Poincaré inequalities for Markov chains: a meeting with Cheeger, Lyapunov and Metropolis

Christophe Andrieu, Anthony Lee, Sam Power, Andi Q. Wang

School of Mathematics, University of Bristol

August 11, 2022

Abstract

We develop a theory of weak Poincaré inequalities to characterize convergence rates of ergodic Markov chains. Motivated by the application of Markov chains in the context of algorithms, we develop a relevant set of tools which enable the practical study of convergence rates in the setting of Markov chain Monte Carlo methods, but also well beyond.

Contents

1 Introduction

This report is the result of a research programme initiated in [1] that aims to understand and develop functional-analytic tools to characterize the rate of convergence to equilibrium of discrete-time Markov chains. While analysis of the right-spectral gap of time-reversible Markov chains is fairly standard and has played an important rôle in the analysis of Markov chain Monte Carlo (MCMC) algorithms, functional-analytic results for nonreversible or subgeometrically convergent Markov chains are scarce. Notable exceptions are [15] and [11], the latter being the closest in spirit to our work. On the other hand, the characterization of the convergence to equilibrium of *continuous-time* processes, both reversible and nonreversible, geometric and subgeometric, is considerably more developed. Study of subgeometric rates of convergence can be traced back to [27], which was later generalized and developed in [38], with a general framework relying on *weak Poincaré inequalities* (WPIs). Further significant contributions to the analysis of diffusion processes were made by the French school in the late 2000s – early 2010s in a series of contributions, for instance [3, 4, 9, 8].

Beyond the scattered nature of this literature, the continuous-time scenario possesses a plethora of specific technical difficulties, which often render it difficult to penetrate for the uninitiated. On the other hand, while the discrete-time Markov chain setup is indeed technically simpler, it has its own subtleties and challenges, which have not thus far been covered in a comprehensive way in the literature. As such, many of our present results are not merely transpositions of existing continuous-time results into the discrete-time setting.

Importantly, the main motivation behind our work being our interest in MCMC methods – and more generally algorithms which utilize ergodic Markov chains – we address numerous questions not addressed in the existing literature, concerning for example optimality and comparison of Markov chains. Our own recent experience shows that these functional-analytic tools we develop are complementary to the classical drift and minorization approach, à la Meyn and Tweedie [32], which has proved particularly useful and fruitful in the context of MCMC algorithms. We provide several concrete examples and applications of our techniques which are relevant for the analysis of MCMC methods; in particular we have been able to answer some questions (see, for instance, [1] or Subsection 5.3) which had eluded us and others previously.

1.1 A roadmap

Beyond an attempt to develop a coherent and self-contained document on WPIs for Markov chains, we also make a number of novel contributions.

This manuscript can be summarized as follows:

- Section 2 focuses on definitions of weak Poincaré inequalities (WPIs) in the discrete-time setting and their immediate implications. In Subsection 2.1 three equivalent parametrizations of WPIs are discussed in detail and we summarise their implications for rates of convergence to equilibrium. In

Subsection 2.2, we connect convergence for bounded functions in L^2 with convergence of L^p functions. We establish in Subsection 2.3 reverse implications: showing that a given rate of convergence implies the existence of a WPI. In Subsection 2.4 we show how WPIs can be used to bound directly the asymptotic variance of ergodic averages. In Subsection 2.5 we draw links between WPIs and subgeometric rates of convergence with spectral properties of the operators involved.

- Section 3 is dedicated to the notion of *optimal* WPIs (Subsection 3.1), lower bounds on rates of convergence (Subsection 3.2), comparison results of the Peskun–Tierney type (Subsection 3.3), optimal sieve functionals (Subsection 3.4) and a form of duality (Subsection 3.5).

- Section 4 develops practical tools for establishing WPIs in practice. In Subsection 4.1 we generalize Cheeger inequalities for Markov chains to establish WPIs. In Subsection 4.2 we establish links between $\mu-$irreducibility and the existence of WPIs via the abstract *RUPI* condition. Subsection 4.3 discusses connections between drift and minorization techniques with Poincaré inequalities. We discuss an alternative strategy to establish WPIs: a *local* Poincaré inequality for a *restricted* version of the Markov chain is combined with a drift condition. Finally in Subsection 4.4, we study how the knowledge of SPIs for restricted versions of a given Markov chain can be used to deduce WPIs for the unrestricted chain.

- In Section 5 we present applications of the theory in particular scenarios. In Subsections 5.1–5.2 we establish lower bounds on the rate of convergence of a type of pseudo-marginal algorithm and the random walk Metropolis (RWM) algorithm targeting heavy-tailed distributions. In Subsection 5.3 we establish dimension dependence of d^{-1} of the spectral gap of the RWM algorithm for a class of light-tailed target distributions, effectively providing the first direct proof of this result. This result is specialized to the Gaussian scenario in Subsection 5.4. In Subsection 5.5 we show how our results can be used to establish the existence of a central limit theorem for ergodic averages.

- Finally the Appendix contains some deferred proofs and miscellaneous results omitted from the main body of the text.

The highlights of this report will ultimately be turned into standard, more succinct and focussed manuscripts for specialists.

1.2 Notation

We will write $\mathbb{N} = \{1, 2, \dots\}$ for the set of natural numbers, $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$, and $\mathbb{R}_+ = (0, \infty)$ for positive real numbers.

Outside of specific examples, we will be working throughout on a general measurable space $(\mathsf{E}, \mathscr{E})$.

- For a set $A \in \mathscr{E}$, its complement in E is denoted by $A^{\complement}$. We denote the corresponding indicator function by $\mathbf{1}_A : \mathsf{E} \to \{0, 1\}$.

- We assume that $(\mathsf{E}, \mathscr{E})$ is equipped with a probability measure μ, and write $\mathrm{L}^2(\mu)$ for the Hilbert space of (equivalence classes of) real-valued μ–square-integrable measurable functions with inner product

$$\langle f, g \rangle = \int_{\mathsf{E}} f(x)\, g(x)\, \mathrm{d}\mu(x) \,,$$

  and corresponding norm $\|\cdot\|_{2,\mu}$, and if there is no ambiguity, we may just write $\|\cdot\|_2$. We write $\mathrm{L}^2_0(\mu)$ for the set of functions $f \in \mathrm{L}^2(\mu)$ which also satisfy $\mu(f) = 0$.

- More generally, for $p \in [1, \infty)$, we write $\mathrm{L}^p(\mu)$ for the Banach space of real-valued measurable functions with finite p-norm, $\|f\|_p := \left(\int_{\mathsf{E}} |f|^p \,\mathrm{d}\mu\right)^{1/p}$, and $\mathrm{L}^p_0(\mu)$ for $f \in \mathrm{L}^p(\mu)$ with $\mu(f) = 0$.

- We assume that the diagonal is measurable in $\mathsf{E} \times \mathsf{E}$, i.e. $\{(x, x) : x \in \mathsf{E}\} \in \mathscr{E} \otimes \mathscr{E}$. This assumption holds, for instance, on a Polish space endowed with its Borel σ-algebra.

- We write $\mathscr{E}_+ := \{A \in \mathscr{E} : \mu(A) > 0\}$.

- For μ and ν probability measures on $(\mathsf{E}, \mathscr{E})$, we let $\|\mu - \nu\|_{\mathrm{TV}} := \sup_{A \in \mathscr{E}} |\mu(A) - \nu(A)|$.

- For a measurable function $f : \mathsf{E} \to \mathbb{R}$, let $\|f\|_{\mathrm{osc}} := \mathrm{ess}_\mu \sup f - \mathrm{ess}_\mu \inf f$.

- For two probability measures μ and ν on $(\mathsf{E}, \mathscr{E})$ we let $\mu \otimes \nu(A \times B) = \mu(A)\,\nu(B)$ for $A, B \in \mathscr{E}$. For a Markov kernel $P(x, \mathrm{d}y)$ on $\mathsf{E} \times \mathscr{E}$, we write for $\bar{A} \in \mathscr{E} \otimes \mathscr{E}$, the minimal product σ-algebra, $\mu \otimes P\left(\bar{A}\right) = \int_{\bar{A}} \mu(\mathrm{d}x)\, P(x, \mathrm{d}y)$.

- A point mass distribution at x will be denoted by $\delta_x(\mathrm{d}y)$.

- $\mathrm{Id} : \mathrm{L}^2(\mu) \to \mathrm{L}^2(\mu)$ denotes the identity mapping, $f \mapsto f$. We also use this symbol for the identity $\mathrm{Id} \colon \mathsf{X} \to \mathsf{X}$.

- Given a bounded linear operator $T : \mathrm{L}^2(\mu) \to \mathrm{L}^2(\mu)$, we let $\mathcal{E}(T, f)$ be the Dirichlet form defined by $\langle (\mathrm{Id} - T)\, f, f \rangle$ for any $f \in \mathrm{L}^2(\mu)$.

- For such an operator T, we write T^* for its adjoint operator $T^* : \mathrm{L}^2(\mu) \to \mathrm{L}^2(\mu)$, which satisfies $\langle f, Tg \rangle = \langle T^* f, g \rangle$ for any $f, g \in \mathrm{L}^2(\mu)$.

- For such an operator T, we denote its spectrum by $\sigma(T)$. We denote the spectrum of the restriction of T to $\mathrm{L}^2_0(\mu)$ by $\sigma_0(T)$.

- For a μ-invariant Markov kernel T we let the right-spectral gap be

$$\mathrm{Gap}_{\mathrm{R}}(T) := \inf_{g \in \mathrm{L}^2_0(\mu), g \neq 0} \frac{\mathcal{E}(T, g)}{\|g\|_2^2} \,.$$

- For a given $f \in \mathrm{L}^2(\mu)$, the asymptotic variance is defined as $\mathrm{var}(T, f) := \lim_{n\to\infty} n\mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n} T^n f\right)$.
- We will write $a \wedge b$ to mean the (pointwise) minimum of real-valued functions a, b and $a \vee b$ for the corresponding maximum. For $s \in \mathbb{R}$, we will write $(s)_+ := s \vee 0$ for the positive part.
- $\inf A$ denotes the infimum of set $A \subset \mathbb{R}$ and $\inf \emptyset = \infty$.
- For a norm $|\cdot|$, which will always be clear from the context, we define the closed ball of radius r around x to be

$$\mathcal{B}(x, r) := \{y \in \mathsf{E} : |y - x| \leq r\}.$$

- We adopt the following $\mathcal{O}$ (resp. Ω) notation to indicate when functions grow no faster than (resp. no slower than) other functions. For $a \in \mathbb{R} \cup \{\infty\}$
  - If $f(x) \in \mathcal{O}(g(x))$ as $x \to a$, this means $\limsup_{x\to a} \left|\frac{f(x)}{g(x)}\right| < \infty$. When $a = +\infty$ then we may drop explicit mention of a.
  - If $f(x) \in \Omega(g(x))$ as $x \to a$, this means $\liminf_{x\to a} \left|\frac{f(x)}{g(x)}\right| > 0$. In particular $f \in \mathcal{O}(g) \iff g \in \Omega(f)$.

2 Fundamentals

2.1 Definitions and basic properties

We first give the basic definitions needed in order to define a weak Poincaré inequality.

Definition 1. a). We call a functional $\Phi : \mathrm{L}^2(\mu) \to [0, \infty]$ a *sieve functional*, or *sieve*, if for any $f \in \mathrm{L}^2(\mu)$, $c > 0$, it holds that

$$\Phi(cf) = c^2\Phi(f), \quad \|f - \mu(f)\|_2^2 \leqslant \mathfrak{a}\Phi(f - \mu(f)),$$

for a finite constant $\mathfrak{a} := \sup_{f \in \mathrm{L}_0^2(\mu)\setminus\{0\}} \|f\|_2^2 / \Phi(f)$.

b). Let P be a μ-invariant Markov kernel. We say that a sieve is *P-non-expansive* if $\Phi(Pf) \leqslant \Phi(f)$ for $f \in \mathrm{L}_0^2(\mu)$.

For simplicity and when no ambiguity is possible, we may refer to a P-non-expansive sieve simply as a sieve.

Example 2. Our main example of a P–non-expansive sieve, for any P, is $\Phi = \|\cdot\|_{\mathrm{osc}}^2$, with $\mathfrak{a} \leq 1$.

There are two ways to parameterize weak Poincaré inequalities for P, which are equivalent under a mild assumption.

Definition 3. We say that a μ-reversible kernel T satisfies a (Φ,α) -*weak Poincaré inequality*, abbreviated (Φ,α) -WPI, if for a sieve Φ and a decreasing function $\alpha:(0,\infty)\to[0,\infty)$,

$$\|f\|_2^2 \le \alpha(r)\,\mathcal{E}(T,f) + r\Phi(f), \quad \forall r>0, f\in \mathrm{L}_0^2(\mu). \tag{1}$$

Secondly, using the same notation, we can parameterize in terms of β: we say that a (Φ,β) -WPI holds if:

$$\|f\|_2^2 \le s\mathcal{E}(T,f) + \beta(s)\,\Phi(f), \quad \forall s>0, f\in \mathrm{L}_0^2(\mu), \tag{2}$$

where $\beta:(0,\infty)\to[0,\infty)$ is a decreasing function with $\beta(s)\to 0$ as $s\to\infty$.

If T satisfies a (Φ,α)-WPI or a (Φ,β)-holds but the specific α or β are not relevant, we may say that a Φ-WPI holds.

In practice, we are interested in bounding the convergence to equilibrium of a given μ-invariant Markov kernel, P. To obtain such bounds, in the framework of Definition 3, we will take $T=P^*P$, or if P is μ-reversible, we may take directly $T=P$.

Remark 4. Given a general μ-invariant Markov kernel T (which is not necessarily reversible), one can still define a WPI for T, namely the requirement that (1) holds for our general kernel T. However, it is enough to define (1) only for reversible kernels, since

$$\begin{aligned}\mathcal{E}(T,f) &= \langle(\mathrm{Id}-T)f,f\rangle\\ &= \langle(\mathrm{Id}-(T+T^*)/2)f,f\rangle\\ &= \mathcal{E}((T+T^*)/2,f),\end{aligned}$$

due to the fact that $(T-T^*)/2$ is antisymmetric, and we are considering real-valued f. Since the kernel $(T+T^*)/2$ is reversible, it is thus sufficient to consider WPIs for reversible kernels.

For any decreasing function $F\colon\mathbb{R}_+\to\mathbb{R}$ we let $F^{-}:\mathbb{R}\to[0,\infty]$ given by $F^{-}(x) := \inf\{y>0\colon F(y)\leqslant x\}$, for $x\in\mathbb{R}$, be its generalized inverse. The following proposition shows that one can straightforwardly move between the two formulations of WPIs.

