}] \;<\; \infty$$

*where $\tau_S$ is the first return time to $S$, and $\mathbf{E}_x$ is expected value conditional on $X_0 = x$.*

$V$-Function Drift Condition:

*vii) There exists a $\pi$-a.e.-finite measurable function $V : \mathcal{X} \to [1, \infty]$, a small set $S \in \mathcal{F}$, and constants $\lambda < 1$ and $b < \infty$ with*

$$PV(x) \;\leq\; \lambda V(x) + b\,\mathbb{1}_S(x) \qquad \text{for all } x \in \mathcal{X}.$$

*viii) For all $j \in \mathbb{N}$, there exists a $\pi$-a.e.-finite measurable function $V : \mathcal{X} \to [1, \infty]$, a small set $S \in \mathcal{F}$, and constants $\lambda < 1$ and $b < \infty$ with $\pi(V^j) < \infty$ and*

$$PV(x) \;\leq\; \lambda V(x) + b\,\mathbb{1}_S(x) \qquad \text{for all } x \in \mathcal{X}.$$

$V$-Uniform Convergence:

*ix) There exists a $\pi$-a.e.-finite measurable function $V : \mathcal{X} \to [1, \infty]$ such that $\Phi$ is $V$-uniformly ergodic. This means there is $\rho < 1$ and $C < \infty$ such that*

$$\sup_{|f| \leq V} \left| P^n f(x) - \pi(f) \right| \;\leq\; C\, V(x)\, \rho^n \qquad \text{for all } x \in \mathcal{X} \text{ and } n \in \mathbb{N}.$$

*x) For all $j \in \mathbb{N}$, there exists a $\pi$-a.e.-finite measurable function $V : \mathcal{X} \to [1, \infty]$ with $\pi(V^j) < \infty$, such that $\Phi$ is $V$-uniformly ergodic. This means there is $\rho < 1$ and $C < \infty$ such that*

$$\sup_{|f| \leq V} \left| P^n f(x) - \pi(f) \right| \;\leq\; C\, V(x)\, \rho^n \qquad \text{for all } x \in \mathcal{X} \text{ and } n \in \mathbb{N}.$$

*xi) There exists a $\pi$-a.e.-finite measurable function $V : \mathcal{X} \to [1, \infty]$, and constants $\rho < 1$ and $C < \infty$, such that for each probability measure $\mu$ on $\mathcal{X}$ with $\mu(V) < \infty$,*

$$\sup_{|f| \leq V} \left| \mu P^n(f) - \pi(f) \right| \;\leq\; C\, \mu(V)\, \rho^n \qquad \text{for all } n \in \mathbb{N}.$$

*xii) For all $j \in \mathbb{N}$, there exists a $\pi$-a.e.-finite measurable function $V : \mathcal{X} \to [1,\infty]$ with $\pi(V^j) < \infty$, and constants $\rho < 1$ and $C < \infty$, such that for each probability measure $\mu$ on $\mathcal{X}$ with $\mu(V) < \infty$,*

$$\sup_{|f|\le V} \left|\mu P^n(f) - \pi(f)\right| \;\le\; C\,\mu(V)\,\rho^n \qquad \text{for all } n \in \mathbb{N}.$$

<u>Spectral Gap:</u>

*xiii) There exists $j \in \mathbb{N}$ and a $\pi$-a.e.-finite measurable function $V\colon \mathcal{X} \to [1,\infty]$ with $\pi(V^j) < \infty$, such that $P$ has a spectral gap as an operator on $L_V^\infty$, meaning 1 is an eigenvalue of $P$ with multiplicity 1, and there is $\rho < 1$ such that*

$$\mathcal{S}_{L_V^\infty}(P) \setminus \{1\} \;\subseteq\; \{z \in \mathbb{C} : |z| \le \rho\}.$$

*xiv) For all $j \in \mathbb{N}$, there exists a $\pi$-a.e.-finite measurable function $V\colon \mathcal{X} \to [1,\infty]$ with $\pi(V^j) < \infty$, such that $P$ has a spectral gap as an operator on $L_V^\infty$, meaning 1 is an eigenvalue of $P$ with multiplicity 1, and there is $\rho < 1$ such that*

$$\mathcal{S}_{L_V^\infty}(P) \setminus \{1\} \;\subseteq\; \{z \in \mathbb{C} : |z| \le \rho\}.$$

<u>Spectral Radius:</u>

*xv) There exists $j \in \mathbb{N}$ and a $\pi$-a.e.-finite measurable function $V : \mathcal{X} \to [1,\infty]$ with $\pi(V^j) < \infty$, such that $P - \Pi$ has spectral radius less than one as an operator on $L_V^\infty$, i.e.*

$$r_{L_V^\infty}(P - \Pi) < 1.$$

*xvi) For all $j \in \mathbb{N}$, there exists a $\pi$-a.e.-finite measurable function $V : \mathcal{X} \to [1,\infty]$ with $\pi(V^j) < \infty$, such that $P - \Pi$ has spectral radius less than one as an operator on $L_V^\infty$, i.e.*

$$r_{L_V^\infty}(P - \Pi) < 1.$$

*xvii) There exists $j \in \mathbb{N}$ and a $\pi$-a.e.-finite measurable function $V : \mathcal{X} \to [1,\infty]$ with $\pi(V^j) < \infty$, such that $P$ has spectral radius less than one as an operator on $L_{V,0}^\infty$, i.e.*

$$r_{L_{V,0}^\infty}(P) < 1.$$

*xviii) For all $j \in \mathbb{N}$, there exists a $\pi$-a.e.-finite measurable function $V : \mathcal{X} \to [1, \infty]$ with $\pi(V^j) < \infty$, such that $P$ has spectral radius less than one as an operator on $L^\infty_{V,0}$, i.e.*

$$r_{L^\infty_{V,0}}(P) < 1.$$

<u>$L^\infty_V$ Operator Norm:</u>

*xix) There exists $j, m \in \mathbb{N}$ and a $\pi$-a.e.-finite measurable function $V : \mathcal{X} \to [1, \infty]$ with $\pi(V^j) < \infty$, such that*

$$\|P^m - \Pi\|_{L^\infty_V} < 1.$$

*xx) For all $j \in \mathbb{N}$, there exists $m \in \mathbb{N}$ and a $\pi$-a.e.-finite measurable function $V : \mathcal{X} \to [1, \infty]$ such that $\pi(V^j) < \infty$ and*

$$\|P^m - \Pi\|_{L^\infty_V} < 1.$$

*xxi) There exists $j, m \in \mathbb{N}$ and a $\pi$-a.e.-finite measurable function $V : \mathcal{X} \to [1, \infty]$ with $\pi(V^j) < \infty$, such that*

$$\|P^m\|_{L^\infty_{V,0}} < 1.$$

*xxii) For all $j \in \mathbb{N}$, there exists $m \in \mathbb{N}$ and a $\pi$-a.e.-finite measurable function $V : \mathcal{X} \to [1, \infty]$ with $\pi(V^j) < \infty$, such that*

$$\|P^m\|_{L^\infty_{V,0}} < 1.$$

*xxiii) There exists $j \in \mathbb{N}$ and a $\pi$-a.e.-finite measurable function $V : \mathcal{X} \to [1, \infty]$ with $\pi(V^j) < \infty$, and constants $\rho < 1$ and $C < \infty$, such that*

$$\|P^n - \Pi\|_{L^\infty_V} \ \le \ C\,\rho^n \qquad \text{for all } n \in \mathbb{N}.$$

*xxiv) For all $j \in \mathbb{N}$, there exists a $\pi$-a.e.-finite measurable function $V : \mathcal{X} \to [1, \infty]$ with $\pi(V) < \infty$, and constants $\rho < 1$ and $C < \infty$, such that*

$$\|P^n - \Pi\|_{L^\infty_V} \ \le \ C\,\rho^n \qquad \text{for all } n \in \mathbb{N}.$$

*xxv) There exists $j \in \mathbb{N}$ and a $\pi$-a.e.-finite measurable function $V : \mathcal{X} \to [1, \infty]$ with $\pi(V^j) < \infty$, and constants $\rho < 1$ and $C < \infty$, such that*

$$\|P^n\|_{L^\infty_{V,0}} \ \le \ C\,\rho^n \qquad \text{for all } n \in \mathbb{N}.$$

*xxvi)* *For all $j \in \mathbb{N}$, there exists a $\pi$-a.e.-finite measurable function $V : \mathcal{X} \to [1, \infty]$ with $\pi(V^j) < \infty$, and constants $\rho < 1$ and $C < \infty$, such that*

$$\|P^n\|_{L^\infty_{V,0}} \;\leq\; C\,\rho^n \qquad \text{for all } n \in \mathbb{N}.$$

Conditions Assuming Reversibility:

*Furthermore, if $\Phi$ is reversible, then the following are also equivalent to the above:*

*xxvii)* *$\Phi$ is $L^2(\pi)$-geometrically ergodic starting from any probability measure $\mu$ in $L^2(\pi)$ with uniform convergence rate. This means there is $\rho < 1$ such that for each probability measure $\mu \in L^2(\pi)$, there is a constant $C_\mu < \infty$ such that*

$$\|\mu P^n(\cdot) - \pi(\cdot)\|_{L^2(\pi)} \;\leq\; C_\mu\,\rho^n \qquad \text{for all } n \in \mathbb{N}.$$

*xxviii)* *There exists $\rho < 1$ such that for each probability measure $\mu \in L^2(\pi)$,*

$$\|\mu P^n(\cdot) - \pi(\cdot)\|_{L^2(\pi)} \;\leq\; \|\mu - \pi\|_{L^2(\pi)}\,\rho^n \qquad \text{for all } n \in \mathbb{N}.$$

*xxix)* *$P$ has a spectral gap as an operator on $L^2(\pi)$, meaning that 1 is an eigenvalue of $P$ with multiplicity 1, and there is $\rho < 1$ with*

$$\mathcal{S}_{L^2(\pi)}(P) \setminus \{1\} \;\subseteq\; \{z \in \mathbb{C} : |z| \leq \rho\}.$$

*xxx)* *$P - \Pi$ has spectral radius less than one as an operator on $L^2(\pi)$, i.e.*

$$r_{L^2(\pi)}(P - \Pi) \;<\; 1.$$

*xxxi)* *$P - \Pi$ has operator norm less than one as an operator on $L^2(\pi)$, i.e.*

$$\|P - \Pi\|_{L^2(\pi)} \;<\; 1.$$

*xxxii)* *$P$ has operator norm less than one as an operator on $\pi^\perp$, i.e.*

$$\|P\|_{\pi^\perp} \;<\; 1.$$

*xxxiii)* *$P|_{\pi^\perp}$ has spectral radius less than one as an operator on $\pi^\perp$, i.e.*

$$r_{\pi^\perp}(P) \;<\; 1.$$

**Remark.** As mentioned, some of these equivalences are already known. The fact that $(v)$ implies $(i)$ was shown in [29] on countable state spaces, and then by [18, Theorem 1] on general state spaces. The equivalence of $(v)$, $(vi)$, and $(vii)$, together with the fact that they imply $(i)$, was presented in [17, Theorem 15.0.1]. The equivalence of $(vii)$, $(ix)$, $(xix)$, and $(xxv)$ was presented in [17, Theorem 16.0.1]. The equivalence of the group $(i)$, $(v)$, $(ix)$, $(x)$, $(xx)$, and $(xxii)$ was presented in [21, Proposition 1], and the equivalence (assuming reversibility) of the group $(xxvii)$, $(xxviii)$, and $(xxxii)$ was presented in [21, Theorem 2], together with the fact that the first group implies the second. The reverse implication, that the second group implies the first, was then shown in [25]. Discussions related to the spectral gap conditions $(xiii)$ and $(xiv)$ and $(xxix)$ appear in [13]. The equivalence of $(xiii)$ and $(vii)$ is shown in [14, Proposition 1.1], and the equivalence of $(xxx)$ and $(i)$ for reversible chains is shown in [14, Proposition 1.2]. Our Theorem 1 is an attempt to combine and bring together all of these various results, plus others too. (Since initiating this work, we also learned of the recent review [1], which presents certain equivalences for reversible chains in terms of mixing conditions and maximal correlations, which complement some of our conditions $(xxvii)$ through $(xxxiii)$.)

Most of the remainder of this paper is devoted to proving Theorem 1. The proof is divided up into different sections below, in terms of which types of conditions are being considered: Section 4 provides some preliminary lemmas, Section 5 relates to various "Geometric" conditions, Section 6 relates to various conditions involving $V$ functions and $L^\infty_V$ bounds, Section 7 relates to various spectral conditions, and Section 8 relates to various conditions for reversible chains. To help the reader (and ourselves) keep track, Figure 1 provides a diagram showing which of our results prove implications between which of the equivalent conditions. Our proofs are mostly self-contained, though we use known results from the literature (especially [17]) where appropriate. Section 9 then presents some future directions and open problems.

# 4 Preliminary Lemmas

We begin with some preliminary lemmas, which are used freely in the sequel, and can be referred to as needed.

**Lemma 4.1.** *Let $P$ be the transition kernel of a $\phi$-irreducible, aperiodic Markov chain with stationary distribution $\pi$ on a countably generated state space $\mathcal{X}$. Then for any measurable subset $A \subseteq \mathcal{X}$ such that $\pi(A) > 0$, there exists a small set $S$, such that $S \subseteq A$.*

*Proof.* This result goes back to [4, 9, 19], and uses that $\mathcal{F}$ is countably generated; see e.g. Theorems 5.2.1 and 5.2.2 in [17]. □

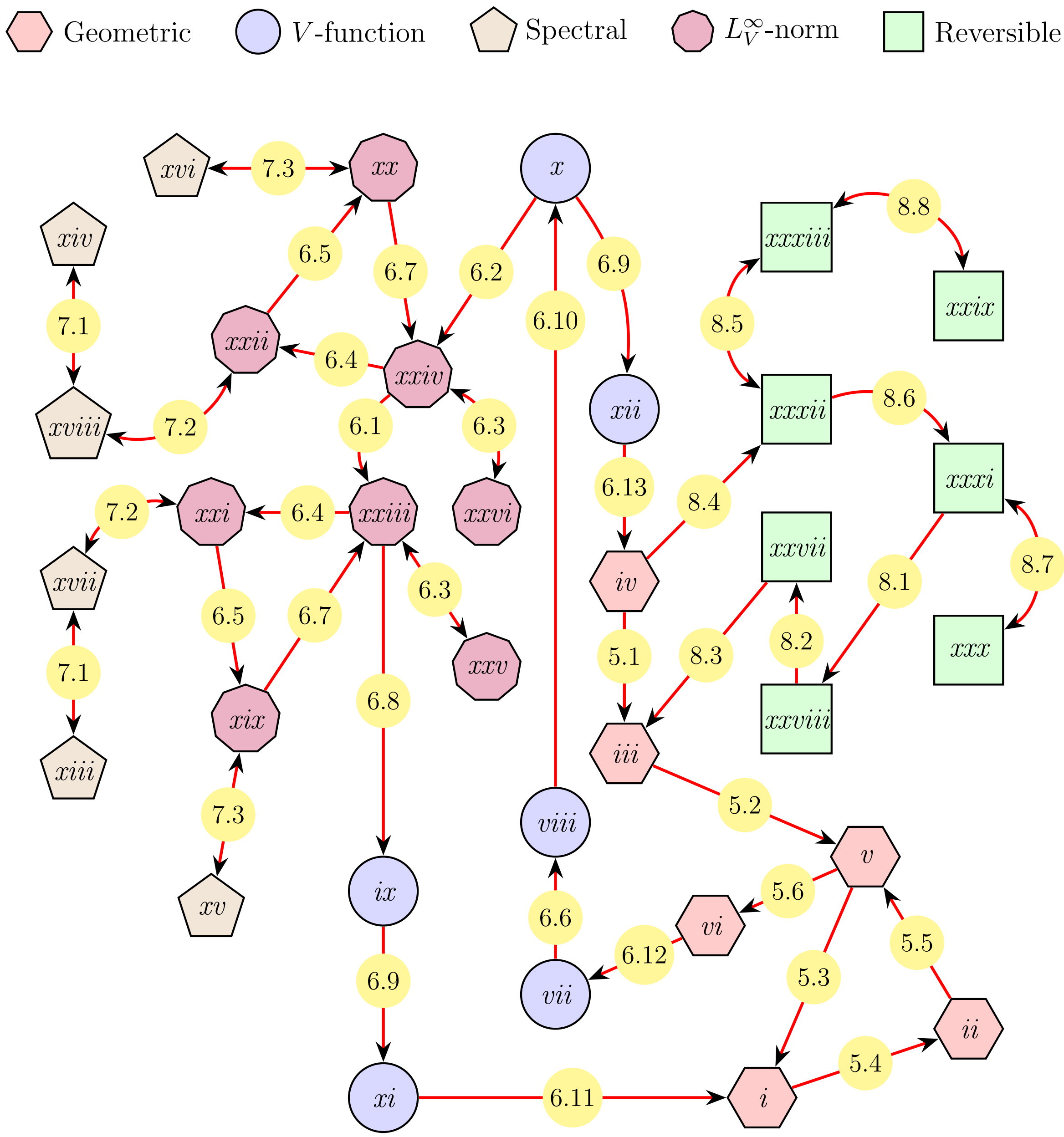


Figure 1: Diagram illustrating which of this paper's results (yellow edge labels) provide proofs of implications between which of the different equivalent conditions (nodes). (All arrows touching a green rectangle assume that the chain is reversible.)