Proposition 5. *Let P be a Markov kernel on $(\mathsf{E},\mathscr{E})$, Φ be a sieve, and $\mathfrak{a} := \sup_{f\in\mathrm{L}_0^2(\mu)\setminus\{0\}}\|f\|_2^2/\Phi(f)$.*

a). If a (Φ,α) -WPI holds with $\alpha(r)=0$ for $r\geqslant\mathfrak{a}$, then a (Φ,β) -WPI, with $\beta:=\alpha^{-}$ on $(0,\infty)$, holds and for any $r,s\geqslant 0$,

i $\alpha^{-}\circ\alpha(r)\leqslant r$ with equality when α is strictly decreasing;

ii $s\leqslant\alpha\circ\alpha^{-}(s)$ if α is right continuous;

iii $\beta\leqslant\mathfrak{a}$.

b). If a (Φ,β) -WPI holds with $\beta\leqslant\mathfrak{a}$, then a (Φ,α) -WPI holds, with $\alpha:=\beta^{-}$ on $(0,\infty)$, and for any $r,s>0$, *i* $\beta^{-} \circ \beta(s) \leqslant s$;

ii $r \leqslant \beta \circ \beta^{-}(r)$ *if* β *is right continuous;*

iii $\alpha(r) = 0$ *for* $r \geqslant \mathfrak{a}$.

c). If α *in* a) *(resp.* β *in* b)*) is right continuous then* $(\alpha^{-})^{-} = \alpha$ *(resp.* $(\beta^{-})^{-} = \beta$*); that is, the two parametrizations are equivalent.*

Proof. Statement a). Assume that a (Φ, α)-WPI holds, let $s > 0$ and $\mathfrak{R}(s) := \{r > 0 : \alpha(r) \leqslant s\} \neq \emptyset$, where the nonemptiness follows from the assumption on α. Then for any $r \in \mathfrak{R}(s)$, it holds that

$$\|f\|_2^2 \leqslant s\mathcal{E}(T, f) + r\Phi(f),$$

and therefore

$$\begin{aligned}\|f\|_2^2 &\leqslant \inf\{s\mathcal{E}(T, f) + r\Phi(f) : r \in \mathfrak{R}(s)\} \\ &= s\mathcal{E}(T, f) + \alpha^{-}(s)\Phi(f).\end{aligned}$$

Note that $\alpha(r) = 0$ for $r \geqslant \mathfrak{a}$ implies that for any $s > 0$,

$$\alpha^{-}(s) := \inf\{r > 0 : \alpha(r) \leqslant s\} = \inf\{r \in (0, \mathfrak{a}] : \alpha(r) \leqslant s\} \leqslant \mathfrak{a}.$$

We use the results of [14], stated for an increasing function T, but directly applicable here by setting, using their notation, $\mathsf{T} = -\alpha$ and noting that $\alpha^{-}(s) = \mathsf{T}^{-}(-s)$. From [14, Proposition 1, (2)], α^{-} is decreasing.

For any $\varepsilon \geqslant 0$, let $s(\varepsilon) := \sup_{r > \varepsilon} \alpha(r)$. If $s(0) < \infty$, then $\alpha^{-}(s) = 0$ for $s > s(0)$. Otherwise, $\lim_{\varepsilon \downarrow 0} s(\varepsilon) = \infty$, since α is decreasing. Therefore for any $\varepsilon > 0$ and any $s \geqslant s(\varepsilon)$, we have $\inf\{r > 0 : \alpha(r) \leqslant s\} \leqslant \inf\{r > 0 : \alpha(r) \leqslant s(\varepsilon)\} \leqslant \varepsilon$ and $\alpha^{-}(s) \leq \alpha^{-}(s(\varepsilon)) \leqslant \varepsilon$. Hence, $\lim_{s \to \infty} \alpha^{-}(s) = 0$, and thus a (Φ, β)-WPI with $\beta := \alpha^{-}$ holds.

The other listed properties are standard for generalized inverse (monotone) functions [14, Proposition 1, (3) and (4)], using that $\alpha^{-} \circ \alpha(r) = \alpha^{-}(-\mathsf{T}(r)) = \mathsf{T}^{-} \circ \mathsf{T}(r)$ and noting that here $\alpha^{-} \leqslant \mathfrak{a} < \infty$.

The second statement b) follows along the same lines.

For statement c) we use that from [14, Proposition 1, (5)], $\alpha(r) \geqslant s \iff r \leqslant \alpha^{-}(s)$, therefore

$$\begin{aligned}(\alpha^{-})^{-}(r) &= \inf\{s > 0 : \alpha^{-}(s) \leqslant r\} \\ &= \inf\{s > 0 : \alpha^{-}(s) < r\} \\ &= \inf\{s > 0 : \alpha(r) \leqslant s\} \\ &= \alpha(s).\end{aligned}$$

The proof for β is identical. □

Definition 6. In the situation where a (Φ, α)-WPI (resp. (Φ, β)-WPI) holds for α (resp. β) right continuous, we refer to it as a (Φ, α, β)-WPI where $\beta = \alpha^{-}$ (resp. $\alpha = \beta^{-}$).

The main interest of WPIs is summarized below:

Theorem 7 (Theorem 8 [1])**.** *Let P be a $\mu-$invariant Markov kernel on $\left(\mathsf{E},\mathscr{E}\right)$ and assume that $T := P^{*}P$ satisfies a $(\Phi,\beta)-$WPI for a sieve Φ. Then for $f \in \mathrm{L}_{0}^{2}(\mu)$ such that $0<\Phi(f)<\infty$ and any $n\in\mathbb{N}$, it holds that*

$$\|P^{n}f\|_{2}^{2} \le \gamma(n)\,\Phi(f)\,, \tag{3}$$

where $\gamma(n) := F_{\mathfrak{a}}^{-1}(n)$, where $F_{\mathfrak{a}}\colon (0,\mathfrak{a}] \to \mathbb{R}$ is the decreasing convex and invertible function

$$F_{\mathfrak{a}}(x) := \int_{x}^{\mathfrak{a}} \frac{\mathrm{d}v}{K^{*}(v)},$$

with $K^{*}\colon [0,\infty)\to[0,\infty]$ defined as $K^{*}(v) := \sup_{u\geq 0}\{uv - K(u)\}$, the convex conjugate of $K\colon[0,\infty)\to[0,\infty)$ given by $K(u) := u\,\beta(1/u)$ for $u>0$ and $K(0) := 0$.

The function γ satisfies $\gamma(n)\to 0$ as $n\to\infty$.

Remark 8. In practice, the precise value of $\mathfrak{a}$ as given in Definition 1 may not be known, however an upper bound $a \geq \mathfrak{a}$ is typically known, as in Example 2. The conclusions of Theorem 7 remain true when we consider $F_{a} := \int_{\cdot}^{a} \mathrm{d}v/K^{*}(v) = F_{\mathfrak{a}} + c$ for $c = \int_{\mathfrak{a}}^{a} \mathrm{d}v/K^{*}(v) \geq 0$, and we obtain the convergence bound in (3) with $\gamma = \gamma(\cdot;a) := F_{a}^{-1} = F_{\mathfrak{a}}^{-1}(\cdot - c) \geq F_{\mathfrak{a}}^{-1}$.

Remark 9. Our proof of this theorem actually supplies a collection of bounds on $\|P^{n}f\|_{2}^{2}$ which trade off tightness for tractability. In particular, writing $v_{n} = \|P^{n}f\|_{2}^{2}/\Phi(f)$, one can deduce (in decreasing order of tightness) the bounds

$$\begin{aligned}
\text{for all } n \geqslant 1, \quad & v_{n} \leqslant v_{n-1} - K^{*}(v_{n-1}) \\
\Longrightarrow \quad & v_{n} \leqslant \left(\mathrm{Id} - K^{*}\right)^{\circ n}(v_{0}) \\
\Longrightarrow \quad & \|P^{n}f\|_{2}^{2} \leqslant \Phi(f)\cdot\left(\mathrm{Id} - K^{*}\right)^{\circ n}\left(\frac{\|f\|_{2}^{2}}{\Phi(f)}\right)
\end{aligned}$$

and

$$\begin{aligned}
\text{for all } n \geqslant 1, \quad & F_{\mathfrak{a}}(v_{n}) - F_{\mathfrak{a}}(v_{n-1}) \geqslant 1 \\
\Longrightarrow \quad & F_{\mathfrak{a}}(v_{n}) \geqslant n + F_{\mathfrak{a}}(v_{0}) \\
\Longrightarrow \quad & \|P^{n}f\|_{2}^{2} \leqslant \Phi(f)\cdot F_{\mathfrak{a}}^{-1}\left(n + F_{\mathfrak{a}}\left(\frac{\|f\|_{2}^{2}}{\Phi(f)}\right)\right).
\end{aligned}$$

Each of these forms will be useful in deducing converse results, i.e. converting rates of convergence into WPIs.

Remark 10. Given only a WPI for P, one can deduce variance dissipation for the continuous-time semigroup obtained by Poissonizing P, i.e. let $P_{t} = \exp(t\mathcal{L})$ with $\mathcal{L} = P - I$, then

$$\|P_{t}f\|^{2} \leqslant \Phi(f)\cdot\gamma(2t)\,.$$

Definition 11. A $\mu-$invariant Markov kernel P satisfying (3) with $\gamma \downarrow 0$ as $n \to \infty$ is said to be (Φ, γ) -convergent. If the specific rate γ is not important, we may say that P is Φ-convergent.

The Dirichlet form $\mathcal{E}(P^* P, f)$ may not be tractable or straightforward to work with. In the reversible scenario, it is possible to deduce a (Φ, β) $-$WPI for $\mathcal{E}\left(P^2, f\right)$ from simpler Dirichlet forms or properties of P.

Theorem 12 ([1], Theorem 21 and Theorem 42)**.** *Let P be a $\mu-$invariant Markov kernel on $(\mathsf{E}, \mathscr{E})$ and assume that P satisfies a $\left(\tilde{\Phi}, \beta_+\right)-$WPI for a sieve $\tilde{\Phi}$. Then,*

a). if, in addition, P is μ-reversible and $(-P)$ satisfies a $\left(\tilde{\Phi}, \beta_-\right)$ $-$WPI, we have that *P^2 satisfies a (Φ, β)* $-$WPI with, for $s > 0$ and $f \in \mathrm{L}_0^2(\mu)$,

$$\beta(s) := \inf\left\{s_1 \beta_+(s_2) + \beta_-(s_1) \,|\, s_1 > 0, s_2 > 0, s_1 s_2 = s\right\},$$
$$\Phi(f) := \tilde{\Phi}(f) \vee \tilde{\Phi}\left((\mathrm{Id} + P)^{1/2} f\right);$$

b). if for any $(x, A) \in \mathsf{E} \times \mathscr{E}$ we have $P(x, A) \geqslant \varepsilon(x) \cdot \int_A \delta_x(\mathrm{d}y)$ for some $\varepsilon \colon \mathsf{E} \to [0,1]$, we have that P^2 satisfies a (Φ, β) $-$WPI with, for $s > 0$ and $f \in \mathrm{L}_0^2(\mu)$,

$$\beta(s) := \inf\left\{s_1 \beta_-(s_2) + \beta_+(s_1) \,|\, s_1 > 0, s_2 > 0, s_1 s_2 = s\right\},$$
$$\Phi(f) := \tilde{\Phi}(f) \vee \|f\|_{\mathrm{osc}}^2,$$

where here $\beta_-(s) := \frac{1}{2}\mu\left(\varepsilon(X)^{-1} \geqslant s\right)$.

For practical purposes it may be useful to note that $K^* = K_+^* \circ K_-^*$ and $K^* = K_-^* \circ K_+^*$ in the respective cases above, with $K_\pm^*$ defined as in Theorem 7, but for $\beta_\pm$.

2.2 $(\|\cdot\|_p^2, \gamma_p)-$convergence from $(\|\cdot\|_{\mathrm{osc}}^2, \gamma)-$convergence

In practice it can sometime be difficult to establish that a candidate sieve Φ, found through calculations, is indeed a sieve. In contrast the cases $\Phi = \|\cdot\|_\infty^2$ or $\Phi = \|\cdot\|_{\mathrm{osc}}^2$ can simplify calculations greatly. This appears at first sight to be at the expense of generality in terms of the class of functions for which convergence can be established. The following, which follows directly from [8, Lemma 5.1], shows that $(\|\cdot\|_{\mathrm{osc}}^2, \gamma)$-convergence automatically implies $(\|\cdot\|_p^2, \gamma_p)$-convergence. (We note that the result of [8, Lemma 5.1] is even more general, but this full generality is not needed here.) We will make use of this result throughout this manuscript in order to simplify presentation. An alternative strategy to handle broader classes of functions is suggested in [1, Proposition 37, Theorems 38, 42], where $\left(\|\cdot\|_p^2, \beta_p := \beta^{1-1/p}\right)$ -WPIs for $p \in [2, \infty]$ are considered directly. We do not know whether either of these two approaches is suboptimal in general but have observed that one recovers similar rates in the polynomial scenario. We note however that we have found the approach given in [1] more difficult to use in practice. We provide a proof of the result of [8, Lemma 5.1] in Appendix A for the reader's convenience.

Proposition 13. *Let P be a μ-invariant Markov kernel, assumed to be $\left(\|\cdot\|_{\mathrm{osc}}^2, \gamma\right)$-convergent. Then P is also $\left(\|\cdot\|_p^2, \gamma_p\right)$-convergent for $p > 2$, with*

$$\gamma_p(n) \leqslant 2^{4+4/p} \left[\gamma(n)\right]^{1-\frac{2}{p}}, \quad n \in \mathbb{N}.$$

Since the bound for γ_2 is not decreasing, the above result does not provide an L^2 convergence rate for all L^2 functions. However, as mentioned in [38], we can deduce uniform L^1 convergence for all L^2 functions from uniform L^2 convergence for all bounded functions.