**Lemma 4.2.** *Let $P$ be the transition kernel of a $\phi$-irreducible, aperiodic Markov chain with stationary distribution $\pi$ on a countably generated state space $\mathcal{X}$. Then, the function $F_n : \mathcal{X} \to [0, \infty)$ defined by $F_n(x) = \|P^n(x, \cdot) - \pi(\cdot)\|_{\mathrm{TV}}$ is measurable.*

*Proof.* This follows from [21, Appendix], which proves that for any bounded signed measure $\nu(\cdot, A)$ on a countably generated space such that the function $x \mapsto \nu(x, A)$ is measurable for each fixed $A \in \mathcal{F}$, the function $x \mapsto \sup_{A \in \mathcal{F}} \nu(x, A)$ is also measurable. □

**Lemma 4.3.** *For probability measures $\mu_1$ and $\mu_2$, $\|\mu_1 - \mu_2\|_{\mathrm{TV}} = \frac{1}{2}\|\mu_1 - \mu_2\|_{L^1(\pi)}$.*

*Proof.* Recall that $\|\mu_1 - \mu_2\|_{\mathrm{TV}} = \sup_{A \in \mathcal{F}} |\mu_1(A) - \mu_2(A)|$. Let $\nu = \mu_1 + \mu_2$ so that $\mu_i \ll \nu$, and let $f_i = \frac{d\mu_i}{d\nu}$. Then $\mu_1(A) - \mu_2(A) = \int_A [f_1(x) - f_2(x)]\, \nu(dx)$. This is maximised when $A = A_+ := \{x : f_1(x) > f_2(x)\}$, and its negative takes the same maximum when $A = A_+^C$. Hence,

$$\|\mu_1 - \mu_2\|_{\mathrm{TV}} \;=\; \mu_1(A_+) - \mu_2(A_+) \;=\; \int_{A_+} [f_1(x) - f_2(x)]\, \nu(dx).$$

But then

$$\begin{aligned}
\|\mu_1 - \mu_2\|_{L^1(\pi)} &= (\mu_1 - \mu_2)^+(\mathcal{X}) + (\mu_1 - \mu_2)^-(\mathcal{X}) \\
&= \int_{A_+} [f_1(x) - f_2(x)]\, \nu(dx) + \int_{A_+^C} [f_2(x) - f_1(x)]\, \nu(dx) \\
&= 2 \int_{A_+} [f_1(x) - f_2(x)]\, \nu(dx) \\
&= 2\, \|\mu_1 - \mu_2\|_{\mathrm{TV}}.
\end{aligned}$$

□

**Lemma 4.4.** *For any signed measure $\mu \ll \pi$, we have $\|\mu\|_{L^1(\pi)} \le \|\mu\|_{L^2(\pi)}$ (though one or both of those quantities might be infinite).*

*Proof.* Recall the definition $\langle \mu, \nu \rangle = \int_{\mathcal{X}} \frac{d\mu}{d\pi} \frac{d\nu}{d\pi}\, d\pi$. Hence, if $|\mu|$ is the measure with $\frac{d|\mu|}{d\pi} = \left|\frac{d\mu}{d\pi}\right|$, then $\langle |\mu|, \pi \rangle = \int_{\mathcal{X}} \left|\frac{d\mu}{d\pi}\right| (1)\, d\pi = \mu^+(\mathcal{X}) + \mu^-(\mathcal{X}) = \|\mu\|_{L^1(\pi)}$. Also

$$\|\, |\mu|\, \|_{L^2(\pi)} \;=\; \sqrt{\int_{\mathcal{X}} \left|\frac{d\mu}{d\pi}\right|^2 d\pi} \;=\; \sqrt{\int_{\mathcal{X}} \left(\frac{d\mu}{d\pi}\right)^2 d\pi} \;=\; \|\mu\|_{L^2(\pi)},$$

and

$$\|\pi\|_{L^2(\pi)} \;=\; \sqrt{\int_{\mathcal{X}} \left(\frac{d\pi}{d\pi}\right)^2 d\pi} \;=\; \sqrt{\int_{\mathcal{X}} (1)^2\, d\pi} \;=\; 1.$$

So, by the Cauchy-Schwarz inequality,

$$\|\mu\|_{L^1(\pi)} \;=\; \langle |\mu|, \pi \rangle \;\le\; \|\, |\mu|\, \|_{L^2(\pi)}\, \|\pi\|_{L^2(\pi)} \;=\; \|\mu\|_{L^2(\pi)}\, (1) \;=\; \|\mu\|_{L^2(\pi)}.$$

□

**Lemma 4.5.** *For all $1 \le p < s < \infty$, we have $L^s(\pi) \subseteq L^p(\pi)$.*

*Proof.* Let $1 \le p < s < \infty$, and let $\mu \in L^s(\pi)$ so $\|\mu\|_{L^s(\pi)} < \infty$. Then,

$$\|\mu\|_{L^p(\pi)}^p \;=\; \int_{\mathcal{X}} \left|\frac{d\mu}{d\pi}\right|^p d\pi \;\le\; \int_{\mathcal{X}} \left(1 + \left|\frac{d\mu}{d\pi}\right|^s\right) d\pi \;=\; 1 + \|\mu\|_{L^s(\pi)}^s \;<\; \infty,$$

so $\mu \in L^p(\pi)$. □

We next present some lemmas which mention spectra of operators.

**Lemma 4.6.** *Suppose an operator $P$ on a Banach space $\mathcal{V}$ can be decomposed as a* direct sum *$P = P_1 \oplus P_2$, where $\mathcal{V} = \mathcal{V}_1 \times \mathcal{V}_2$ and each $P_i$ is an operator on $\mathcal{V}_i$, meaning that $P(h_1, h_2) = (P_1 h_1, P_2 h_2)$ for all $h_1 \in \mathcal{V}_1$ and $h_2 \in \mathcal{V}_2$. Then $\mathcal{S}_{\mathcal{V}}(P) = \mathcal{S}_{\mathcal{V}_1}(P_1) \cup \mathcal{S}_{\mathcal{V}_2}(P_2)$, i.e. the spectrum of $P$ is the union of the spectra of the sub-operators $P_1$ and $P_2$.*

*Proof.* Since $P = P_1 \oplus P_2$, therefore $P$ has the block decomposition

$$P \;=\; \begin{pmatrix} P_1 & 0 \\ 0 & P_2 \end{pmatrix}$$

with respect to $\mathcal{V} = \mathcal{V}_1 \times \mathcal{V}_2$. If $\lambda \notin \mathcal{S}_{\mathcal{V}_1}(P_1) \cup \mathcal{S}_{\mathcal{V}_2}(P_2)$, then there are inverse operators $A_i$ on $\mathcal{V}_i$ such that $(\lambda I_i - P_i)A_i = A_i(\lambda I_i - P_i) = I_i$ for $i = 1, 2$, whence $(\lambda I - P)(A_1, A_2) = (A_1, A_2)(\lambda I - P) = I_1 \oplus I_2 = I$, so $\lambda \notin \mathcal{S}_{\mathcal{V}}(P)$. Conversely, if $\lambda \notin \mathcal{S}_{\mathcal{V}}(P)$, then $\lambda I - P$ has some inverse operator, so in block form we have

$$\begin{pmatrix} \lambda I_1 - P_1 & 0 \\ 0 & \lambda I_2 - P_2 \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \;=\; \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} \lambda I_1 - P_1 & 0 \\ 0 & \lambda I_2 - P_2 \end{pmatrix} \;=\; I \;=\; \begin{pmatrix} I_1 & 0 \\ 0 & I_2 \end{pmatrix}.$$

It follows that $(\lambda I_1 - P_1)A = A(\lambda I_1 - P_1) = I_1$ and $(\lambda I_2 - P_2)D = D(\lambda I_2 - P_2) = I_2$, so that $\lambda \notin \mathcal{S}_{\mathcal{V}_1}(P_1) \cup \mathcal{S}_{\mathcal{V}_2}(P_2)$. □

**Lemma 4.7.** *Let $P$ be the transition kernel of a $\phi$-irreducible Markov chain with stationary distribution $\pi(\cdot)$, and let $V : \mathcal{X} \to [1, \infty]$ be a $\pi$-a.e.-finite measurable function. Then, the following hold:*

*1) $|f|_V \le 1$ if and only if $|f(x)| \le V(x)$ for all $x \in \mathcal{X}$.*

*2) If $P$ is a bounded operator on $L_V^\infty$, then $\mathcal{S}_{L_V^\infty}(P) \setminus \{1\} \subseteq \mathcal{S}_{L_{V,0}^\infty}(P)$.*

*3) If there is $j \in \mathbb{N}$ with $\pi(V^j) < \infty$, then $\pi(V) < \infty$.*

*4) If there are $\lambda < 1$ and $b < \infty$ and a small set $S \in \mathcal{F}$ with $PV(x) \leq \lambda V(x) + b\,\mathbb{1}_S(x)$ for all $x \in \mathcal{X}$, then $\pi(V) < \infty$.*

*Proof.* 1) If $|f|_V \leq 1$, then for each $x \in \mathcal{X}$,

$$\frac{|f(x)|}{V(x)} \leq |f|_V \leq 1,$$

from which we conclude that $|f| \leq V$. Conversely, if $|f| \leq V$, then for each $x \in \mathcal{X}$, $\dfrac{|f(x)|}{V(x)} \leq 1$, and thus $|f|_V = \sup_{x \in \mathcal{X}} \dfrac{|f(x)|}{V(x)} \leq 1$.

2) Any $f \in L_V^\infty$ can be written as $f = f_0 + c$ where $f_0 \in L_{V,0}^\infty$ and $c = \pi(f)$. Then $Pf = Pf_0 + c$. It follows that $P$ has the direct sum representation $P = P_0 \oplus I_\mathbb{R}$, where $I_\mathbb{R}$ is the identity operator on $\mathbb{R}$. Hence, by Lemma 4.6, $\mathcal{S}_{L_V^\infty}(P) = \mathcal{S}_{L_{V,0}^\infty}(P) \cup \mathcal{S}_\mathbb{R}(I_\mathbb{R}) = \mathcal{S}_{L_{V,0}^\infty}(P) \cup \{1\}$. So, $\mathcal{S}_{L_V^\infty}(P) \setminus \{1\} \subseteq \mathcal{S}_{L_{V,0}^\infty}(P)$, as claimed.

3) This follows since we always have $V(x) \leq V^j(x) + 1$. [In fact, since $V \geq 1$, the "+1" is not actually necessary.]

4) The implication "(*iii*) $\Rightarrow$ (*i*)" of [17, Theorem 14.0.1] with the choice $f(x) = (1-\lambda)\,V(x)$ shows that $\pi(f) < \infty$, i.e. $(1-\lambda)\,\pi(V) < \infty$, hence $\pi(V) < \infty$. [In fact, once we know that $\pi(V) < \infty$, then since $PV \leq \lambda V + b$, it follows that $\pi(PV) \leq \pi(\lambda V + b)$, i.e. $\pi(V) \leq \lambda\,\pi(V) + b$, and hence $\pi(V) \leq b\big/(1-\lambda)$.] □

**Lemma 4.8.** *Let $P$ be the transition kernel of a $\phi$-irreducible Markov chain on a state space $\mathcal{X}$, and let $\mathcal{V}$ be either a vector space of signed measures on $\mathcal{X}$ which includes $\pi$, or a vector space of functions $\mathcal{X} \to \mathbb{R}$ which includes the constant functions. Then $P$ as an operator on $\mathcal{V}$ has eigenvalue 1 with multiplicity 1.*

*Proof.* It follows that 1 is an eigenvalue since $\pi P = \pi$ and also $Pc = c$ for constant functions $c$. To show multiplicity 1, suppose first that there is a signed measure $\pi'$, linearly independent from $\pi$, such that also $\pi' P = \pi'$. Then there is some linear combination $\mu = a\pi + b\pi'$ such that both $\mu^+$ and $\mu^-$ are non-zero measures.

□

Finally, we present some lemmas which are specific to reversible chains.

**Lemma 4.9.** *If $P$ is a transition kernel of a reversible Markov chain, then for all $n \geq 1$,*

$$\pi(dx)\,P^n(x,dy) \;=\; \pi(dy)\,P^n(y,dx). \tag{1}$$

*Proof.* For $n = 1$, (1) holds by reversibility. Furthermore, if (1) holds for $n$, then for $n+1$,

$$
\begin{aligned}
\pi(dx)P^{n+1}(x,dy) = \pi(dx)\int_{z\in\mathcal{X}} P(z,dy)P^n(x,dz) &= \int_{z\in\mathcal{X}} P(z,dy)\pi(dx)P^n(x,dz) \\
\text{(Induction hypothesis)} \quad &= \int_{z\in\mathcal{X}} P(z,dy)\pi(dz)P^n(z,dx) \\
\text{(Reversibility)} \quad &= \int_{z\in\mathcal{X}} P^n(z,dx)\pi(dy)P(y,dz) \\
&= \pi(dy)\, P^{n+1}(y,dx),
\end{aligned}
$$

so (1) holds for $n+1$. Hence, by induction, (1) holds for all $n \geq 1$. $\square$

**Lemma 4.10.** *If $P$ is a transition kernel of a reversible Markov chain with stationary distribution $\pi$, then for each $\mu \in L^2(\pi)$ and each $n \in \mathbb{N}$,*

$$\frac{d(\mu P^n)}{d\pi} = P^n \frac{d\mu}{d\pi}.$$

*Proof.* For all $A \in \mathcal{F}$ and $n \in \mathbb{N}$,

$$
\begin{aligned}
\mu P^n(A) \;=\; \int_{x\in\mathcal{X}} P^n(x,A)\,\mu(dx) \;&=\; \int_{x\in\mathcal{X}}\int_{y\in A} \frac{d\mu}{d\pi}(x)\,P^n(x,dy)\,\pi(dx) \\
\text{(Fubini's Theorem)} \quad &=\; \int_{y\in A}\int_{x\in\mathcal{X}} \frac{d\mu}{d\pi}(x)\,\pi(dx)\,P^n(x,dy) \\
\text{(Lemma 4.9)} \quad &=\; \int_{y\in A}\int_{x\in\mathcal{X}} \frac{d\mu}{d\pi}(x)\,\pi(dy)\,P^n(y,dx) \\
&=\; \int_{y\in A}\left(\int_{x\in\mathcal{X}} \frac{d\mu}{d\pi}(x)\,P^n(y,dx)\right)\pi(dy) \\
&=\; \int_{y\in A}\left(P^n\frac{d\mu}{d\pi}(y)\right)\pi(dy).
\end{aligned}
$$

Since this is true for all $A \in \mathcal{F}$, the result follows. $\square$

**Lemma 4.11.** *Let $P$ be the transition kernel of a reversible Markov chain with stationary distribution $\pi$, such that $P$ is a bounded operator on $L^2(\pi)$. Then, the following holds:*

1) *The operator $P - \Pi$ is self-adjoint.*

2) *For each $\mu \in L^2(\pi)$, the signed measure $\mu - \mu(\mathcal{X})\pi$ is orthogonal to $\pi$.*

3) *For each $\mu \in L^2(\pi)$, $\|\mu - \mu(\mathcal{X})\pi\|^2_{L^2(\pi)} = \|\mu\|^2_{L^2(\pi)} - \mu(\mathcal{X})^2$.*

4) *$\mathcal{S}_{L^2(\pi)}(P) \setminus \{1\} \subseteq \mathcal{S}_{\pi^\perp}(P)$.*

*Proof.* 1) For $\mu, \nu \in L^2(\pi)$, we have $\langle \mu(P - \Pi), \nu \rangle = \langle \mu P, \nu \rangle - \langle \mu \Pi, \nu \rangle$. Now, since $P$ is reversible, it is self-adjoint on $L^2(\pi)$, so $\langle \mu P, \nu \rangle = \langle \nu P, \mu \rangle$. Also, we compute that $\langle \mu \Pi, \nu \rangle = \langle \mu(\mathcal{X})\pi, \nu \rangle = \mu(\mathcal{X})\, \nu(\mathcal{X}) = \langle \nu \Pi, \mu \rangle$. Hence, $\langle \mu(P - \Pi), \nu \rangle = \langle \nu(P - \Pi), \mu \rangle$, so $P - \Pi$ is self-adjoint.

2) Let $\mu \in L^2(\pi)$, then,

$$\langle \mu - \mu(\mathcal{X})\pi, \pi \rangle = \langle \mu, \pi \rangle - \mu(\mathcal{X})\langle \pi, \pi \rangle = \langle \mu, \pi \rangle - \langle \mu, \pi \rangle \|\pi\|_{L^2(\pi)} = \langle \mu, \pi \rangle - \langle \mu, \pi \rangle = 0.$$

3) Let $\mu \in L^2(\pi)$. Then,

$$\begin{aligned}
\|\mu - \mu(\mathcal{X})\pi\|^2_{L^2(\pi)} &= \int_{\mathcal{X}} \left| \frac{d\mu}{d\pi}(y) - \mu(\mathcal{X})\,(1) \right|^2 \pi(dy) \\
&= \int_{\mathcal{X}} \left[ \left( \frac{d\mu}{d\pi}(y) \right)^2 - 2\,\mu(\mathcal{X}) \frac{d\mu}{d\pi}(y) + \mu(\mathcal{X})^2 \right] \pi(dy) \\
&= \|\mu\|^2_{L^2(\pi)} - 2\mu(\mathcal{X})^2 + \mu(\mathcal{X})^2 \\
&= \|\mu\|^2_{L^2(\pi)} - \mu(\mathcal{X})^2.
\end{aligned}$$

4) Any signed measure $\mu \in L^2(\pi)$ can be decomposed as $\mu = \mu_0 + c\,\pi$, where $c = \langle \mu, \pi \rangle = \mu(\mathcal{X})$, and $\langle \mu_0, \pi \rangle = \mu_0(\mathcal{X}) = 0$, so $\mu_0 \in \pi^\perp$. Then $\mu P = \mu_0 P + c\,\pi$. It follows that $P$ has the direct sum representation $P = P|_{\pi^\perp} \oplus I_{\mathbb{R}}$ with respect to $L^2(\pi) = \pi^\perp \times \mathbb{R}$. Hence, by Lemma 4.6, $\mathcal{S}_{L^2(\pi)}(P) = \mathcal{S}_{\pi^\perp}(P) \cup \{1\}$, so $\mathcal{S}_{L^2(\pi)}(P) \setminus \{1\} \subseteq \mathcal{S}_{\pi^\perp}(P)$, as claimed.