Proposition 14. *The following are equivalent:*

$$\lim_{n\to\infty} \sup_{f:\mu(f^2)\leq 1} \|P^n f - \mu(f)\|_1 = 0, \tag{4}$$

and

$$\lim_{n\to\infty} \sup_{f:\|f\|_\infty\leq 1} \|P^n f - \mu(f)\|_2 = 0. \tag{5}$$

Proof. We start with (4)⇒(5). So consider f with $\|f\|_\infty \leq 1$.

$$\begin{aligned}\|P^n f - \mu(f)\|_2^2 &= \int |P^n f - \mu(f)| \cdot |P^n f - \mu(f)| \, \mathrm{d}\mu \\ &\leq 2 \int |P^n f - \mu(f)| \, \mathrm{d}\mu,\end{aligned}$$

and this final expression converges uniformly over f to 0 by (4), since $\{f : \|f\|_\infty \leq 1\} \subset \{f : \mu(f^2) \leq 1\}$. We now consider the converse, (5)⇒(4). Without loss of generality we may consider $f \in \mathcal{F} = \{g \in \mathrm{L}_0^2(\mu) : \|g\|_2 \leq 1\}$. Let $\epsilon > 0$ be arbitrary; we will show that for n large enough, $\sup_{f:\|f\|_2\leq 1} \int |P^n f| \, \mathrm{d}\mu \leq \epsilon$. Take $K = 4/\epsilon$ and N large enough such that

$$\sup_{g:\|g\|_\infty \leq K} \|P^N(g) - \mu(g)\|_2 \leq \frac{\epsilon}{2},$$

which is valid due to (5). Decomposing an arbitrary $f \in \mathcal{F}$ as $f = f \cdot \mathbf{1}_A + f \cdot \mathbf{1}_{A^\complement}$ for $A \in \mathscr{E}$, we have

$$\int \left|P^N f\right| \, \mathrm{d}\mu \leq \int \left|P^N (f \cdot \mathbf{1}_A)\right| \, \mathrm{d}\mu + \int \left|P^N (f \cdot \mathbf{1}_{A^\complement})\right| \, \mathrm{d}\mu,$$

by Minkowski's inequality. Now by Jensen's inequality, μ-invariance of P^N, and Cauchy–Schwarz,

$$\int \left|P^N (f \cdot \mathbf{1}_{A^\complement})\right| \, \mathrm{d}\mu \leq \int |f \cdot \mathbf{1}_{A^\complement}| \, \mathrm{d}\mu \leq \|f\|_2 \, \mu(A^\complement)^{1/2} \leq \mu(A^\complement)^{1/2}.$$

Take $A = \{x \in \mathsf{E} : |f(x)| \leq K\}$, and we obtain by Markov's inequality

$$\mu(A^{\complement}) = \mu(\mathbf{1}_{|f|^2 > K^2}) \leq \frac{1}{K^2}.$$

From

$$|\mu(f \cdot \mathbf{1}_{A^{\complement}})| \leq \mu\left(|f \cdot \mathbf{1}_{A^{\complement}}|\right) \leq 1/K,$$

and $\mu(f) = 0$ we also obtain $|\mu(f \cdot \mathbf{1}_A)| \leq 1/K$. Finally, we deduce that

$$\begin{aligned}
\int \left|P^N f\right| \mathrm{d}\mu &\leq \int \left|P^N \left(f \cdot \mathbf{1}_A\right)\right| \mathrm{d}\mu + \int \left|P^N \left(f \cdot \mathbf{1}_{A^{\complement}}\right)\right| \mathrm{d}\mu. \\
&\leq \int \left|P^N \left(f \cdot \mathbf{1}_A\right) - \mu(f \cdot \mathbf{1}_A)\right| \mathrm{d}\mu + |\mu(f \cdot \mathbf{1}_A)| + \frac{1}{K} \\
&\leq \frac{\epsilon}{2} + \frac{2}{K} \\
&\leq \epsilon.
\end{aligned}$$

Since $f \in \mathcal{F}$ was arbitrary, the result follows. □

2.3 Deducing WPIs from subgeometric rates of convergence

Given a quantitative estimate of the convergence of $\|P^n f\|_2^2$, it is possible to deduce a quantitative WPI for $\mathcal{E}\left(P^* P, f\right)$.

Proposition 15 ([1, Proposition 24; see also Remark 25])**.** *Let P be a $\mu-$invariant Markov kernel on $\left(\mathsf{E}, \mathscr{E}\right)$, and let Φ be a sieve.*

a). Suppose that for some K^* nonnegative, increasing, convex, and satisfying $K^*(0) = 0$, there holds for all $f \in \mathrm{L}_0^2(\mu)$ such that $0 < \Phi(f) < \infty$ and for all $n \geqslant 0$ an estimate of the form $\|P^n f\|_2^2 \leqslant \Phi(f) \cdot \left(\mathrm{Id} - K^*\right)^{\circ n}\left(\frac{\|f\|_2^2}{\Phi(f)}\right)$. It then follows that $\mathcal{E}\left(P^* P, f\right) \geqslant \Phi(f) \cdot K^*\left(\frac{\|f\|_2^2}{\Phi(f)}\right)$

b). Suppose that for a function $F : \mathbb{R}_+ \to (0, \infty)$ which is decreasing, continuous, divergent at 0, with an inverse function F^{-1} which is decreasing, continuous, and convex, and such that $\log\left(-\mathrm{D}F^{-1}\right)$ is convex, there holds for all $f \in \mathrm{L}_0^2(\mu)$ such that $0 < \Phi(f) < \infty$ and for all $n \geqslant 0$ an estimate of the form $\|P^n f\|_2^2 \leqslant \Phi(f) \cdot F^{-1}\left(n + F\left(\frac{\|f\|_2^2}{\Phi(f)}\right)\right)$. It then follows that $\mathcal{E}\left(P^* P, f\right) \geqslant \Phi(f) \cdot K^*\left(\frac{\|f\|_2^2}{\Phi(f)}\right)$, where $K^* = \mathrm{Id} - F^{-1}\left(1 + F(\cdot)\right)$ is nonnegative, increasing, convex, and satisfies $K^*(0) = 0$.

c). Suppose that for a function $\gamma : \mathbb{R}_+ \to (0, \infty)$ which is decreasing and has limit 0 at ∞, there holds for all $f \in \mathrm{L}_0^2(\mu)$ such that $0 < \Phi(f) < \infty$ and for all $n \geqslant 0$ an estimate of the form $\|P^n f\|_2^2 \leqslant \Phi(f) \cdot \gamma(n)$. Suppose also that P is μ-reversible. It then follows that $\mathcal{E}\left(P^* P, f\right) \geqslant \Phi(f) \cdot K^*\left(\frac{\|f\|_2^2}{\Phi(f)}\right)$, for some K^* which is nonnegative, increasing, convex, and satisfies $K^*(0) = 0$.

Remark 16. Note that for reversible kernels P, it holds for all $f \in \mathrm{L}_0^2(\mu)$ that the sequence $\gamma_f : n \mapsto \|P^n f\|_2^2$ is decreasing, continuous, convex, and that $\log(-\mathrm{D}\gamma_f)$ is convex, and hence that the assumption in Part 2 of the above Proposition holds.

2.4 Bounds on the Asymptotic Variance

A by-product of the WPI analysis is that the asymptotic variance of ergodic averages of the Markov chain in question can be upper-bounded for suitable functions.

Theorem 17. *Let P be a $\mu-$reversible Markov kernel on $(\mathsf{E}, \mathscr{E})$ and let Φ be a sieve such that for all $f \in \mathrm{L}_0^2(\mu)$ such that $0 < \Phi(f) < \infty$, the optimized WPI holds:*

$$\frac{\mathcal{E}(P^* P, f)}{\Phi(f)} \geqslant K^* \left(\frac{\|f\|_2^2}{\Phi(f)} \right).$$

Assume also that the map $v \mapsto v - K^*(v)$ is increasing on $(0, \mathfrak{a}]$. Define $B(v) = \int_0^v \frac{w}{K^*(w)} \, \mathrm{d}w$, which is assumed to be finite for $v \in [0, \mathfrak{a}]$. Then the asymptotic variance of f can be bounded as

$$\operatorname{var}(P, f) \leqslant 4 \cdot \Phi(f) \cdot B\left(\frac{\|f\|_2^2}{\Phi(f)} \right).$$

Proof. Using reversibility of the kernel, we write the asymptotic variance of f as

$$\operatorname{var}(P, f) = \int_{-1}^{1} \nu_f(\mathrm{d}\lambda) \cdot \frac{1+\lambda}{1-\lambda}.$$

Bounding $\frac{1+\lambda}{1-\lambda} = \frac{(1+\lambda)^2}{1-\lambda^2} \leqslant 4 \cdot \frac{1}{1-\lambda^2}$, we can thus bound

$$\begin{aligned} \operatorname{var}(P, f) &\leqslant 4 \cdot \int_{-1}^{1} \nu_f(\mathrm{d}\lambda) \cdot \frac{1}{1-\lambda^2} \\ &= 4 \cdot \sum_{n \geqslant 0} \|P^n f\|_2^2. \end{aligned}$$

Recall now our tightest discrete-time bound on the variance of the semigroup, with $S := \mathrm{Id} - K^*$,

$$\|P^n f\|_2^2 \leqslant \Phi(f) \cdot S^{\circ n} \left(\frac{\|f\|_2^2}{\Phi(f)} \right),$$

we write $v = \frac{\|f\|_2^2}{\Phi(f)}$ and bound the asymptotic variance as

$$
\begin{aligned}
\operatorname{var}(P, f) &\leqslant 4 \cdot \Phi(f) \cdot \sum_{n \geqslant 0} S^{\circ n}(v) \\
&=: 4 \cdot \Phi(f) \cdot \tilde{B}(v) .
\end{aligned}
$$

We now control the growth of $\tilde{B}$. Noting that S is nonnegative, increasing, and concave, a simple induction argument proves that $S^{\circ n}$ also has these properties, and since $\tilde{B}$ is a nonnegative combination of these functions, it too has these properties.

Now, isolating the first term in the sum which defines $\tilde{B}$, we have the recursion $\tilde{B}(v) = v + \tilde{B}(S(v))$, which allows us to write

$$
\begin{aligned}
v &= \tilde{B}(v) - \tilde{B}(S(v)) \\
&= \int_{S(v)}^{v} \tilde{B}'(w) \, \mathrm{d}w.
\end{aligned}
$$

By concavity, it holds that for $w \in [S(v), v]$, $\tilde{B}'(w) \geqslant \tilde{B}'(v)$, whence

$$
\begin{aligned}
&& v &\geqslant (v - S(v)) \cdot \tilde{B}'(v) \\
&& &= K^*(v) \cdot \tilde{B}'(v) \\
\implies && \tilde{B}'(v) &\leqslant \frac{v}{K^*(v)}.
\end{aligned}
$$

Now, arguing that $\tilde{B}(0) = 0$ and integrating, we obtain the expression

$$
\tilde{B}(v) \leq B(v) := \int_0^v \frac{w}{K^*(w)} \, \mathrm{d}w
$$

from which the result follows. □

Remark 18. An analogous result can be shown for a continuous-time Markov process $\{P_t : t \geqslant 0\}$, by defining the Dirichlet form in terms of the infinitesimal generator.

Remark 19. It is plausible that the assumption that $v \mapsto v - K^*(v)$ is increasing might follow from the defining properties of K and/or β, but we have been unable to establish this directly. In all of our explicit examples, this condition holds.

2.5 Towards spectral interpretations

In the reversible scenario, spectral representations of the operator P can provide useful insights. Subgeometric convergence naturally implies that the spectral radius of P is one and therefore that the spectrum accumulates at -1 or 1. The following are attempts to make these ideas more concrete.

2.5.1 Concentration of the spectrum

When P is reversible, we can utilize the spectral projection-valued measure representation of P. Thus for a given $f \in \mathrm{L}_0^2(\mu)$, let $\nu_f(\mathrm{d}\lambda)$ be the positive measure on $\sigma(P)$ which satisfies

$$\langle P^n f, f\rangle = \int_{\sigma(P)} \lambda^n \, \nu_f(\mathrm{d}\lambda).$$

Note that ν_f is a probability measure precisely when $\|f\|_2 = 1$. From our $(\Phi, \beta)-$WPI, we can conclude $(\Phi, \gamma)-$ convergence of $\|P^n f\|_2^2$ for some $\gamma : \mathbb{N}_0 \to \mathbb{R}$ with $\gamma(n) \downarrow 0$ as $n \to \infty$. This gives some control on the moments of ν_f: for any $f \in \mathrm{L}_0^2(\mu)$ with $\|f\|_2 = 1$,

$$\|P^n f\|_2^2 = \int_{\sigma(P)} \lambda^{2n} \, \nu_f(\mathrm{d}\lambda) \leq \Phi(f)\, \gamma(n). \tag{6}$$

In particular, we have

$$\sup_{f:\|f\|_2=1} \left\{ \frac{\int_{\sigma(P)} \lambda^{2n} \, \nu_f(\mathrm{d}\lambda)}{\Phi(f)} \right\} \leq \gamma(n),$$

from which we may deduce by Markov's inequality

$$\sup_{f:\|f\|_2=1} \left\{ \frac{\mathbb{P}_{\nu_f}\left(\lambda^2 > \exp(-\delta)\right)}{\Phi(f)} \right\} \leq \inf_{n\geq 1} \left\{ \frac{\gamma(n)}{\exp(-\delta n)} \right\}.$$

For example, if $\gamma(n) \leq cn^{-k}$, then there exists C such that

$$\sup_{f:\|f\|_2=1} \left\{ \frac{\mathbb{P}_{\nu_f}\left(\lambda^2 > \exp(-\delta)\right)}{\Phi(f)} \right\} \leq C\delta^k.$$

This may be viewed as the subgeometric counterpart to the fact that if $\gamma(n) = \rho^n$ then this implies by the same reasoning that $\mathbb{P}_{\nu_f}(\lambda^2 > \rho) = 0$ for all f with $\Phi(f) < \infty$ and $\|f\|_2 = 1$.

2.5.2 Spectrum of the Independent Metropolis–Hastings algorithm

Consider the Independent Metropolis–Hastings (IMH), also known as an *independence sampler*, on a countable state space $\mathsf{E} = \mathbb{N}_0$. For a fixed target distribution π and proposal distribution q on E, at position $X_n = x$, the chain proposes a move to $Y \sim q$, and conditional on $Y = y$, accepts this move with probability $1 \wedge \frac{\pi(y)q(x)}{\pi(x)q(y)}$ and sets $X_{n+1} = y$, otherwise the move is rejected and $X_{n+1} = x$. For brevity, we define

$$w(x) := \frac{\pi(x)}{q(x)}, \quad x \in \mathsf{E}.$$

For the IMH, the spectrum of the transition kernel P has been characterized in [16]:

$$\sigma(P) = \{\mathsf{r}_w : w \in \mathcal{W}\} \cup \{1\},$$

where $\mathcal{W} = \{w(x) : x \in \mathsf{E}\}$, $\mathsf{r}_w := \mathbb{P}(X_1 = x \,|\, X_0 = x, w(x) = w)$ are the rejection probabilities.

In order to be concrete, we consider a specific choice of π, q: we take geometric $\pi(x) = (1-a) \cdot a^x$ and $q(x) = (1-b) \cdot b^x$ for $x \in \mathsf{E} = \mathbb{N}_0$, where $0 < b < a < 1$. In this case, the Markov chain will converge subgeometrically, with rate $n^{-\frac{b}{a-b}}$ for bounded functions (this can be seen by a straightforward adaptation of the example in [1, Section 2.3.1]). In this countable state space setting, it is furthermore possible to explicitly characterize the spectrum [16]. By computing explicitly the rejection probabilities r_w, we find that

$$\sigma(P) = \left\{\Lambda_m := 1 - \frac{1-b}{1-a} \cdot \left(\frac{b}{a}\right)^m + \frac{a-b}{1-a} \cdot b^m : m \in \mathbb{N}_0\right\} \cup \{1\}. \tag{7}$$

Since $\Lambda_m \uparrow 1$ as $m \to \infty$, we see there is no spectral gap, and indeed choosing a smaller value of b – which leads to a slower rate of convergence for bounded functions – causes the spectrum to concentrate even more tightly around 1.