□

**Lemma 4.12.** *Let $P$ be a transition kernel from a reversible Markov chain with stationary distribution $\pi$. Then,*

$$\|P - \Pi\|_{L^2(\pi)} \;=\; \|P\|_{\pi^\perp}.$$

*Proof.* Any $\mu \in L^2(\pi)$ can be written as $\mu = \mu_0 + c\,\pi$, where $c = \mu(\mathcal{X})$ and $\mu_0 \in \pi^\perp$ so $\mu_0(\mathcal{X}) = 0$. Then $\mu_0 \Pi = \mu_0(\mathcal{X})\,\pi = 0$, so

$$\mu(P - \Pi) \;=\; (\mu_0 + c\,\pi)(P - \Pi) \;=\; \mu_0 P + c\,\pi - 0 - c\,\pi \;=\; \mu_0 P.$$

Also $\|\mu\|_{L^2(\pi)} = \|\mu_0\|_{L^2(\pi)} + c^2 \geq \|\mu_0\|_{L^2(\pi)}$. Hence,

$$\|P - \Pi\|_{L^2(\pi)} \;=\; \sup_{0<\|\mu\|_{L^2(\pi)}<\infty} \frac{||\mu(P - \Pi)||_{L^2(\pi)}}{\|\mu\|_{L^2(\pi)}} \;=\; \sup_{0<\|\mu\|_{L^2(\pi)}<\infty} \frac{||\mu_0 P||_{L^2(\pi)}}{\|\mu_0\|_{L^2(\pi)} + c^2}.$$

This supremum is achieved when $c = 0$, i.e. when $\mu = \mu_0 \in \pi^\perp$, so that

$$\|P - \Pi\|_{L^2(\pi)} \;=\; \sup_{\substack{0<\|\mu_0\|_{L^2(\pi)}<\infty \\ \mu_0 \in \pi^\perp}} \frac{||\mu_0 P||_{L^2(\pi)}}{\|\mu_0\|_{L^2(\pi)}} \;=\; \|P\|_{\pi^\perp}.$$

□

## 5 Proofs for Geometric Conditions

We now begin proving the actual equivalences of the various conditions in Theorem 1, as per the plan illustrated in Figure 1. We begin with some results related to some of the "geometric" conditions.

**Proposition 5.1.** *(iv) $\Rightarrow$ (iii).*

*Proof.* Immediate upon e.g. choosing $p = 2$ and setting $C_\mu = C_{2,\mu}$ and $\rho_\mu = \rho_2$ for each probability measure $\mu \in L^2(\pi)$. □

**Proposition 5.2.** *(iii) $\Rightarrow$ (v).*

*Proof.* By Lemma 4.1, there exists a small set $S \subset \mathcal{X}$. Since by assumption $P$ is geometrically ergodic starting from all probability measures in $L^p(\pi)$ it suffices to show that $\pi_S \in L^p(\pi)$. Now for any measurable $A \subset \mathcal{X}$ we have,

$$\pi_S(A) = \frac{\pi(S \cap A)}{\pi(S)} = \frac{1}{\pi(S)} \int_A 1_S d\pi$$

which implies $\frac{d\pi_S}{d\pi} = 1_S/\pi(S)$. Thus

$$\int_{\mathcal{X}} \left| \frac{d\pi_S}{d\pi} \right|^p d\pi = \int_{\mathcal{X}} \frac{1}{\pi(S)^p} 1_S d\pi = \frac{1}{\pi(S)^{p-1}} < \infty$$

□

**Proposition 5.3.** *(v) $\Rightarrow$ (i).*

*Proof.* This is the result of [18, Theorem 1], which generalizes the countable state space result of [29]. □

**Proposition 5.4.** *(i) $\Rightarrow$ (ii).*

*Proof.* Immediate upon choosing $A = \mathcal{X}$, and $\rho_x = \rho$ for all $x \in \mathcal{X}$. □

**Proposition 5.5.** *(ii) $\Rightarrow$ (v).*

*Proof.* Let $A \in \mathcal{F}$ with $\pi(A) > 0$ and $\|P^n(x,\cdot) - \pi(\cdot)\|_{\text{TV}} \leq C_x \rho_x^n$ for all $x \in A$ and $n \in \mathbb{N}$. For each $n \in \mathbb{N}$ define $F_n : A \to [0, \infty)$ by $F_n(x) = \|P^n(x,\cdot) - \pi(\cdot)\|_{\text{TV}}$. Since each $F_n$ is measurable (Lemma 4.2) so are the functions $\rho, C : A \to [0, \infty]$ defined by

$$\rho(x) \;=\; \limsup_{n\to\infty} [F_n(x)^{1/n}], \qquad C(x) \;=\; \sup_n [F_n(x)/r(x)^n].$$

Since $\rho$ and $C$ are measurable, so are the nested subsets

$$B_k \;:=\; \{x \in A : \rho(x) \leq 1 - \frac{1}{k},\; C(x) \leq k\}, \qquad k \in \mathbb{N}.$$

On the other hand, $(ii)$ implies that $\rho(x) < 1$ and $C(x) < \infty$ for each $x \in A$, so $\bigcup_k B_k = A$. Continuity of measures then implies that $\lim_{k\to\infty} \pi(B_k) = \pi(A) > 0$, so there is $K \in \mathbb{N}$ with $\pi(B_K) > 0$. By Lemma 4.1, there exists a small set $S \subseteq B_K$. It follows that for $x \in S$ and $n \in \mathbb{N}$,

$$\|P^n(x,\cdot) - \pi(\cdot)\|_{\mathrm{TV}} \;=\; F_n(x) \;\leq\; C(x)\,\rho(x)^n \;\leq\; K\,(1 - \frac{1}{K})^n,$$

and therefore

$$\begin{aligned}
\|\pi_S P^n(\cdot) - \pi(\cdot)\|_{\mathrm{TV}} \;&=\; \sup_D |\pi_S P^n(D) - \pi(D)| \\
&=\; \sup_D |\frac{1}{\pi(S)} \int_S [P^n(x,D) - \pi(D)]\pi(dx)| \\
&\leq\; \sup_D \sup_{x\in S} |P^n(x,D) - \pi(D)| \\
&=\; \sup_{x\in S} \|P^n(x,\cdot) - \pi(\cdot)\|_{\mathrm{TV}} \\
&\leq\; K\,(1 - \frac{1}{K})^n,
\end{aligned}$$

which shows $(v)$ with $C_S = K$ and $\rho_S = 1 - \frac{1}{K}$. □

**Proposition 5.6.** *(v) $\Rightarrow$ (vi).*

*Proof.* This is the content of the "$(i) \Rightarrow (ii)$" implication of [17, Theorem 15.0.1]. □

# 6 Proofs for $V$-function and $L_V^\infty$ Conditions

**Proposition 6.1.** *(xxiv) $\Rightarrow$ (xxiii).*

*Proof.* Immediate (just choose $j = 1$). □

**Proposition 6.2.** *(x) $\Rightarrow$ (xxiv).*

*Proof.* Let $f \in L_V^\infty$ such that $|f|_V = 1$. Then, $|f| \leq V$, and, if $(x)$ holds, for each $x \in \mathcal{X}$ and each $n \in \mathbb{N}$,

$$|P^n f(x) - \Pi(f)(x)| = |P^n f(x) - \pi(f)| \leq \sup_{|f|\leq V} |P^n(x,\cdot) - \pi(\cdot)| \leq CV(x)\rho^n,$$

which implies

$$|(P^n - \Pi)f|_V = \sup_{x\in\mathcal{X}} \frac{|P^n f(x) - \Pi(f)(x)|}{V(x)} = \sup_{x\in\mathcal{X}} \frac{|P^n f(x) - \pi(f)|}{V(x)} \leq C\rho^n,$$

and therefore,

$$\|P^n - \Pi\|_{L_V^\infty} = \sup_{\substack{f\in L_V^\infty \\ |f|_V=1}} |(P^n - \Pi)f|_V \leq C\rho^n. \qquad \square$$

**Proposition 6.3.** *(xxiii) ⇔ (xxv),* and *(xxiv) ⇔ (xxvi).*

*Proof.* (⇒) Let $n \in \mathbb{N}$. Given that $L_{V,0}^\infty \subseteq L_V^\infty$, $\|P^n\|_{L_{V,0}^\infty} \leq \|P^n - \Pi\|_{L_V^\infty} \leq C\,\rho^n$.