Given a test function $f \in \mathrm{L}_0^2(\pi)$ with $\|f\|_2 = 1$, we can consider its spectral measure $\nu_f(\cdot)$ on $\sigma(P)$, which has the property that $\langle P^n f, f\rangle = \int_{\sigma(P)} \lambda^n \, \nu_f(\mathrm{d}\lambda)$ for all $n \in \mathbb{N}_0$. Since f has unit norm, ν_f is a probability mass function supported on $\{\Lambda_m : m \in \mathbb{N}_0\}$. The function f is thus entirely characterized by the measure ν_f, and many of its properties can be read off from this.

For example, if

$$\int_{\sigma(P)} (1-\lambda)^{-1} \, \nu_f(\mathrm{d}\lambda) = \sum_{m \in \mathbb{N}_0} (1-\Lambda_m)^{-1} \, \nu_f(\Lambda_m) < \infty, \tag{8}$$

then f will have a finite asymptotic variance. Given our expression for the Λ_m (7), we see this will be the case when the masses $\nu_f(\Lambda_m)$ decay strictly faster than $(a/b)^m$, to ensure the sum in (8) is finite.

3 Optimal choices of α, β, Φ and ordering

Given our formulation of a WPI in Definition 3, it is natural to ask how one might optimize the constituent components: that is, how to make formal the notion of a "best" possible α, β or Φ.

3.1 Optimal α and β

We start by fixing a given sieve Φ, and seeking an optimal α and β. We assume that Φ is such that there exist functions f such that $0 < \Phi(f) < \infty$. Since $\mathrm{var}_\mu(f) \le \mathfrak{a}\Phi(f)$, $\Phi(f) = 0 \Rightarrow \mathrm{var}_\mu(f) = 0$ and so this assumption means only that we avoid the scenario where the only functions such that $\Phi(f) < \infty$ are constant functions.

We define minimal α and β functions, for a given sieve Φ, as the (pointwise) minimal functions satisfying Definition 3.

Definition 20. For a $\mu-$invariant Markov kernel T and sieve Φ define,

a). for any $r > 0$,

$$\alpha^{\star}(r;\Phi) := \sup\left\{\frac{\|g\|_2^2}{\mathcal{E}(T,g)}\left(1-\frac{r}{\|g\|_2^2}\right) : g \in \mathrm{L}_0^2(\mu), \Phi(g) = 1\right\} \vee 0,$$

noting that if $r \geq \mathfrak{a}$, $\alpha^{\star}(r;\Phi) = 0$;

b). for any $s > 0$,

$$\beta^{\star}(s;\Phi) := \sup\left\{\|g\|_2^2 - s\mathcal{E}(T,g) : g \in \mathrm{L}_0^2(\mu), \Phi(g) = 1\right\} \vee 0 \quad .$$

When $\Phi = \|\cdot\|_{\mathrm{osc}}^2$ we shall plainly write $\alpha^{\star}(\cdot) := \alpha^{\star}(\cdot;\Phi)$ and $\beta^{\star}(\cdot) := \beta^{\star}(\cdot;\Phi)$.

Despite their definitions it is not clear that the functions $\alpha^{\star}$ and $\beta^{\star}$ satisfy all the conditions required for a WPI to hold. The following theorem clarifies this point and also establishes that $\alpha^{\star}$ and $\beta^{\star}$ are inverses of each other when restricted to appropriate domains. The statement requires the existence of some (Φ,α)- or (Φ,β)-WPI, which we note can be established with the results of Subsection 4.2 for $\Phi = \|\cdot\|_{\mathrm{osc}}^2$. In particular Corollary 63 establishes that $\mu-$irreducibility is a sufficient condition for the existence of a WPI.

Theorem 21. *Suppose that the $\mu-$invariant kernel T possesses some (Φ,α)- or (Φ,β)-WPI. Then $\alpha^{\star}(\cdot;\Phi)$ defines a $(\Phi,\alpha^{\star})$-WPI and $\beta^{\star}(\cdot;\Phi)$ defines a $(\Phi,\beta^{\star})$-WPI. Furthermore, the functions $\alpha^{\star}(\cdot;\Phi) : (0,\mathfrak{a}] \to [0,\infty)$ and $\beta^{\star}(\cdot;\Phi) : [0,\infty) \to [0,\mathfrak{a}]$ are convex and continuous. In addition, $\beta^{\star}$ is strictly decreasing to 0 and $\alpha^{\star} = (\beta^{\star})^{-1}$ is the inverse function, which is well-defined on $(0,\mathfrak{a}]$ and strictly decreasing.*

Proof. We consider the β formulation, and drop explicit reference to the fixed Φ under consideration; the α formulation is analogous. By assumption, we know that T possesses a (Φ,β)-WPI, for some function β as in Definition 3 (*c.f.* Proposition 5). By definition of $\beta^{\star}$, we have that $0 \leq \beta^{\star} \leq \beta$ pointwise and so $\beta^{\star}(s) \to 0$ as $s \to \infty$. Since the pointwise supremum of affine functions (of s) is convex, we obtain convexity and continuity of $\beta^{\star}$, from the fact that it is the composition of a nondecreasing convex continuous function, $s \mapsto \max\{0,s\}$, with a convex function. We observe that $\beta^{\star}(0) = \mathfrak{a}$. Now, let $s_0 := \inf\{s > 0 : \beta^{\star}(s) = 0\}$, which may be infinite. Since $\beta^{\star}$ is convex and continuous, it is strictly decreasing on $(0,s_0)$. It follows that $\beta^{\star}$ is invertible on $(0,s_0)$ with inverse $(\beta^{\star})^{-1} : (0,\mathfrak{a}] \to [0,\infty)$ that is also convex and strictly decreasing.

Now we show that $\alpha^{\star} = (\beta^{\star})^{-1}$. For $r \in (0,\mathfrak{a}]$, let $s := (\beta^{\star})^{-1}(r)$. For any $f \in \mathrm{L}_0^2(\mu)$ with $\Phi(f) = 1$ we have

$$\|f\|_2^2 - s\mathcal{E}(T,f) \leq \beta^{\star}(s) = r,$$

and this implies

$$\alpha^\star(r) = \sup_{f:\Phi(f)=1} \frac{\|f\|_2^2}{\mathcal{E}(T,f)} - \frac{r}{\mathcal{E}(T,f)} \leq s.$$

Assume for the sake of contradiction that $\alpha^\star(r) = t < s$. For any $f \in \mathrm{L}_0^2(\mu)$ with $\Phi(f) = 1$ we have

$$\frac{\|f\|_2^2}{\mathcal{E}(T,f)} - \frac{r}{\mathcal{E}(T,f)} \leq t,$$

and so

$$\beta^\star(t) = \sup_{f:\Phi(f)=1} \|f\|_2^2 - t\mathcal{E}(T,f) \leq r = \beta^\star(s),$$

which is a contradiction since $\beta^\star$ is decreasing, and we conclude. $\square$

Remark 22. The function $\alpha^\star$ may be upper and lower bounded using the function $\psi : \mathbb{R}_+ \to [0,\infty)$,

$$\psi(t;\Phi) := \inf_{f:\Phi(f)=1, \|f\|_2^2 > t} \frac{\mathcal{E}(T,f)}{\|f\|_2^2},$$

which is nondecreasing. The behaviour of $\psi(\cdot;\Phi)$ as t decreases to 0 gives bounds on $\alpha^\star(\cdot;\Phi)$. Indeed, we find that for any $t > r$,

$$\frac{1}{\psi(t;\Phi)}\left(1 - \frac{r}{t}\right) \leq \alpha^\star(r;\Phi) \leq \frac{1}{\psi(r;\Phi)}.$$

Taking $t = 2r$ we obtain

$$\frac{1}{2\psi(2r;\Phi)} \leq \alpha^\star(r;\Phi) \leq \frac{1}{\psi(r;\Phi)},$$

and we may also deduce that $\lim_{r\downarrow 0} \alpha^\star(r;\Phi) = \psi(0;\Phi)^{-1}$. We see that $\alpha^\star$ is intimately connected to the rate at which ψ decreases as t decreases, i.e. as the variance of functions f with $\Phi(f) = 1$ is allowed to decrease to 0. We will see in Theorem 38 that, when $\Phi = \|\cdot\|_{\mathrm{osc}}^2$, upper and lower bounds may also be obtained by considering only indicator functions. One can also bound $\beta^\star$ in a similar manner using the function $\psi^-(u) := \sup\{t : \psi(t) \leq u\}$, in which case one finds

$$\frac{1}{2}\psi^-\left(\frac{1}{2s};\Phi\right) \leq \beta^\star(s;\Phi) \leq \psi^-\left(\frac{1}{s};\Phi\right).$$

In fact, if Φ defines a subspace $\mathcal{F}$ of $\mathrm{L}_0^2(\mu)$ then one may view $\psi(0;\Phi)$ as the right spectral gap associated with T as an operator on the closure of $\mathcal{F}$; see Lemma 116. In the case where $T = P^*P$ and $\psi(0;\Phi) > 0$ then this implies $\|P^n f\|_2^2 \leq \{1 - \psi(0;\Phi)\}^n \|f\|_2^2$ for functions $f \in \mathcal{F}$; see Remark 117. This is also natural by observing that if we define $\alpha^\star(0;\Phi) := \lim_{r\downarrow 0} \alpha^\star(r;\Phi) = \psi(0;\Phi)^{-1}$ we observe that a $(\Phi,\alpha^\star)$-WPI implies that $\|f\|_2^2 \leq \alpha^\star(0;\Phi)\mathcal{E}(T,f)$ for all $f \in \mathcal{F}$, from which the same bound on $\|P^n f\|_2^2$ may be directly obtained. Finally, when $\Phi = \|\cdot\|_{\mathrm{osc}}^2$ then $\psi(0;\Phi)$ is the $\mathrm{L}_0^2(\mu)$ spectral gap; see Lemma 116.

3.2 Lower bounds on convergence rates

In principle, noting that $\alpha^\star$ and $\beta^\star$ are pointwise minimal functions, any function $f \in \mathrm{L}^2_0(\mu)$ with $\Phi(f) = 1$ may be used to construct a lower bound. For example, for any such function, $\beta^\star$ satisfies

$$\beta^\star(s) \geq \|f\|_2^2 - s\mathcal{E}(T, f), \qquad s > 0.$$

In practice, to produce an informative lower bound for the whole function $\beta^\star$, one will need to identify an appropriate sequence of functions. Indicator functions of measurable sets are always in $\mathrm{L}^2_0(\mu)$, have finite oscillation, and they can provide a tractable source of such functions as $\mathcal{E}(P, \mathbf{1}_A)$ has a natural probabilistic interpretation. We show that such functions can provide both lower and upper bounds for $\beta^\star$ in Section 4.1.

We now show that a lower bound on β_1 in a (Φ, β_1)-WPI for P can imply a lower bound on β_2 in a (Φ, β_2)-WPI for P^*P.

Lemma 23 ([11, Remark 3.1])**.** *Let P be μ-invariant. Then*

$$\mathcal{E}(P^*P, f) \leq 2\mathcal{E}(P, f), \qquad f \in \mathrm{L}^2_0(\mu).$$

Remark 24. If P is μ-reversible, one can obtain $\mathcal{E}(P^2, f) \leq (1 + \lambda_\star)\mathcal{E}(P, f)$ by using the spectral theorem, where $\lambda_\star = \sup \sigma_0(P)$. However, since the focus here is on WPIs, the case $\lambda_\star < 1$ is less relevant.

We note that a converse may be obtained when P, and therefore P^*, satisfies $P(x, \{x\}) \geq \varepsilon$ on a μ-full set; see Lemma 49.

Lemma 25. *Let P be μ-invariant, and assume it satisfies a $(\Phi, \beta_1^\star)$-WPI, where $\beta_1^\star$ is pointwise minimal. Assume P^*P satisfies a $(\Phi, \beta_2^\star)$-WPI where $\beta_2^\star$ is pointwise minimal. Then $\beta_2^\star(s) \geq \beta_1^\star(2s)$.*

Proof. Let $\mathcal{F} = \{f \in \mathrm{L}^2_0(\mu) : \Phi(f) = 1\}$. By Lemma 23 we have $\mathcal{E}(P^*P, f) \leq 2\mathcal{E}(P, f)$. We may write

$$\beta_1^\star(s) = 0 \vee \sup_{f \in \mathcal{F}} \mathrm{var}_\mu(f) - s\mathcal{E}(P, f).$$

We then have

$$\begin{aligned}
\beta_2^\star(s) &= 0 \vee \sup_{f \in \mathcal{F}} \mathrm{var}_\mu(f) - s\mathcal{E}(P^*P, f) \\
&\geq 0 \vee \sup_{f \in \mathcal{F}} \mathrm{var}_\mu(f) - 2s\mathcal{E}(P, f) \\
&= \beta_1^\star(2s),
\end{aligned}$$

and we conclude. ◻

In the case where P is μ-reversible, we can then deduce from a (Φ, β_1)-WPI for P a lower bound on a separable rate of convergence for $\|P^n f\|$.

Proposition 26. *Assume P is μ-reversible, satisfies (2) and the pointwise minimal $\beta^\star$ satisfies $\beta^\star(s) \in \Omega(s^{-p})$ for some $p > 0$. Then it cannot hold that with $q > p$, $\|P^n f\|_2^2 \in \mathcal{O}(n^{-q})$ for all $f \in \mathrm{L}_0^2(\mu)$ with $\Phi(f) < \infty$.*

Proof. If $\beta^\star(s) \in \Omega(s^{-p})$ then we may deduce by Lemma 25 that if P^2 satisfies (2), its pointwise minimal $\beta_2^\star$ also satisfies $\beta_2^\star(s) \in \Omega(s^{-p})$. Now assume for the sake of contradiction that $\|P^n f\|_2^2 \in \mathcal{O}(n^{-q})$ for all $f \in \mathrm{L}_0^2(\mu)$ such that $\Phi(f) < \infty$. Then by [1, Proposition 24 and Remark 25], we deduce that a WPI for P^2 holds with $\beta_2(s) \in \mathcal{O}(s^{-q})$, which contradicts $\beta_2^\star(s) \in \Omega(s^{-p})$ being pointwise minimal. $\square$

The following result establishes a lower bound on $\beta^\star$ for Markov kernels P that can exhibit sticky behaviour in regions of the state space. [37, Theorem 5.1] showed that for a μ-invariant Markov kernel P with μ not concentrated at a single point, that $\mathrm{ess}_\mu \sup_x P(x, \{x\}) = 1$ implies that P cannot converge geometrically. In [26, Theorem 1] conductance is used to prove the same when P is μ-reversible, and the following provides a quantitative refinement.

Theorem 27. *Let P be $\mu-$reversible satisfying a (Φ, β)-WPI for $\Phi = \|\cdot\|_{\mathrm{osc}}^2$. For any $\varepsilon > 0$, define the set $A_\varepsilon := \{x \in \mathsf{X} \colon P(x, \{x\}) \geq 1 - \varepsilon\}$. Then for any $s > 0$,*

$$\beta(s) \geq \beta^\star(s) \geq \sup_{\varepsilon \in (0,1)} \{\mu(A_\varepsilon)(1 - s\varepsilon - \mu(A_\varepsilon))\}.$$

Proof. For any $A \subset \mathsf{X}$, from Lemma 118, we have $\mathcal{E}(P, \mathbf{1}_A) = \mu \otimes P(A \times A^\complement)$ and $\mathrm{var}(\mathbf{1}_A) = \mu \otimes \mu(A \times A^\complement)$. Since $\Phi(\mathbf{1}_{A_\varepsilon}) \leq 1$, for any $\varepsilon > 0$ we have

$$\begin{aligned}
\beta^\star(s) &:= \sup\left\{\|f\|_2^2 - s\mathcal{E}(P, f) \colon f \in \mathrm{L}_0^2(\mu), \Phi(f) \leq 1\right\} \\
&\geq \mathrm{var}_\mu(\mathbf{1}_{A_\varepsilon}) - s\int \mu(\mathrm{d}x)P(x, \mathrm{d}y)\mathbf{1}_{A_\varepsilon}(x)\mathbf{1}_{A_\varepsilon^\complement}(y) \\
&\geq \mathrm{var}_\mu(\mathbf{1}_{A_\varepsilon}) - s\int \mu(\mathrm{d}x)P(x, \{x\}^\complement)\mathbf{1}_{A_\varepsilon}(x) \\
&\geq \mu(A_\varepsilon)\mu(A_\varepsilon^\complement) - s\mu(A_\varepsilon)\varepsilon \\
&= \mu(A_\varepsilon)(1 - s\varepsilon - \mu(A_\varepsilon)).
\end{aligned}$$

Thus we conclude. $\square$

Example 28. Assume $C\varepsilon^\alpha \geq \mu(A_\varepsilon) \geq c\varepsilon^\alpha$ for some $\alpha, c, C > 0$ and for all $\varepsilon > 0$ sufficiently small. Then for $s > 0$, we seek to maximize $\zeta(\varepsilon) = \varepsilon^\alpha(1 - s\varepsilon - C\varepsilon^\alpha)$. One can check that

$$\zeta'(\varepsilon) = (1 + \alpha)\varepsilon^{\alpha - 1}\left[\frac{\alpha}{1 + \alpha} - s\varepsilon - C\varepsilon^\alpha\right],$$

and since $\mathbb{R}_+ \ni \varepsilon \mapsto s\varepsilon + c\varepsilon^\alpha$ is increasing, there is a unique ε^* such that $\zeta'(\varepsilon_*) = 0$, $\zeta(\varepsilon) > 0$ (resp. $\zeta(\varepsilon) < 0$) for $\varepsilon < \varepsilon_*$ (resp. $\varepsilon > \varepsilon_*$). Note that for $s \geq \alpha/(1 + \alpha)$, $\varepsilon_* \in (0, 1)$ and let

$$\varepsilon_0 := \frac{\alpha}{1 + \alpha}s^{-1},$$

from above. Then notice that $\zeta(\varepsilon_0) \leq 0$ and $\varepsilon_0' = \varepsilon_0 - Cs^{-1}\varepsilon_0$ is such that $\zeta'(\varepsilon_0^{'}) \geq 0$, implying $\varepsilon_0 - cs^{-1}\varepsilon_0 \leq \varepsilon_* \leq \varepsilon_0$ and we obtain the lower bound, for $s > 0$

$$\beta^\star(s) \geq \underline{\beta}^\star(s) := c\left(\frac{\alpha}{1+\alpha}\right)^\alpha s^{-\alpha}\left[\frac{1}{1+\alpha} - C\left(\frac{\alpha}{1+\alpha}\right)^\alpha s^{-\alpha}\right],$$

which is positive for s sufficiently large. Therefore, since from earlier results $\beta^\star \geq \underline{\beta}^\star$ implies $\underline{\gamma}^\star(n) \leq \gamma^\star(n)$ if $\mu(A_\varepsilon) \geq c\varepsilon^\alpha$ then the corresponding Markov chain cannot converge at a rate faster than the polynomial rate $\underline{\gamma}^\star(n) \propto n^{-\alpha}$.

Example 29. In the case of the Independent Metropolis-Hastings (IMH) we are interested in lower bounding the probability

$$\varpi(\varepsilon) := \pi\left(\int \pi(\mathrm{d}y)\min\left\{w^{-1}(X), w^{-1}(y)\right\} < \varepsilon\right).$$

Note that for any $x \in \mathsf{X}$ we have

$$\int \pi(\mathrm{d}y)\min\left\{w^{-1}(x), w^{-1}(y)\right\} \leq w^{-1}(x),$$

therefore, since for random variables $Z(\omega) \leq Z'(\omega)$ implies $\mathbb{P}(Z(\omega) < \varepsilon) \geq \mathbb{P}(Z'(\omega) < \varepsilon)$

$$\varpi(\varepsilon) \geq \pi\left(w^{-1}(X) < \varepsilon\right) = \pi\left(w(X) > \varepsilon^{-1}\right).$$

As a result for $s > 0$

$$\beta^\star(s) \geq \sup_{\varepsilon\in(0,1)}\left\{\pi\left(w(X) > \varepsilon^{-1}\right)\left(1 - s\varepsilon - \pi\big(w(X) > \varepsilon^{-1}\big)\right)\right\},$$

therefore implying a lower bound on the fastest rate of convergence possible.

3.3 Ordering of α's, β's and γ's and Peskun–Tierney ordering

Theorem 30. *Let P_1 and P_2 be $\mu-$invariant Markov kernels such that for a sieve Φ, $P_1^*P_1$ satisfies a $(\Phi, \alpha_1, \beta_1)$-WPI and $P_2^*P_2$ a $(\Phi, \alpha_2, \beta_2)$-WPI respectively. Then we have*

a). $\alpha_2(\cdot;\Phi) \geq \alpha_1(\cdot;\Phi)$ if and only if $\beta_2(\cdot;\Phi) \geq \beta_1(\cdot;\Phi)$;

b). $\beta_2(\cdot;\Phi) \geq \beta_1(\cdot;\Phi)$ implies $\gamma_2(\cdot;\Phi) \geq \gamma_1(\cdot;\Phi)$.

Proof. First statement: we drop Φ for notational simplicity. For the direction ($\Longrightarrow$) : for any $s > 0$ we have $\{r > 0\colon \alpha_2(r) \leq s\} \subset \{r > 0\colon \alpha_1(r) \leq s\}$ and hence $\beta_2 = \alpha_2^{\text{-}} \geq \alpha_1^{\text{-}} = \beta_1$; ($\Longleftarrow$) follows along the same lines. For the second statement: from their definitions, $K_1 \leq K_2$ and hence $K_1^* \geq K_2^*$. As a result, $F_{1,\mathfrak{a}} \leq F_{2,\mathfrak{a}}$ and consequently $\gamma_1 := F_{1,\mathfrak{a}}^{-1} \leq F_{2,\mathfrak{a}}^{-1} =: \gamma_2$. $\square$

We know from [42] that for P_1, P_2 $\mu-$reversible, then $\mathcal{E}(P_1, g) \geq \mathcal{E}(P_2, g)$ for any $g \in \mathrm{L}^2(\mu)$ implies $\mathrm{var}(P_1, f) \leq \mathrm{var}(P_2, f)$ for $f \in \mathrm{L}^2(\mu)$ and $\mathrm{Gap}_{\mathrm{R}}(P_1) \geq \mathrm{Gap}_{\mathrm{R}}(P_2)$, the latter being useful when $\mathrm{Gap}_{\mathrm{R}}(P_1) > 0$, and say P_1 and P_2 are positive, since this implies faster convergence to equilibrium in most scenarios of interest. The following generalizes the latter statement to the subgeometric setup – the statement on asymptotic the variances remains naturally true.

Theorem 31. *Let P_1, P_2 be $\mu-$invariant Markov kernels such that for a sieve Φ,*

a). $P_1^* P_1$ (resp. $P_2^* P_2$) satisfies a $(\Phi, \alpha_1, \beta_1)$- WPI (resp. a $(\Phi, \alpha_2, \beta_2)$–WPI),

b). $\mathcal{E}(P_1^* P_1, g) \geq \mathcal{E}(P_2^* P_2, g)$ for any $g \in \mathrm{L}^2(\mu)$ such that $\Phi(g) \leq 1$.

Then with $\alpha_i^\star(\cdot; \Phi)$ and $\beta_i^\star(\cdot; \Phi)$ for $i = 1, 2$ defined as in Definition 20, a $(\Phi, \alpha_i^\star, \beta_i^\star)$- WPI holds for $i = 1, 2$ and we have for the corresponding convergence rates $\gamma_1^\star \leq \gamma_2^\star$.

Proof. From the ordering of Dirichlet forms we have for any $g \in \mathrm{L}^2(\mu)$

$$\|g\|_2^2 - s\mathcal{E}(P_1^* P_1, g) \leq \|g\|_2^2 - s\mathcal{E}(P_2^* P_2, g),$$

from Definition 20 we deduce $\beta_1^\star(\cdot; \Phi) \leq \beta_2^\star(\cdot; \Phi)$ and from Theorem 30 we conclude $\gamma_1^\star \leq \gamma_2^\star$. □

3.4 Optimal Φ

On the other hand, we can fix a bounded β, say and seek the optimal class of functions defined by a sieve Φ for this β. As a starting point, we assume that some $(\Phi, \beta)-$WPI holds for $T = P^* P$:

$$\|f\|_2^2 \leq s\mathcal{E}(P^* P, f) + \beta(s)\Phi(f), \quad \forall s > 0, f \in \mathrm{L}_0^2(\mu),$$

for a given Φ. By Theorem 7, we obtain the convergence bound:

$$\|P^n f\|_2^2 \leq \Phi(f)\gamma(n),$$

for a function $\gamma : \mathbb{N}_0 \to \mathbb{R}_+$ which satisfies $\gamma(n) \to 0$ as $n \to \infty$.

We now seek the smallest sieve $\Phi_\beta^\star$ such that a $(\Phi_\beta^\star, \beta)-$WPI still holds.

Definition 32. We define for any $f \in \mathrm{L}_0^2(\mu)$,

$$\Phi_\beta^\star(f) := \sup_{n \in \mathbb{N}_0} \Phi_\beta(P^n f),$$

where

$$\Phi_\beta(f) := \sup_{s>0} \frac{\|f\|_2^2 - s\mathcal{E}(P^* P, f)}{\beta(s)} = \|f\|_2^2 \cdot \sup_{s>0} \frac{1 - s\delta(f)}{\beta(s)},$$

where $\delta(f) := \mathcal{E}(P^* P, f)/\|f\|_2^2$ and satisfies $0 < \delta(f) \leq 1$.

Lemma 33. *The functional $\Phi_\beta^\star$ is a nonexpansive sieve for P.*

Proof. Note that for any $f \in \mathrm{L}_0^2(\mu)$, and $\Phi_\beta(cf) = c^2\Phi_\beta(f)$, and furthermore $\Phi_\beta(f) \geq \|f\|_2^2/\beta(0)$, where $\beta(0) := \lim_{s\to 0}\beta(s)$, which exists and is finite and nonzero by monotonicity and boundedness of β. Thus Φ_β satisfies condition a) from Definition 1.

Now $\Phi_\beta^\star(f) \geq \Phi_\beta(f)$, and hence $\Phi_\beta^\star$ also satisfies condition a) from Definition 1. Finally, $\Phi_\beta^\star$ is nonexpansive for P by construction, and hence is a nonexpansive sieve. □

With this definition of $\Phi_\beta^\star$, it is clear that we have a $(\Phi_\beta^\star, \beta)-$WPI: for all $s > 0$, $f \in \mathrm{L}_0^2(\mu)$,

$$\|f\|_2^2 \leq s\mathcal{E}(P^*P, f) + \beta(s)\Phi_\beta^\star(f),$$

and so we can obtain the convergence bound

$$\|P^n f\|_2^2 \leq \Phi_\beta^\star(f)\gamma(n),$$

for the same γ, and by construction $\Phi_\beta^\star \leq \Phi$.

Example 34. When $\beta(s) = s^{-\alpha}$, we can calculate that

$$\Phi_\beta(f) = \|f\|_2^2 \frac{\alpha^\alpha}{(\alpha+1)^{\alpha+1}} \left[\delta(f)\right]^{-\alpha}.$$

Then we have

$$\begin{aligned}\delta(P^n f) &= \frac{\mathcal{E}(P^2, P^n f)}{\|P^n f\|_2^2} = \frac{\langle(\mathrm{Id} - P^2)P^n f, P^n f\rangle}{\|P^n f\|_2^2} \\ &= 1 - \frac{\int_{\sigma(P)} \lambda^{2n+2}\,\nu_f(\mathrm{d}\lambda)}{\int_{\sigma(P)} \lambda^{2n}\,\nu_f(\mathrm{d}\lambda)}.\end{aligned}$$

Thus the mapping $n \mapsto \int_{\sigma(P)} \lambda^{2n}\,\nu_f(\mathrm{d}\lambda)$ will dictate for a given $f \in \mathrm{L}_0^2(\mu)$ whether or not $\Phi_\beta^*(f)$ is finite or infinite. As a concrete example, consider the situation when $\sigma(P) = [0,1]$ and when $\nu_f(\mathrm{d}\lambda)$ has density proportional to $\lambda^{a-1}\mathrm{d}\lambda$ for some $a > 1$. Then the $2n$th moment is $\prod_{r=0}^{2n-1} \frac{a+r}{a+1+r} = \frac{a}{a+2n}$, and so the ratio is

$$\frac{\int_{\sigma(P)} \lambda^{2n+2}\,\nu_f(\mathrm{d}\lambda)}{\int_{\sigma(P)} \lambda^{2n}\,\nu_f(\mathrm{d}\lambda)} = \frac{a+2n}{a+2+2n}.$$

In particular, we find

$$1 - \frac{\int_{\sigma(P)} \lambda^{2n+2}\,\nu_f(\mathrm{d}\lambda)}{\int_{\sigma(P)} \lambda^{2n}\,\nu_f(\mathrm{d}\lambda)} = \frac{2}{a+2+2n}.$$

Thus we see that asymptotically, $\Phi_\beta\left(P^n f\right)$ must grow like $\|P^n f\|_2^2 \cdot n^\alpha$. This will diverge to infinity as $n \to \infty$ if n^α dominates the rate of convergence to 0 of

$\|P^n f\|_2^2$. So informally speaking, if we consider the set $\left\{f \in \mathrm{L}_0^2(\mu) : \Phi_\beta^\star(f) < \infty\right\}$, $\Phi_\beta^\star$ is in effect 'sieving out' functions $f \in \mathrm{L}_0^2(\mu)$ whose spectral measures $\nu_f(\mathrm{d}\lambda)$ place too much mass close to 1.