(⇐) If $f \in L_V^\infty$ such that $|f|_V = 1$, we have

$$\begin{aligned} |(P^n - \Pi)f|_V &= |(P^n - 1_{\mathcal{X}} \otimes P^n\pi)f|_V \\ &= |P^n f - (P^n\pi)f|_V \\ &= |P^n(f - \pi(f))|_V \\ &\leq |P^n(f - \pi(f))|_V \\ (f - \pi(f) \in L_{V,0}^\infty) \quad &\leq \|P^n\|_{L_{V,0}^\infty} \, |f - \pi(f))|_V \\ &\leq \|P^n\|_{L_{V,0}^\infty} \, (|f|_V + |\pi(f)|_V) \\ &\leq C\,\rho^n\,(1 + \pi(V)) \\ &\leq C'\,\rho^n, \end{aligned}$$

where $C' = C(1 + \pi(V)) < \infty$. $\square$

**Proposition 6.4.** (*xxiii*) ⇒ (*xxi*), and (*xxiv*) ⇒ (*xxii*).

*Proof.* Let $f \in L_{V,0}^\infty$. If (*xxiii*) holds,

$$|P^m f|_V = |P^m f - \pi(f)|_V = |(P^m - \Pi)f|_V,$$

and given this,

$$\|P^m\|_{L_{V,0}^\infty} = \sup_{\substack{f\in L_{V,0}^\infty \\ |f|_V=1}} |P^m f|_V = \sup_{\substack{f\in L_{V,0}^\infty \\ |f|_V=1}} |(P^m - \Pi)f|_V \leq \sup_{\substack{f\in L_V^\infty \\ |f|_V=1}} |(P^m - \Pi)f|_V = \|P^m - \Pi\|_{L_V^\infty} < 1.$$

Thus, (*xxi*) holds. $\square$

**Proposition 6.5.** *(xxi) ⇒ (xix),* and *(xxii) ⇒ (xx).*

*Proof.* Let $f$ such that $|f|_V \leq 1$ and let $k \in \mathbb{N}$. Let $s = \|P^m\|_{L^\infty_{V,0}} < 1$. Then,

$$\begin{aligned}
|(P^{mk} - \Pi)f|_V &= |P^{mk}(f - \pi(f))|_V \\
&\leq \left\|P^{mk}\right\|_{L^\infty_{V,0}} |f - \pi(f)|_V \\
&= \left\|(P^m)^k\right\|_{L^\infty_{V,0}} |f - \pi(f)|_V \\
&\leq s^k |f - \pi(f)|_V \\
&\leq s^k (|f|_V + |\pi(f)|_V) \\
&\leq s^k (1 + \pi(V)).
\end{aligned}$$

From Lemma 4.7, we must have $\pi(V) < \infty$. Hence, there exists $n \in \mathbb{N}$ such that $s^n(1+\pi(V)) < 1$. Thus, taking $m^* = mn$, we have that, for each $|f|_V \leq 1$,

$$|(P^{m^*} - \Pi)f|_V < 1$$

and therefore, $\left\|P^{m^*} - \Pi\right\|_{L^\infty_V} < 1$. □

**Proposition 6.6.** *(vii) $\Rightarrow$ (viii).*

*Proof.* First of all, we must have $\pi(V) < \infty$ by Lemma 4.7. Then, given $j \in \mathbb{N}$, let $\widehat{V} = V^{1/j}$, so $\pi(\hat{V}^j) = \pi(V) < \infty$. It follows from Jensen's inequality and concavity that

$$P\,\widehat{V} \;\leq\; (PV)^{1/j} \;\leq\; (\lambda V + b 1_S)^{1/j} \;\leq\; \hat{\lambda}\,\hat{V} + \hat{b}\,1_S,$$

with $\hat{\lambda} = \lambda^{1/j} < 1$ and $\hat{b} = b^{1/j} < \infty$, thus showing $(vii)$. □

**Proposition 6.7.** *(xix) $\Rightarrow$ (xxiii),* and *(xx) $\Rightarrow$ (xxiv).*

*Proof.* Let $m \in \mathbb{N}$ such that $s = \|P^m - \Pi\|_{L^\infty_V} < 1$. Let $\alpha = \|P - \Pi\|_{L^\infty_V}$, and let $n \in \mathbb{N}$. If $n \leq m$, we have that

$$\|P^n - \Pi\|_{L^\infty_V} = \|(P - \Pi)^n\|_{L^\infty_V} \leq \alpha^n \leq \alpha^n s^{-1} (s^{1/m})^n.$$

If $n > m$, then $n = mt + \ell$, for some $t \in \mathbb{N}$ and $0 \leq \ell < m$, and hence

$$\begin{aligned}
\|P^n - \Pi\|_{L^\infty_V} &= \left\|(P - \Pi)^{mt+\ell}\right\|_{L^\infty_V} \leq \alpha^l \left\|(P - \Pi)^{mt}\right\|_{L^\infty_V} \\
&= \alpha^l \left\|(P^m - \Pi)^t\right\|_{L^\infty_V} \leq \alpha^l s^t \leq \alpha^l s^{-1} (s^{1/m})^n.
\end{aligned}$$

So, taking $C = \max_{1 \leq r \leq m} \alpha^r s^{-1}$ and $\rho = s^{1/m}$, we conclude that for each $n \in \mathbb{N}$,

$$\|P^n - \Pi\|_{L^\infty_V} \leq C \rho^n.$$ □

**Proposition 6.8.** *(xxiii) $\Rightarrow$ (ix).*

*Proof.* Since $\|P^n - \Pi\|_{L_V^\infty} \le C\rho^n$, we have for each $n \in \mathbb{N}$ and $|f|_V \le 1$ that

$$|(P^n - \Pi)f|_V \le \|P^n - \Pi\|_{L_V^\infty} |f|_V \le C\rho^n |f|_V \le C\rho^n.$$

Hence, for each $n \in \mathbb{N}$, $|f|_V \le 1$ and $x \in \mathcal{X}$,

$$\frac{|P^n f(x) - \pi(f)|}{V(x)} = \frac{|P^n f(x) - \Pi(f)(x)|}{V(x)} \le C\rho^n.$$

By Lemma 4.7, $|f|_V \le 1 \Leftrightarrow |f| \le V$, so for each $n \in \mathbb{N}$ and $x \in \mathcal{X}$,

$$\sup_{|f|\le V} |P^n f(x) - \pi(f)| = \sup_{|f|_V \le 1} |P^n f(x) - \pi(f)| \le CV(x)\rho^n.$$

□

**Proposition 6.9.** *(ix) $\Rightarrow$ (xi),* and *(x) $\Rightarrow$ (xii).*

*Proof.* This follows from the triangle inequality. If $\mu(V) < \infty$ and $|f| \le V$, then

$$\begin{aligned} |\mu P^n f - \pi(f)| = \left| \int_{\mathcal{X}} P^n f(y)\mu(dy) - \pi(f) \right| &= \left| \int_{\mathcal{X}} P^n f(y)\mu(dy) - \int_{\mathcal{X}} \pi(f)\mu(dy) \right| \\ &\le \int_{\mathcal{X}} \left| P^n f(y) - \pi(f) \right| \mu(dy) \\ &\le \int_{\mathcal{X}} \sup_{|f|\le V} \left| P^n f(y) - \pi(f) \right| \mu(dy) \\ &\le \int_{\mathcal{X}} C\, V(y)\, \rho^n\, \mu(dy) \\ &= C\,\mu(V)\,\rho^n. \end{aligned}$$

Hence, $\sup_{|f|\le V} |\mu P^n f - \pi(f)| \le C\,\mu(V)\,\rho^n$ for all $n \in \mathbb{N}$. □

**Proposition 6.10.** *(viii) $\Rightarrow$ (x).*

*Proof.* This is the content of [18, Theorem 1], following [29]; proofs also appear in [17, Theorem 15.0.1(iii)] and [22, Theorem 9]. And since the same function $V$ is used in both conditions, its moments are preserved. □

**Proposition 6.11.** *(xi) $\Rightarrow$ (i).*

*Proof.* Let $\mu$ be a point-mass at $x$, so that $\mu(A) = 1$ if $x \in A$ otherwise $\mu(A) = 0$. Then $\mu(V) = V(x)$, so from $(i)$,

$$\|P^n(x,\cdot) - \pi(\cdot)\|_{\mathrm{TV}} \;=\; \|\mu P^n(\cdot) - \pi(\cdot)\|_{\mathrm{TV}} \;=\; \sup_{|f|\le V} \left|\mu P^n(f) - \pi(f)\right|$$

$$\leq \sup_{|f|\leq 1} \left|\mu P^n(f) - \pi(f)\right| \;\leq\; C\,\mu(V)\,\rho^n \;=\; C\,V(x)\,\rho^n\,.$$

Hence, $(i)$ holds with $C_x = C\,V(x)$. □

**Proposition 6.12.** *(vi) ⇒ (vii).*

*Proof.* The existence of a drift function $V$ satisfying the condition $(vii)$ follows from [17, Theorem 15.2.4]. □

**Proposition 6.13.** *(xii) ⇒ (iv).*

*Proof.* Let $p \in (1,\infty)$, and let $\mu \in L^p(\pi)$ be a probability measure. Let $j \in \mathbb{N}$ be large enough that $1 + \frac{1}{j} < p$, so that $\mu \in L^{1+\frac{1}{j}}(\pi)$ by Lemma 4.5. Then choose $V$ in $(xii)$ such that $\pi(V^{j+1}) < \infty$. Then, using the notation

$$\|f\|_r \;:=\; \left(\int_{\mathcal{X}} |f|^r \, d\pi\right)^{1/r}$$

for functions $f : \mathcal{X} \to \mathbb{R}$, since $\frac{1}{j+1} + \frac{1}{1+\frac{1}{j}} = 1$, we have by Hölder's inequality that

$$\begin{aligned}\mu(V) \;&=\; \int_{\mathcal{X}} V(x)\,\mu(dx) \;=\; \int_{\mathcal{X}} V(x)\left(\frac{d\mu}{d\pi}(x)\right)\pi(dx) \\ &\leq\; \|V\|_{j+1}\left\|\frac{d\mu}{d\pi}\right\|_{1+\frac{1}{j}} \;=\; \pi(V^{j+1})^{1/(j+1)}\;\|\mu\|_{L^{1+\frac{1}{j}}(\pi)} \;<\; \infty.\end{aligned}$$