To be more explicit, by applying Chernoff's inequality to (6), we can conclude that for $f \in \mathrm{L}_0^2(\mu)$ with $\|f\|_2^2 = 1$, for any $\delta > 0$,

$$\int_{1-\delta}^{1} \lambda^2 \, \nu_f(\mathrm{d}\lambda) \leq C \cdot \Phi(f) \cdot \delta^\alpha,$$

for a constant $C > 0$ independent of f, thus demonstrating that ν_f cannot place mass in an arbitrary fashion in a neighbourhood of 1.

3.5 Duality

The preceding two sections suggest the following natural approach to deriving convergence bounds and then refining them:

a). Choose a class of functions we seek convergence bounds for, and the corresponding Φ. For example, we could consider the class of bounded functions and correspondingly take $\Phi = \|\cdot\|_{\mathrm{osc}}^2$. As argued in Section 2.2, this choice is in a sense canonical.

b). Given this function class and its Φ, derive an optimal $\beta^\star(\cdot;\Phi)$ for this class, as given in Definition 20.

c). Given this optimal $\beta^\star(\cdot;\Phi)$, find the optimal $\Phi^\star := \Phi^\star_{\beta^\star(\cdot;\Phi)}$, given in Definition 32.

This procedure in fact is optimal after a single iteration; recursing these steps does not lead to any improvement.

Proposition 35. *We have that*

$$\beta^\star(\cdot;\Phi^\star) = \beta^\star(\cdot;\Phi).$$

Proof. By definition,

$$\beta^\star(s;\Phi^\star) = \sup_{f \in \mathrm{L}_0^2(\mu), \Phi^\star(f) \leq 1} \left\{\|f\|_2^2 - s\mathcal{E}(P^*P, f)\right\}. \tag{9}$$

Firstly, note that since $\Phi^\star$ is optimal,

$$\Phi^\star(f) \leq \Phi(f), \quad \forall f \in \mathrm{L}_0^2(\mu).$$

Therefore,

$$\left\{f \in \mathrm{L}_0^2(\mu) : \Phi(f) \leq 1\right\} \subset \left\{f \in \mathrm{L}_0^2(\mu) : \Phi^\star(f) \leq 1\right\}.$$

Thus the supremum in the definition of $\beta^\star(s;\Phi^\star)$ (9) is over a larger class of functions than that of $\beta^\star(s;\Phi)$ in Definition 20. Therefore we can immediately conclude that

$$\beta^\star(s;\Phi^\star) \geq \beta^\star(s;\Phi), \quad \forall s > 0. \tag{10}$$

However, by definition, if $\Phi^\star(f) \leq 1$, we have that

$$\sup_{s>0, n\in\mathbb{N}_0} \frac{\|P^n f\|_2^2 - s\mathcal{E}(P^*P, P^n f)}{\beta^\star(s;\Phi)} \leq 1,$$

which in particular (taking $n = 0$) implies that for any $s > 0$,

$$\|f\|_2^2 - s\mathcal{E}(P^*P, f) \leq \beta^\star(s;\Phi).$$

Thus

$$\beta^\star(s;\Phi^\star) \leq \beta^\star(s;\Phi), \quad \forall s > 0,$$

which taken together with (10), establishes the result. $\square$

4 Establishing WPIs

4.1 Cheeger meets Poincaré

In this section we discuss the connections between weak Poincaré inequalities and methods based on the concept of conductance. In particular, we define the notion of weak conductance, which extends the traditional definition of conductance to the subgeometric setting. Similar ideas were proposed in [38, Sections 4, 5] in the (continuous time) diffusion setting, but our arguments differ significantly and are inspired by the discrete-time proofs of [25, 12]. We fix a μ-reversible Markov transition kernel P on our measure space $(\mathsf{E}, \mathscr{E})$.

Definition 36. For a μ-reversible kernel P, we define the *weak conductance* $\kappa : [0,\infty) \to [0,\infty]$ to be

$$\kappa(u) := \inf_{A\in\mathscr{E}: u<\mu\otimes\mu(A\times A^\complement)} \frac{\mathcal{E}(P, \mathbf{1}_A)}{\|\mathbf{1}_A - \mu(A)\|_2^2} = \inf_{A\in\mathscr{E}: u<\mu\otimes\mu(A\times A^\complement)} \frac{\mu\otimes P(A\times A^\complement)}{\mu\otimes\mu(A\times A^\complement)}.$$

The last inequality follows from Lemma 118 in the Appendix. Note that since for any $A \in \mathcal{E}$, $\mu\otimes\mu(A\times A^\complement) \leq 1/4$, by convention we have $\kappa(u) = \infty$ for $u \geq 1/4$.

The definition of (strong) conductance [25] is recovered by taking $u = 0$; $\kappa(0)$ is Cheeger's constant, which in the subgeometric case is 0.

Remark 37. Following [23] rather than [25], some authors use a slightly different definition of conductance:

$$\kappa_* := \inf_{A\in\mathscr{E}, \mu(A)\leq 1/2} \frac{\mu\otimes P\big(A\times A^\complement\big)}{\mu(A)},$$

which possesses a clear probabilistic interpretation. We note however that $\kappa_* \leq \kappa(0) \leq 2\kappa_*$, and the key quantity used to establish Cheeger's inequalities, and our generalization, relies on κ as in Definition 36.

There is some resemblance between the weak conductance κ and the s-conductance introduced by [29]. However, it is not straightforward to compare the two or the type of convergence results obtained; see, e.g., [2, Lemma 2.1].

Cheeger's inequality [25] obtains a lower bound on $\mathcal{E}(P,f)/\|f\|_2^2$ for all $f \in \mathrm{L}_0^2(\mu)$, $f \neq 0$, from a lower bound on this same quantity when restricted to functions $f = \mathbf{1}_A - \mu(A)$ for $A \in \mathscr{E}$ (namely, $\kappa(0)$). This leads to the following celebrated inequalities when $\kappa(0) > 0$:

$$\kappa^2(0)/8 \le \mathrm{Gap}_{\mathrm{R}}(P) \le \kappa(0). \tag{11}$$

We generalize this idea to the scenario where the quantity $\kappa(0)$ is zero, so there is no right-spectral gap. As we shall see, this generalization involves an upper and lower bound for the function α in (1).

This generalization will be particularly useful when we seek to establish the existence of WPIs from the abstract RUPI condition in Section 4.2.

Theorem 38. *Let P be a μ-reversible kernel and $\Phi = \|\cdot\|_{\mathrm{osc}}^2$.*

*Provided that $\kappa(u) > 0$ for all $u \in (0, 1/4)$, a (Φ, α)-*WPI *holds for P, with*

$$\alpha(r) := \frac{16}{\kappa^2(r/16)}, \quad r > 0.$$

Conversely, if a $(\|\cdot\|_{\mathrm{osc}}^2, \alpha)$-WPI holds for some $\alpha : (0,\infty) \to [0,\infty)$, we have the bound

$$\frac{1}{\alpha(r)} \le \inf_{u>1} \left\{ \kappa(ur) \frac{u}{u-1} \right\} \le 2\kappa(2r), \quad r > 0. \tag{12}$$

Remark 39. In the notation of Section 3, and in analogue with (11), we can succinctly express this theorem in terms of the optimal $\alpha^\star$ as:

$$\frac{\kappa^2(r/16)}{16} \le 1/\alpha^\star(r) \le \inf_{s>1} \left\{ \frac{s}{s-1} \kappa(sr) \right\} \le 2\kappa(2r), \quad r > 0.$$

From Theorem 30, inequality (12) implies that convergence to equilibrium cannot occur at a rate γ faster fast than that obtained with $\underline{\alpha}(r) = [2\kappa(2r)]^{-1}$.

The proof is a direct consequence of Propositions 40 and 41. We first show that the conductance always provides a lower bound for α if a $(\|\cdot\|_{\mathrm{osc}}^2, \alpha)$-WPI holds.

Proposition 40. *Let P be a μ-reversible kernel satisfying a $(\|\cdot\|_{\mathrm{osc}}^2, \alpha)$-WPI. We have the bound (12).*

Proof. Consider the function

$$f = \frac{\mathbf{1}_A - \mu(A)}{\sqrt{\mu(A)\mu(A^{\complement})}},$$

for a measurable set $A \in \mathscr{E}$ such that $1 > \mu(A) > 0$. By construction, $f \in \mathrm{L}^2_0(\mu)$ with $\|f\|_2^2 = 1$. Plugging this into the weak Poincaré inequality, we find that for any $r > 0$,

$$1 \leq \alpha(r) \frac{\mu \otimes P(A \times A^{\complement})}{\mu \otimes \mu(A \times A^{\complement})} + \frac{r}{\mu \otimes \mu(A \times A^{\complement})}.$$

Rearranging this, we obtain that for any $r > 0$,

$$\frac{1}{\alpha(r)} \left(1 - \frac{r}{\mu \otimes \mu(A \times A^{\complement})}\right) \leq \frac{\mu \otimes P(A \times A^{\complement})}{\mu \otimes \mu(A \times A^{\complement})}.$$

Now for any $s > r > 0$, we consider only $A \in \mathscr{E}$ such that $\mu \otimes \mu(A \times A^{\complement}) > s$, yielding

$$1/\alpha(r) \leq \left(1 - \frac{r}{s}\right)^{-1} \frac{\mu \otimes P(A \times A^{\complement})}{\mu \otimes \mu(A \times A^{\complement})}.$$

Therefore for $r > 0$ we have

$$1/\alpha(r) \leq \inf_{s>r} \frac{s}{s-r} \kappa(s) = \inf_{u>1} \frac{u}{u-1} \kappa(ru) \leq 2\kappa(2r).$$

where we have used the change of variable $u = s/r$ for the equality and taken $u = 2$ for the final inequality. □

We now prove the trickier converse: we show that the weak conductance gives rise to an α such that a $(\|\cdot\|^2_{\mathrm{osc}}, \alpha)$-WPI holds. We make use of the fundamental Lemma 119 of [25] which provides a bridge between Dirichlet forms of indicator functions and general functions and can be found in the appendix for the reader's convenience.

Proposition 41. *Let P be a μ-reversible kernel. Then provided $\kappa(u) > 0$ for all $u \in (0, 1/4)$, a $(\|\cdot\|^2_{\mathrm{osc}}, \alpha)-$WPI holds with*

$$\alpha(r) := \frac{16}{\kappa^2\left(r/16\right)}, \quad r > 0. \tag{13}$$

Proof. Let us fix $f \in \mathrm{L}^2_0(\mu)$ with $\|f\|_2^2 = 1$. Our goal is to show that α as defined in (13) gives rise to a valid weak Poincaré inequality for P with $\Phi = \|\cdot\|^2_{\mathrm{osc}}$; since we have fixed $\|f\|_2^2 = 1$ this amounts to showing that for $r > 0$,

$$1 \leq \frac{16}{\kappa^2\left(r/16\right)} \mathcal{E}(P, f) + r\|f\|^2_{\mathrm{osc}}.$$

We make use of the following two results, the proof of which can be found in [25, 12, 39]. Let $g := f + c$ for $c \in \mathbb{R}$. Firstly, it can be shown using the Cauchy–Schwarz inequality that

$$\frac{\mathbb{E}_{\mu\otimes P}\left[\left|g^2(X) - g^2(Y)\right|\right]^2}{\mathbb{E}_{\mu}\left[g^2(X)\right]} \leq 8\mathcal{E}(P, f). \tag{14}$$

Note that since $\|f\|_2^2 = 1$ and $\mu(f) = 0$, $\mathbb{E}_\mu\left[g^2(X)\right] = \|g - c + c\|^2 = 1 + c^2$. Secondly, it can also be established (following the proof in [39], say) that

$$\max\left\{\lim_{c\to\infty}\frac{\mathbb{E}_{\mu\otimes\mu}\left[|g^2(X)-g^2(Y)|\right]^2}{\mathbb{E}_\mu[g^2(X)]}, \frac{\mathbb{E}_{\mu\otimes\mu}\left[|f^2(X)-f^2(Y)|\right]^2}{1}\right\} \geq 1, \quad (15)$$

where the second term in the braces corresponds to the choice $c = 0$. The bound in (15) is used below to lower bound the left-hand side of (14). Consider the family of sets $\mathcal{T}_s := \{t \geq 0 : \mu\otimes\mu(A_t, A_t^\complement) > s\} \subset [0, \infty)$ for $s > 0$. Then using successively Lemma 119 with $\nu = \mu\otimes\mu$, the bound $(a+b)^2 \leq 2a^2 + 2b^2$, the definition of $\kappa(s)$, Lemma 119 with $\nu = \mu\otimes P$ and (14), we obtain for any $c \in \mathbb{R}$ and $s > 0$,

$$\begin{aligned}
&\frac{\mathbb{E}_{\mu\otimes\mu}\left[|g^2(X)-g^2(Y)|\right]^2}{2\mathbb{E}_\mu[g^2(X)]} \\
&\quad = \frac{\left(2\int_0^\infty \mu\otimes\mu(A_t, A_t^\complement)\,\mathrm{d}t\right)^2}{2\mathbb{E}_\mu[g^2(X)]} \\
&\quad = \frac{\left(2\int_{\mathcal{T}_s} \mu\otimes\mu(A_t, A_t^\complement)\,\mathrm{d}t + 2\int_{\mathcal{T}_s^\complement} \mu\otimes\mu(A_t, A_t^\complement)\,\mathrm{d}t\right)^2}{2\mathbb{E}_\mu[g^2(X)]} \\
&\quad \leq \frac{\left(2\int_{\mathcal{T}_s} \mu\otimes\mu(A_t, A_t^\complement)\,\mathrm{d}t\right)^2}{\mathbb{E}_\mu[g^2(X)]} + \frac{\left(2\int_{\mathcal{T}_s^\complement} \mu\otimes\mu(A_t, A_t^\complement)\,\mathrm{d}t\right)^2}{\mathbb{E}_\mu[g^2(X)]} \\
&\quad \leq \frac{\left(\frac{1}{\kappa(s)}2\int_{\mathcal{T}_s} \mu\otimes P(A_t \times A_t^\complement)\,\mathrm{d}t\right)^2}{\mathbb{E}_\mu[g^2(X)]} + \frac{\left(2\int_{\mathcal{T}_s^\complement} \mu\otimes\mu(A_t, A_t^\complement)\,\mathrm{d}t\right)^2}{\mathbb{E}_\mu[g^2(X)]} \\
&\quad \leq \frac{\frac{1}{\kappa^2(s)}\mathbb{E}_{\mu\otimes P}\left[|g^2(X)-g^2(Y)|\right]^2}{\mathbb{E}_\mu[g^2(X)]} + \frac{\left(2\int_{\mathcal{T}_s^\complement} \mu\otimes\mu(A_t, A_t^\complement)\,\mathrm{d}t\right)^2}{\mathbb{E}_\mu[g^2(X)]} \\
&\quad \leq \frac{8}{\kappa^2(s)}\mathcal{E}(P, f) + \frac{\left(2\int_{\mathcal{T}_s^\complement} \mu\otimes\mu(A_t, A_t^\complement)\,\mathrm{d}t\right)^2}{\mathbb{E}_\mu[g^2(X)]}.
\end{aligned}$$