Then,

$$\begin{aligned}\|\mu P^n(\cdot) - \pi(\cdot)\|_{\mathrm{TV}} \;&=\; \frac{1}{2}\sup_{|f|\leq 1} |\mu P^n(f) - \pi(f)| \\ &\leq\; \frac{1}{2}\sup_{|f|\leq V} |\mu P^n(f) - \pi(f)| \\ &\leq\; \frac{1}{2}\,C\,\mu(V)\,\rho^n,\end{aligned}$$

so $(iv)$ holds with $C_{p,\mu} = \dfrac{1}{2}\,C\,\mu(V) < \infty$. □

# 7 Proofs for Spectral Conditions

**Proposition 7.1.** *(xiii) ⇔ (xvii),* and *(xiv) ⇔ (xviii).*

*Proof.* (⇒) Since 1 is an eigenvalue of multiplicity 1, with corresponding eigenvectors the non-zero constant functions which are not in $L^\infty_{V,0}$, we must have $\mathcal{S}_{L^\infty_{V,0}}(P) \subseteq \mathcal{S}_{L^\infty_V}(P) \setminus \{1\}$. So, if (*xiii*) holds, then $\mathcal{S}_{L^\infty_{V,0}}(P) \subseteq \mathcal{S}_{L^\infty_V}(P) \setminus \{1\} \subseteq \{z \in \mathbb{C} : |z| \leq \rho\}$ for some $\rho < 1$. This implies that $r_{L^\infty_{V,0}}(P) \leq \rho < 1$.

(⇐) If $\rho := r_{L^\infty_{V,0}}(P) < 1$, then since $\mathcal{S}_{L^\infty_V}(P) \setminus \{1\} \subseteq \mathcal{S}_{L^\infty_{V,0}}(P)$ by Lemma 4.7, we have

$$\mathcal{S}_{L^\infty_V}(P) \setminus \{1\} \;\subseteq\; \mathcal{S}_{L^\infty_{V,0}}(P) \;\subseteq\; \{z \in \mathbb{C} : |z| \leq \rho\}.$$

□

**Proposition 7.2.** *(xvii) ⇔ (xxi),* and *(xviii) ⇔ (xxii).*

*Proof.* (⇒) By the spectral radius formula ([27], Theorem 10.13), $\rho = r(P|_{L^\infty_{V,0}}) = \inf_{n\geq 1} \|P^n\|^{1/n}_{L^\infty_{V,0}}$. Hence, for $\rho_0 < 1$ such that $\rho < \rho_0$, there exists $m \in \mathbb{N}$ such that $\|P^m\|_{L^\infty_{V,0}} < \rho^m < 1$.
(⇐) If $\|P^m\|_{L^\infty_{V,0}} < 1$ for some $m \in \mathbb{N}$,

$$r(P|_{L^\infty_{V,0}}) = \inf_{n\geq 1} \|P^n\|^{1/n}_{L^\infty_{V,0}} \leq \|P^m\|^{1/m}_{L^\infty_{V,0}} < 1,$$

and thus, (*xxi*) holds. □

**Proposition 7.3.** *(xv) ⇔ (xix),* and *(xvi) ⇔ (xx).*

*Proof.* (⇒) Given that $\rho_0 = r(P - \Pi) = \inf_{n\geq 1} ||P^n - \Pi||^{1/n}_{L^\infty_V}$, for $\rho_0 < \rho < 1$, there exists $m \in \mathbb{N}$ such that $||P^m - \Pi||_{L^\infty_V} < \rho^m < 1$. Therefore, for some $m \in \mathbb{N}$,

$$||P^m - \Pi||_{L^\infty_V} < 1.$$

(⇐) If $||P^m - \Pi||_{L^\infty_V} < 1$ for some $m \in \mathbb{N}$, given that $r(P - \Pi) = \inf_{n\geq 1} \|P^n - \Pi\|^{1/n}_{L^\infty_V}$, we have

$$r(P - \Pi) = \inf_{n\geq 1} \|P^n - \Pi\|^{1/n}_{L^\infty_V} \leq \|P^m - \Pi\|^{1/m}_{L^\infty_V} < 1.$$

□

# 8 Proofs for Reversible Conditions

**Proposition 8.1.** *(xxxi) ⇒ (xxviii).*

*Proof.* Let $\rho = \|P - \Pi\|_{L^2(\pi)} < 1$. Then for each signed measure $\mu \in L^2(\pi)$,

$$\|\mu(P - \Pi)(\cdot)\|_{L^2(\pi)} \leq \rho\|\mu\|_{L^2(\pi)}.$$

Let $\mu \in L^2(\pi)$ be a probability measure and let $n \in \mathbb{N}$. By Lemma 4.11, $\mu - \mu(\mathcal{X})\pi = \mu - \pi$ is orthogonal to $\pi$, so $(\mu - \pi)\Pi = 0$, and hence

$$\mu P^n - \pi = (\mu - \pi)P^n = (\mu - \pi)(P^n - \Pi) = (\mu - \pi)(P - \Pi)^n.$$

Therefore,

$$\begin{aligned}\|\mu P^n(\cdot) - \pi(\cdot)\|_{L^2(\pi)} &= \|(\mu - \pi)(P - \Pi)^n(\cdot)\|_{L^2(\pi)} \\ &\le \|\mu - \pi\|_{L^2(\pi)} \|P^n - \Pi\|_{L^2(\pi)} \\ &\le \|\mu - \pi\|_{L^2(\pi)} \rho^n.\end{aligned}$$

□

**Proposition 8.2.** *(xxviii) $\Rightarrow$ (xxvii).*

*Proof.* If (*xxviii*) holds, for each probability measure $\mu \in L^2(\pi)$ and $n \in \mathbb{N}$,

$$\|\mu P^n(\cdot) - \pi(\cdot)\|_{L^2(\pi)} \le \|\mu - \pi\|_{L^2(\pi)}\rho^n = C_\mu \rho^n,$$

with $C_\mu = \|\mu - \pi\|_{L^2(\pi)}$. □

**Proposition 8.3.** *(xxvii) $\Rightarrow$ (iii).*

*Proof.* If (*xxvii*) holds, then by Lemmas 4.3 and 4.4, for each $n \in \mathbb{N}$ and $\mu \in L^2(\pi)$ we have

$$\|\mu P^n(\cdot) - \pi(\cdot)\|_{\mathrm{TV}} \;=\; \frac{1}{2}\|\mu P^n(\cdot) - \pi(\cdot)\|_{L^1(\pi)} \;\le\; \frac{1}{2}\|\mu P^n(\cdot) - \pi(\cdot)\|_{L^2(\pi)} \;\le\; \frac{1}{2} C_\mu \rho^n.$$

This shows (*iii*) with $p = 2$ and $\rho_\mu = \rho$. □

**Proposition 8.4.** *(iv) $\Rightarrow$ (xxxii).*

*Proof.* Take $p = 2$ in (*iv*). Then it follows from the "(*iii*) $\Rightarrow$ (*ii*)" implication of [21, Theorem 2] (which is proven by contradiction, using reversibility and the spectral measure of $P$ acting on $L^2(\pi)$) that there is $\rho < 1$ such that

$$\|\mu P\|_{L^2(\pi)} \;\le\; \rho\, \|\mu\|_{L^2(\pi)}$$

for all probability measures $\mu \in L^2(\pi)$ with $\mu(\mathcal{X}) = 0$. Hence, $\|P\|_{\pi^\perp} \le \rho < 1$. □

**Proposition 8.5.** *(xxxii) $\Leftrightarrow$ (xxxiii).*

*Proof.* This follows immediately from the fact (e.g. [3, Proposition VIII.1.11(e)]) that, by reversibility, $r_{\pi^\perp}(P) = ||P||_{\pi^\perp}$. □

**Proposition 8.6.** *(xxxii) ⇒ (xxxi).*

*Proof.* From Lemma 4.12 it follows that

$$\|P - \Pi\|_{L^2(\pi)} \;=\; \|P\|_{\pi^\perp}.$$

Hence, if $\|P\|_{\pi^\perp} < 1$, then $\|P - \Pi\|_{L^2(\pi)} < 1$. □

**Proposition 8.7.** *(xxx) ⇔ (xxxi).*

*Proof.* Since $P$ is reversible, $P - \Pi$ is self-adjoint by Lemma 4.11. Therefore, $r_{L^2(\pi)}(P - \Pi) = ||P - \Pi||_{L^2(\pi)}$ (e.g. [3, Proposition VIII.1.11(e)]). Hence, $||P - \Pi||_{L^2(\pi)} < 1$ if and only if $r_{L^2(\pi)}(P - \Pi) < 1$. □

**Proposition 8.8.** *(xxix) ⇔ (xxxiii).*

*Proof.* (⇒) If (*xxix*) holds, there is $\rho < 1$ such that

$$\mathcal{S}_{L^2(\pi)}(P) \;\subseteq\; \{1\} \cup \{\lambda \in \mathbb{C} : |\lambda| \leq \rho\}.$$

Since 1 is an eigenvalue of multiplicity 1, with corresponding eigenvectors the non-zero constant multiples of $\pi$ which are not in $\pi^\perp$, we must have $\mathcal{S}_{\pi^\perp}(P) \subseteq \mathcal{S}_{L^2(\pi)}(P) \setminus \{1\}$. Hence, $\mathcal{S}_{\pi^\perp}(P) \subseteq \{\lambda \in \mathbb{C} : |\lambda| \leq \rho\}$. Therefore, $r(P|_{\pi^\perp}) \leq \rho < 1$.

(⇐) If $r_{\pi^\perp}(P) < 1$, there is $\rho < 1$ with $\mathcal{S}_{\pi^\perp}(P) \subseteq \{\lambda \in \mathbb{C} : |\lambda| \leq \rho\}$. So, by Lemma 4.11,

$$\mathcal{S}_{L^2(\pi)}(P) \setminus \{1\} \;\subseteq\; \mathcal{S}_{\pi^\perp}(P) \;\subseteq\; \{\lambda \in \mathbb{C} : |\lambda| \leq \rho\}.$$ □

# 9 Future Directions and Open Problems

Our Theorem 1 above provides a fairly complete picture of equivalences of geometric ergodicity. However, it does lead to some additional questions which remain, including:

**Q9.1.** We have assumed throughout that the chain is *$\phi$-irreducible* and *aperiodic*. Those properties are certainly required for, and implied by, geometric ergodicity. But do they need to be assumed explicitly? Many of our equivalent conditions *imply* them, so that they do not actually need to be mentioned. But some of our conditions do not, e.g. the drift conditions (*vii*) and (*viii*). So, which of our equivalences continue to hold without assuming $\phi$-irreducibility and aperiodicity?

**Q9.2.** Related, does the assumption of $\phi$-irreducibility directly imply the "multiplicity 1" assumption in the spectral gap conditions (*xiii*) and (*xiv*) and (*xxix*), so that it does not need to be mentioned explicitly? (As a start, combining e.g. Theorems 10.0.1 and 10.1.1 of [17] shows that $\phi$-irreducibility implies that the stationary probability distribution $\pi$ must be unique; does this uniqueness extend to signed measures and functions too?)

**Q9.3.** We also assumed that our state space $(\mathcal{X}, \mathcal{F})$ is *countably generated*, which holds for e.g. the Borel subsets of $\mathbb{R}$ and of $\mathbb{R}^d$, but *not* for e.g. the Lebesgue-measurable subsets. It is a very standard assumption (e.g. [17, p. 66]), used to ensure the existence of small sets [4, 9, 19] and the measurability of certain functions (e.g. [21, Appendix]). But which of our equivalences would continue to hold without it?

**Q9.4.** The property of *aperiodicity* is not necessary for other important properties such as Central Limit Theorems which involve averages of functional values like $\frac{1}{M}\sum_{i=1}^{M} h(X_i)$. The weaker notion of *variance bounding* essentially corresponds to geometric ergodicity without aperiodicity, and still implies CLTs. Many equivalences to variance bounding have been proven for *reversible* chains; see [23]. But can equivalences similar to our Theorem 1 be derived for the variance bounding property without assuming reversibility?

**Q9.5.** Our later conditions (*xxvii*) through (*xxxiii*) were only shown to be equivalent for *reversible* chains. But are there explicit counter-examples to show that they are *not* equivalent in the absence of reversibility? Or are some of them are still equivalent to geometric ergodicity, even without assuming reversibility? (For a start on this, [14, Theorem 1.3] proves that without reversibility the implication $(xxx) \Rightarrow (i)$ still holds, but [14, Theorem 1.4] makes use of [6] to show that the converse might fail.)

**Q9.6.** Our equivalences are for the fairly strong property of geometric ergodicity. But are there similar equivalences for the even stronger property of *uniform ergodicity*, i.e. the property that $\|P^n(x,\cdot) - \pi(\cdot)\|_{\mathrm{TV}} \leq C\,\rho^n$ from $\pi$-a.e. $x \in \mathcal{X}$ where $C$ does not depend on $x$? (For a start on this, see [17, Theorem 16.0.2].)

**Q9.7.** In the other direction, are there similar equivalences for the weaker property of *polynomial ergodicity*, i.e. the property that $\|P^n(x,\cdot) - \pi(\cdot)\|_{\mathrm{TV}} \leq C_x\, n^{-\alpha}$ for some $\alpha > 0$? (For some discussion and results related to this property, see e.g. [5, 10].)

**Q9.8.** And, are there similar equivalences for the even weaker property of *simple ergodicity*, i.e. the property that just $\|P^n(x,\cdot) - \pi(\cdot)\|_{\mathrm{TV}} \to 0$ as $n \to \infty$ from $\pi$-a.e. $x \in \mathcal{X}$, without specifying any rate? (For a start on this, see e.g. [17, Theorem 13.0.1].)

We leave these questions as open problems for future work.

**Acknowledgements.** We thank Jim Hobert, Galin Jones, and Gareth Roberts for encouraging us to write this paper.

## References

[1] R. C. Bradley. An exposition of some basic features of strictly stationary, reversible Markov chains. *Journal of Time Series Analysis*, 42(5–6):499–533, 2021.

[2] S. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng, editors. *Handbook of Markov chain Monte Carlo.* Chapman & Hall, 2011.

[3] J. B. Conway. *A course in functional analysis, 2nd ed.* Springer Science Business Media, New York, 1990.

[4] W. Doeblin. Eléments d'une théorie générale des chaînes simples constantes de Markov. *Annales Scientifiques de l'Ecole Normale Supérieure*, 57(III):61–111, 1940.

[5] G. Fort and E. Moulines. Polynomial ergodicity of markov transition kernels. *Stoch. Proc. Appl.*, 103:57–99, 2003.

[6] O. Häggström. On the central limit theorem for geometrically ergodic Markov chains. *Prob. Th. Rel. Fields*, 132(1):74–82, 2005.

[7] J. P. Hobert, G. L. Jones, B. Presnell, and J. S. Rosenthal. On the applicability of regenerative simulation in Markov chain Monte Carlo. *Biometrika*, 89:731–743, 2002.

[8] I. A. Ibragimov and Y. V. Linnik. *Independent and stationary sequences of random variables.* Wolters-Noordhoff, 1971.

[9] N. Jain and B. Jamison. Contributions to doeblin's theory of Markov processes. *Z. Wahrsch. Verw. Geb.*, 8:19–40, 1967.

[10] S. F. Jarner and G. O. Roberts. Polynomial convergence rates of Markov chains. *Ann. Appl. Prob.*, 12:224–247, 2002.

[11] G. L. Jones. On the Markov chain central limit theorem. *Probability surveys*, 1:299–320, 2004.

[12] G. L. Jones and J. P. Hobert. Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Stat. Sci.*, 16:312–334, 2001.

[13] I. Kontoyiannis and S. P. Meyn. Spectral theory and limit theorems for geometrically ergodic Markov processes. *Ann. Appl. Prob.*, 13(1):304–362, 2003.

[14] I. Kontoyiannis and S. P. Meyn. Geometric ergodicity and the spectral gap of non-reversible Markov chains. *Prob. Th. Rel. Fields*, 154(1):327–339, 2009.

[15] J. S. Liu. Metropolized independent sampling with comparisons to rejection sampling and importance sampling. *Statistics and Computing*, 6(2):113–119, 1996.

[16] K. L. Mengersen and R. L. Tweedie. Rates of convergence of the Hastings and Metropolis algorithms. *Annals of Statistics*, 24:101–121, 1996.

[17] S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Springer-Verlag, London, 1993. Available at: probability.ca/MT.

[18] E. Nummelin and R. L. Tweedie. Geometric ergodicity and R-positivity for general Markov chains. *The Annals of Probability*, 6(3):404–420, 1978.

[19] S. Orey. *Lecture notes on limit theorems for Markov chain transition probabilities*. Van Nostrand Reinhold, 1971.

[20] G. O. Roberts. A note on acceptance rate criteria for CLTs for Metropolis-Hastings algorithms. *J. Appl. Prob.*, 36:1210–1217, 1999.

[21] G. O. Roberts and J. S. Rosenthal. Geometric ergodicity and hybrid Markov chains. *Electronic Communications in Probability*, 2:13–25, 1997.

[22] G. O. Roberts and J. S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probability surveys*, 1:20–71, 2004.

[23] G. O. Roberts and J. S. Rosenthal. Variance bounding Markov chains. *Annals of Applied Probability*, 18(3):1201–1214, 2008.

[24] G. O. Roberts and R. L. Tweedie. Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, 83:95–110, 1996.

[25] G. O. Roberts and R. L. Tweedie. Geometric L2 and L1 convergence are equivalent for reversible Markov chains. *Journal of Applied Probability*, 38:37–41, 2001.

[26] J. S. Rosenthal. Quantitative convergence rates of Markov chains: A simple account. *Electronic Communications in Probability*, 7:123–128, 2002.

[27] W. Rudin. *Functional Analysis*. International series in pure and applied mathematics. McGraw-Hill, 1991.

[28] L. Tierney. Markov chains for exploring posterior distributions. *Annals of Statistics*, 22:1701–1728, 1994.

[29] D. Vere-Jones. Geometric ergodicity in denumerable Markov chains. *The Quarterly Journal of Mathematics*, 13(1):7–28, 1962.