We now focus on the second term. We begin with the case $c = 0$. For $t \in \mathcal{T}_s^\complement$,

$$\mu(A_t)\mu(A_t^\complement) = \mu(g^2(X) \geq t)\mu(g^2(X) < t) \leq s.$$

In particular, since we are assuming that $\|f\|_\infty < \infty$, if $t > (\|f\|_\infty + |c|)^2$, then $\mu(g^2(X) \geq t) = 0$. This enables us to bound, in the case $c = 0$: since we have $\|f\|_\infty^2 \leq \|f\|_{\mathrm{osc}}^2$,

$$\begin{aligned}
\int_{\mathcal{T}_s^\complement} \mu\otimes\mu(A_t, A_t^\complement)\,\mathrm{d}t &\leq \int_0^{\|f\|_\infty^2} s\,\mathrm{d}t \\
&\leq s\|f\|_{\mathrm{osc}}^2.
\end{aligned}$$

From Lemma 119 we also have the bound

$$\begin{aligned}\int_{\mathcal{T}_s^{\complement}} \mu\otimes\mu(A_t, A_t^{\complement})\,\mathrm{d}t &\leq \mathbb{E}_{\mu\otimes\mu}\left[|g^2(X)-g^2(Y)|\right] \\ &\leq 2\mathbb{E}_\mu\left[f^2(X)\right] = 2\,.\end{aligned}$$

Using these two bounds to upper bound the square below, we obtain that for $c = 0$,

$$\frac{\left(2\int_{\mathcal{T}_s^{\complement}} \mu\otimes\mu(A_t, A_t^{\complement})\,\mathrm{d}t\right)^2}{\mathbb{E}_\mu[g^2(X)]} \leq 4s\|f\|_{\mathrm{osc}}^2 \cdot 2.$$

We now consider the case $c \to \infty$. Since we are interested in the case when $c > \|f\|_\infty$, we know that $g > 0$ everywhere. In particular, this implies that if $t > (\mathrm{ess}_\mu \sup f + c)^2$, then $\mu(g^2(X) \geq t) = 0$. Similarly, if $t < (c + \mathrm{ess}_\mu \inf f)^2$, then $\mu(g^2(X) < t) = 0$. Thus we bound

$$\begin{aligned}&\int_{\mathcal{T}_s^{\complement}} \mu\otimes\mu(A_t, A_t^{\complement})\,\mathrm{d}t \\ &\qquad = \int_0^\infty \mu\otimes\mu(A_t, A_t^{\complement})1_{\mathcal{T}_s^{\complement}}(t)\,\mathrm{d}t \\ &\qquad = \int_{(c+\mathrm{ess}_\mu \inf f)^2}^{(\mathrm{ess}_\mu \sup f + c)^2} \mu\otimes\mu(A_t, A_t^{\complement})1_{\mathcal{T}_s^{\complement}}(t)\,\mathrm{d}t \\ &\qquad \leq s\left[(\mathrm{ess}_\mu \sup f + c)^2 - (c + \mathrm{ess}_\mu \inf f)^2\right] \\ &\qquad = s\left[(\mathrm{ess}_\mu \sup f)^2 - (\mu\text{-}\mathrm{ess}_\mu \inf f)^2 + 2c\,(\mathrm{ess}_\mu \sup f - \mathrm{ess}_\mu \inf f)\right] \\ &\qquad = s\left[(\mathrm{ess}_\mu \sup f)^2 - (\mathrm{ess}_\mu \inf f)^2 + 2c\|f\|_{\mathrm{osc}}\right].\end{aligned}$$

So ultimately we obtain

$$\begin{aligned}\frac{\left(2\int_{\mathcal{T}_s^{\complement}} \mu\otimes\mu(A_t, A_t^{\complement})\,\mathrm{d}t\right)^2}{\mathbb{E}_\mu[g^2(X)]} \leq s\,&\frac{4\left[(\mathrm{ess}_\mu \sup f)^2 - (\mathrm{ess}_\mu \inf f)^2 + 2c\|f\|_{\mathrm{osc}}\right]}{\sqrt{1+c^2}} \\ &\times \frac{\mathbb{E}_{\mu\otimes\mu}\left[|g^2(X)-g^2(Y)|\right]}{\sqrt{\mathbb{E}_\mu[g^2(X)]}}.\end{aligned}$$

Then taking the limit, we get

$$\begin{aligned}\limsup_{c\to\infty} \frac{\left(2\int_{\mathcal{T}_s^{\complement}} \mu\otimes\mu(A_t, A_t^{\complement})\,\mathrm{d}t\right)^2}{\mathbb{E}_\mu[g^2(X)]} &\leq s\cdot 8\|f\|_{\mathrm{osc}}\cdot 2\|f\|_{\mathrm{osc}} \\ &= s\cdot 16\|f\|_{\mathrm{osc}}^2.\end{aligned}$$

Rearranging then gives the desired bound. $\square$

4.2 WPIs from RUPI and μ-irreducibility

Given our notion of a WPI in Definition 3, a natural question to ask is under what general conditions on a kernel T, a WPI for T will hold. In particular, a WPI for the kernel $T = (P^*)^k P^k$ for $k \in \mathbb{N}$ enables one to deduce (subgeometric) convergence bounds for $\|P^{kn} f\|_2$, where $f \in \mathrm{L}_0^2(\mu)$ is such that $\Phi(f) < \infty$. Thus, we seek simple conditions on a Markov kernel T under which (1) will hold, for sieve $\Phi = \|\cdot\|_{\mathrm{osc}}^2$, with $T = P$ or $T = (P^*)^k P^k$ for $k \in \mathbb{N}$, for a finite-valued function α.

We will see that for a Markov operator T, a necessary and sufficient condition for a $(\|\cdot\|_{\mathrm{osc}}^2, \alpha)-$WPI to hold is the *resolvent-uniform-positivity-improving* (RUPI) property. This property appeared in [18], and in [43] it was suggested that an equivalence between the RUPI property and the existence of a WPI was already established in an unpublished manuscript by L. Wu. However, we have not been able to access this manuscript, and so in Section 4.2.1 we provide a direct proof of this equivalence.

In Section 4.2.2 we will demonstrate that arbitrarily small, uniform holding probabilities allow one to relate the existence of $\|\cdot\|_{\mathrm{osc}}^2$-WPIs for P, P^*P and $\|\cdot\|_{\mathrm{osc}}^2$-convergence of P (see Proposition 58), and also to deduce that $\|\cdot\|_{\mathrm{osc}}^2$-convergence of P and its additive reversibilization can similarly be closely related with a non-zero holding probabilities (see Proposition 55).

Furthermore, a simple sufficient condition for RUPI (and hence a WPI) is *μ-irreducibility*, which we discuss in detail in Section 4.2.3; see Corollary 63.

Hereafter we may omit the statement $A, B \in \mathscr{E}$ to alleviate notation; no confusion should be possible.

4.2.1 Equivalence of $\|\cdot\|_{\mathrm{osc}}^2$-WPI and RUPI

Definition 42 (UPI and RUPI)**.** A kernel T is *uniform-positivity-improving* (UPI) if for each $\epsilon > 0$,

$$\inf\{\langle \mathbf{1}_A, T\mathbf{1}_B\rangle : \mu(A) \wedge \mu(B) \geq \epsilon\} > 0.$$

A Markov kernel T is said to be *resolvent-uniform-positivity-improving* (RUPI) if for some (and hence all) $0 < \lambda < 1$, we have that the resolvent

$$R(\lambda, T) := \sum_{n=0}^{\infty} \lambda^n T^n = (\mathrm{Id} - \lambda T)^{-1},$$

is UPI.

Theorem 43. *Suppose that T is a μ–invariant Markov kernel. Then T satisfies an $\|\cdot\|_{\mathrm{osc}}^2$-WPI if and only if T is RUPI.*

Proof. This follows from Proposition 47 and Proposition 51 below. □

We follow [43] and give an equivalent condition for RUPI which will be convenient to work with.

Lemma 44. *An equivalent condition for a Markov kernel T to be RUPI is the following: for any $\epsilon > 0$, there exists $m \in \mathbb{N}$ such that*

$$\inf \left\{ \left\langle \mathbf{1}_A, \sum_{n=0}^{m} T^n \mathbf{1}_B \right\rangle : \mu(A) \wedge \mu(B) \geq \epsilon \right\} > 0. \tag{16}$$

Proof. The condition in Lemma 44 directly implies RUPI. To see this, take $\lambda \in (0,1)$, $\epsilon > 0$ and let $m \in \mathbb{N}$ such that $\inf \{\langle \mathbf{1}_A, \sum_{n=0}^{m} T^n \mathbf{1}_B \rangle : \mu(A) \wedge \mu(B) \geq \epsilon\} > 0$, which exists by assumption. Write

$$\begin{aligned} \langle \mathbf{1}_A, R(\lambda, T)\, \mathbf{1}_B \rangle &= \left\langle \mathbf{1}_A, \sum_{n=0}^{\infty} \lambda^n T^n \mathbf{1}_B \right\rangle \\ &\geqslant \left\langle \mathbf{1}_A, \sum_{n=0}^{m} \lambda^n T^n \mathbf{1}_B \right\rangle \\ &\geqslant \lambda^m \cdot \left\langle \mathbf{1}_A, \sum_{n=0}^{m} T^n \mathbf{1}_B \right\rangle \end{aligned}$$

to deduce that

$$\begin{aligned} &\inf\{\langle \mathbf{1}_A, R(\lambda, T)\, \mathbf{1}_B \rangle : \mu(A) \wedge \mu(B) \geq \epsilon\} \\ &\qquad \geqslant \lambda^m \cdot \inf \left\{ \left\langle \mathbf{1}_A, \sum_{n=0}^{m} T^n \mathbf{1}_B \right\rangle : \mu(A) \wedge \mu(B) \geq \epsilon \right\} > 0 \end{aligned}$$

from which the RUPI condition follows.

Conversely, suppose that T is RUPI, fix $\lambda \in (0,1)$ and assume that for some $\epsilon > 0$, (16) does not hold for any $m \in \mathbb{N}$. We show that this leads to a contradiction. By the RUPI assumption we have that $\delta := \inf\{\langle \mathbf{1}_A, \sum_{n=0}^{\infty} \lambda^n T^n \mathbf{1}_B \rangle : \mu(A) \wedge \mu(B) \geq \epsilon\} > 0$. Choose $m \in \mathbb{N}$ large enough so that

$$\sum_{n=m+1}^{\infty} \lambda^n < \delta/2.$$

Since we have assumed that (16) is violated for $\epsilon > 0$ and $m \in \mathbb{N}$ as chosen above there exists a sequence $\{(A_j, B_j)\}_{j=1}^{\infty}$ of sets all with mass at least ϵ such that $\langle \mathbf{1}_{A_j}, \sum_{n=0}^{m} T^n \mathbf{1}_{B_j} \rangle \to 0$, therefore implying for any $j \in \mathbb{N}$,

$$\begin{aligned}\delta \le \left\langle \mathbf{1}_{A_j}, \sum_{n=0}^{\infty} \lambda^n T^n \mathbf{1}_{B_j} \right\rangle &= \left\langle \mathbf{1}_{A_j}, \sum_{n=0}^{m} \lambda^n T^n \mathbf{1}_{B_j} \right\rangle + \left\langle \mathbf{1}_{A_j}, \sum_{n=m+1}^{\infty} \lambda^n T^n \mathbf{1}_{B_j} \right\rangle \\ &\leqslant \left\langle \mathbf{1}_{A_j}, \sum_{n=0}^{m} T^n \mathbf{1}_{B_j} \right\rangle + \sum_{n=m+1}^{\infty} \lambda^n \\ &\leqslant \left\langle \mathbf{1}_{A_j}, \sum_{n=0}^{m} T^n \mathbf{1}_{B_j} \right\rangle + \frac{\delta}{2} \\ &\overset{j\to\infty}{\to} \frac{\delta}{2},\end{aligned}$$

therefore leading to a contradiction. The conclusion follows. □

We first establish that for reversible kernels, RUPI implies a WPI for the resolvent.

Lemma 45. *Suppose that a reversible Markov kernel T is RUPI. Then for any $\lambda \in (0,1)$, the resolvent Markov kernel $S_\lambda := (1-\lambda)R(\lambda, T)$ is reversible and has the following property: for any $\epsilon > 0$,*

$$\inf_{A:\mu(A)\mu(A^{\complement})\ge\epsilon} \frac{\mathcal{E}(S_\lambda, \mathbf{1}_A)}{\mu(A)\mu(A^{\complement})} > 0.$$

Thus by Theorem 38, S_λ satisfies an $\|\cdot\|_{\mathrm{osc}}^2$-WPI.

Proof. Fix $\epsilon > 0$ and $\lambda \in (0,1)$. By the RUPI condition, $\inf\{\langle \mathbf{1}_A, S_\lambda \mathbf{1}_B\rangle : \mu(A) \wedge \mu(B) \ge \epsilon\} > 0$. In particular, if A is such that $\mu(A)\mu(A^{\complement}) \ge \epsilon$, we must have that both $\mu(A) \ge \epsilon$ and $\mu(A^{\complement}) \ge \epsilon$. Thus since

$$\mathcal{E}(S_\lambda, \mathbf{1}_A) = \langle \mathbf{1}_A, S_\lambda \mathbf{1}_{A^{\complement}} \rangle,$$

we must have that

$$\langle \mathbf{1}_A, S_\lambda \mathbf{1}_{A^{\complement}} \rangle \ge \delta > 0$$

for some $\delta > 0$, whenever $\mu(A)\mu(A^{\complement}) \ge \epsilon$. □

We now establish one direction of Theorem 43 through a sequence of lemmas: we first consider the case when T is reversible, and then deduce the case for general T; see Remark 4.

Lemma 46. *Suppose T is a reversible Markov kernel that is RUPI. Then T satisfies an $\|\cdot\|_{\mathrm{osc}}^2$-WPI.*

Proof. Since T is RUPI, we have established above in Lemma 45 that the resolvent $S_\lambda := (1-\lambda)R(\lambda, T)$ satisfies a WPI. In other words, we can find some $\alpha_\lambda : (0,\infty) \to [0,\infty)$ such that for any $f \in \mathrm{L}_0^2(\mu)$ and $r > 0$,

$$\|f\|_2^2 \le \alpha_\lambda(r)\langle (\mathrm{Id} - S_\lambda)f, f\rangle + r\|f\|_{\mathrm{osc}}^2.$$

Now, given a function $g \in \mathrm{L}^2_0(\mu)$, define $f := \frac{\mathrm{Id}-\lambda T}{1-\lambda} g \Leftrightarrow g = (1-\lambda)(\mathrm{Id}-\lambda T)^{-1} f$. (Note that since $0 < \lambda < 1$, the operator is $(\mathrm{Id} - \lambda T)$ invertible.)

Now since $g \in \mathrm{L}^2_0(\mu)$, we have that $f \in \mathrm{L}^2_0(\mu)$; for instance, consider the power series representation of $R(\lambda, T)$. Furthermore, we have that

$$\begin{aligned} \|g\|_2^2 &= (1-\lambda)^2 \|(\mathrm{Id} - \lambda T)^{-1} f\|_2^2 \\ &\le (1-\lambda)^2 \|(\mathrm{Id} - \lambda T)^{-1}\|^2 \|f\|_2^2 \\ &\le \|f\|_2^2, \end{aligned}$$

since the operator norm $\|(\mathrm{Id} - \lambda T)^{-1}\| \le \frac{1}{\lambda} \cdot \frac{1}{1/\lambda - 1} = \frac{1}{1-\lambda}$, by standard norm bounds for resolvents based on the distance to the spectrum.

Thus we have

$$\begin{aligned} \|g\|_2^2 &\le \|f\|_2^2 \le \alpha_\lambda(r) \langle (\mathrm{Id} - S_\lambda) f, f\rangle + r\|f\|_{\mathrm{osc}}^2 \\ &= \alpha_\lambda(r) \langle (\mathrm{Id} - (1-\lambda)(\mathrm{Id} - \lambda T)^{-1}) f, f\rangle + r\|f\|_{\mathrm{osc}}^2 \\ &= \alpha_\lambda(r) \left\langle \frac{\mathrm{Id} - \lambda T}{1-\lambda} g - g, \frac{\mathrm{Id} - \lambda T}{1-\lambda} g \right\rangle + r\|f\|_{\mathrm{osc}}^2 \\ &= \alpha_\lambda(r) \left\langle \frac{\lambda}{1-\lambda}(\mathrm{Id} - T) g, g + \frac{\lambda}{1-\lambda}(\mathrm{Id} - T) g \right\rangle + r\|f\|_{\mathrm{osc}}^2 \\ &= \alpha_\lambda(r) \left\{ \frac{\lambda}{1-\lambda} \langle (\mathrm{Id} - T) g, g\rangle + \left(\frac{\lambda}{1-\lambda}\right)^2 \|(\mathrm{Id} - T) g\|^2 \right\} + r\|f\|_{\mathrm{osc}}^2. \end{aligned}$$

Now we have that

$$\begin{aligned} \|(\mathrm{Id} - T) g\|^2 &= \langle (\mathrm{Id} - T) g, (\mathrm{Id} - T) g\rangle \\ &= \langle (\mathrm{Id} - T) g, g\rangle - \langle (\mathrm{Id} - T) g, T g\rangle. \end{aligned}$$

It is enough to bound this final term by

$$-\langle (\mathrm{Id} - T) g, T g\rangle \le \langle (\mathrm{Id} - T) g, g\rangle.$$

To see why this inequality is true, note that it is equivalent to

$$\begin{aligned} 0 &\le \langle (\mathrm{Id} - T) g, (\mathrm{Id} + T) g\rangle \\ &= \langle (\mathrm{Id} + T)(\mathrm{Id} - T) g, g\rangle \\ &= \langle (\mathrm{Id} - T^2) g, g\rangle, \end{aligned}$$

where we have made use of reversibility of T. And we certainly have that $0 \le \langle (\mathrm{Id} - T^* T) g, g\rangle = \langle (\mathrm{Id} - T^2) g, g\rangle$.

Overall, this gives us that

$$\begin{aligned} \|g\|^2 &\le \alpha_\lambda(r) \left(\frac{\lambda}{1-\lambda} + 2\left(\frac{\lambda}{1-\lambda}\right)^2 \right) \langle (\mathrm{Id} - T) g, g\rangle + r \left\| \frac{\mathrm{Id} - \lambda T}{1-\lambda} g \right\|_{\mathrm{osc}}^2 \\ &\le \alpha_\lambda(r) \left(\frac{\lambda}{1-\lambda} + 2\left(\frac{\lambda}{1-\lambda}\right)^2 \right) \langle (\mathrm{Id} - T) g, g\rangle + r \cdot \frac{(1+\lambda)^2}{(1-\lambda)^2} \|g\|_{\mathrm{osc}}^2. \end{aligned}$$

By reparameterizing with $r' = r \cdot \frac{(1+\lambda)^2}{(1-\lambda)^2}$, this is a standard WPI for T. □

Proposition 47. *Suppose a μ-invariant Markov kernel T is RUPI. Then T satisfies an $\|\cdot\|^2_{\mathrm{osc}}$-WPI.*

Proof. It suffices to show that $(T+T^*)/2$ is RUPI, as then by Lemma 46, $(T+T^*)/2$ possesses a WPI, which is equivalent to T possessing a WPI (see Remark 4). Since T is RUPI, for any $\epsilon > 0$, we can find some $\delta > 0$ and $N \in \mathbb{N}$ such that whenever $\mu(A) \wedge \mu(B) \geq \epsilon$,

$$\left\langle \mathbf{1}_A, \sum_{n=0}^{N} T^n \mathbf{1}_B \right\rangle \geq \delta > 0. \tag{17}$$

So now we wish to obtain such a statement for the kernel $(T+T^*)/2$. So fix $\epsilon > 0$, and consider

$$\left\langle \mathbf{1}_A, \sum_{n=0}^{N} \left(\frac{T+T^*}{2}\right)^n \mathbf{1}_B \right\rangle = \left\langle \mathbf{1}_A, \sum_{n=0}^{N} \frac{T^n}{2^n} \mathbf{1}_B \right\rangle + \langle \mathbf{1}_A, R\mathbf{1}_B \rangle,$$

where R is a sum of operators of the form $cT^{a_1}(T^*)^{b_1} \cdot \dots \cdot T^{a_r}(T^*)^{b_r}$ for some $r \in \mathbb{N}$, $a_i, b_i \in \mathbb{N}_0$ for all $i = 1, \dots, r$ and $c \geq 0$. Thus since T and T^* are Markov kernels, we have that $\langle \mathbf{1}_A, R\mathbf{1}_B \rangle \geq 0$. So we can continue and have, for any sets with $\mu(A) \wedge \mu(B) \geq \epsilon$,

$$\begin{aligned}
\left\langle \mathbf{1}_A, \sum_{n=0}^{N} \left(\frac{T+T^*}{2}\right)^n \mathbf{1}_B \right\rangle &\geq \left\langle \mathbf{1}_A, \sum_{n=0}^{N} \frac{T^n}{2^n} \mathbf{1}_B \right\rangle \\
&\geq 2^{-N} \langle \mathbf{1}_A, \sum_{n=0}^{N} T^n \mathbf{1}_B \rangle \\
&\geq \delta/2^N > 0,
\end{aligned}$$

since each summand is positive, and we have used the fact that T is RUPI (17). ☐

For the other direction, we first prove some auxiliary lemmas.

Lemma 49 is a general state space extension of the argument referenced by [34, Remark 2.16].

Lemma 48. $P(x, \{x\}) = P^*(x, \{x\})$ *for μ-almost all x.*

Proof. Let $D = \{(x,y) \in \mathsf{E}^2 : (x = y)\}$, $s(x) := P(x, \{x\})$ and $s^*(x) := P^*(x, \{x\})$ for $x \in \mathsf{E}$. For any $B \in \mathcal{E}$, we have

$$\mu(\mathbf{1}_B \cdot s) = \mu \otimes P(D \cap B^2) = \mu \otimes P^*(D \cap B^2) = \mu(\mathbf{1}_B \cdot s^*),$$

and so taking $B_+ = \{x \in \mathsf{E} : s(x) > s^*(x)\}$ and $B_- = \{x \in \mathsf{E} : s(x) < s^*(x)\}$ we deduce

$$\mu\left((s-s^*)^+\right) = 0 = \mu\left((s-s^*)^-\right),$$

and hence $s = s^*$ μ-almost everywhere. ☐

Lemma 49. *Assume $P(x, \{x\}) \geq \varepsilon$ for μ-almost all x. Then $\mathcal{E}(P^*P, f) \geq 2\varepsilon\mathcal{E}(P, f)$.*

Proof. We have $P^*(x, \{x\}) \geq \varepsilon$ for μ-almost all x by Lemma 48. Hence,

$$\begin{aligned}\mathcal{E}(P^*P, f) &= \frac{1}{2}\int \mu(\mathrm{d}x)P^*P(x, \mathrm{d}y)\left\{f(x) - f(y)\right\}^2 \\ &\geq \frac{1}{2}\int \mu(\mathrm{d}x)\left\{\varepsilon P(x, \mathrm{d}y) + \varepsilon P^*(x, \mathrm{d}y)\right\}\left\{f(x) - f(y)\right\}^2 \\ &= 2\varepsilon\mathcal{E}(P, f).\end{aligned}$$

□

The following is a useful implication of $\|\cdot\|_{\mathrm{osc}}^2$-convergence, that we will rely on below and also in Section 4.2.2.

Lemma 50. *Assume T is $\|\cdot\|_{\mathrm{osc}}^2$-convergent. Then for any $\epsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that for any $N \geq n_0$*

$$\inf\left\{\left\langle \mathbf{1}_A, T^N \mathbf{1}_B\right\rangle : \mu(A) \wedge \mu(B) \geq \epsilon\right\} > 0.$$

In particular, for all $k \in \mathbb{N}$, T^k is RUPI and T^k satisfies an $\|\cdot\|_{\mathrm{osc}}^2$-WPI.

Proof. Let $\epsilon > 0$ be arbitrary. Since T is $\|\cdot\|_{\mathrm{osc}}^2$-convergent, we may take $n_0 \in \mathbb{N}$ large enough such that $\|T^N f\|_2 \leq \|f\|_{\mathrm{osc}}\epsilon^2/2$ for all $N \geq n_0$. Let $A, B \in \mathscr{E}$ be such that $\mu(A) \wedge \mu(B) \geq \epsilon$. For any $N \geq n_0$ we have

$$\begin{aligned}\langle \mathbf{1}_A, T^N \mathbf{1}_B\rangle &= \langle \mathbf{1}_A, (T^N - \mu)\mathbf{1}_B\rangle + \langle \mathbf{1}_A, \mu\mathbf{1}_B\rangle \\ &= \langle \mathbf{1}_A, (T^N - \mu)\mathbf{1}_B\rangle + \mu(A)\mu(B).\end{aligned}$$

Let $f = \mathbf{1}_B - \mu(B)$ and we have by Cauchy–Schwarz,

$$|\langle \mathbf{1}_A, (T^N - \mu)\mathbf{1}_B\rangle| = |\langle \mathbf{1}_A, T^N f\rangle| \leq \mu(A)^{1/2}\|f\|_{\mathrm{osc}}\epsilon^2/2 \leq \epsilon^2/2,$$

and therefore

$$\langle \mathbf{1}_A, T^N \mathbf{1}_B\rangle \geq -\epsilon^2/2 + \epsilon^2 = \epsilon^2/2 > 0,$$

from which we can conclude. Now let $k \in \{1, 2, \ldots\}$ be arbitrary. Since we may choose N to be a multiple of k it follows from Lemma 44 that T^k is RUPI. Hence, by Proposition 47 T^k satisfies an $\|\cdot\|_{\mathrm{osc}}^2$-WPI. □

Proposition 51. *Let T be a μ-invariant Markov kernel satisfying a $(\|\cdot\|_{\mathrm{osc}}^2, \alpha)-$WPI for some $\alpha : (0, \infty) \to [0, \infty)$. Then T is RUPI.*

Proof. Consider the Markov operator $\tilde{T} := \frac{1}{2}(\mathrm{Id}+T)$, which satisfies $\tilde{T}(x, \{x\}) \geq 1/2$ by construction. Note that

$$\begin{aligned}\mathcal{E}\left(\tilde{T}, f\right) &= \langle(\mathrm{Id} - (\mathrm{Id} + T)/2)f, f\rangle \\ &= \frac{1}{2}\langle(\mathrm{Id} - T)f, f\rangle \\ &= \frac{1}{2}\mathcal{E}(T, f).\end{aligned}$$

Therefore, since T satisfies a $(\|\cdot\|_{\mathrm{osc}}^2, \alpha)-$WPI, we have that $\tilde{T}$ satisfies a $(\|\cdot\|_{\mathrm{osc}}^2, 2\alpha)-$WPI:

$$\begin{aligned}\|f\|_2^2 &\leq \alpha(r)\mathcal{E}(T,f) + r\|f\|_{\mathrm{osc}}^2 \\ &= 2\alpha(r)\mathcal{E}\left(\tilde{T},f\right) + r\|f\|_{\mathrm{osc}}^2.\end{aligned}$$

Since $\mathrm{ess}_\mu \inf_x \tilde{T}(x,\{x\}) \geq 1/2$, by Lemma 49 we have the inequality $\mathcal{E}(\tilde{T},f) \leq \mathcal{E}(\tilde{T}^*\tilde{T},f)$, so we deduce a $(\|\cdot\|_{\mathrm{osc}}^2, 2\alpha)-$WPI for $\tilde{T}^*\tilde{T}$. Hence, $\tilde{T}$ is $\|\cdot\|_{\mathrm{osc}}^2$-convergent by Theorem 7.

We will now verify the condition for RUPI in Lemma 44. Let $\epsilon \in (0,1)$ be arbitrary. Since $\tilde{T}$ is $\|\cdot\|_{\mathrm{osc}}^2$-convergent, Lemma 50 implies that there exists $N \in \mathbb{N}$ such that

$$\delta = \inf\left\{\left\langle \mathbf{1}_A, \tilde{T}^N \mathbf{1}_B \right\rangle : \mu(A) \wedge \mu(B) \geq \epsilon\right\} > 0.$$

Now,

$$\tilde{T}^N = \left(\frac{\mathrm{Id}+T}{2}\right)^N = \frac{1}{2^N}\sum_{k=0}^N a_k T^k,$$

for binomial coefficients $\{a_i\}$. Since $\sum_{i=0}^N a_i = 2^N$ and $\mathbf{1}_A$, $\mathbf{1}_B$ are non-negative, we have

$$\left\langle \mathbf{1}_A, \sum_{k=0}^N T^k \mathbf{1}_B \right\rangle \geq \left\langle \mathbf{1}_A, \left(\frac{\mathrm{Id}+T}{2}\right)^N \mathbf{1}_B \right\rangle, \qquad A, B \in \mathscr{E},$$

and this implies that

$$\inf\left\{\left\langle \mathbf{1}_A, \sum_{k=0}^N T^k \mathbf{1}_B \right\rangle : \mu(A) \wedge \mu(B) \geq \epsilon\right\} \geq \delta > 0,$$

so T is RUPI. □

4.2.2 Holding probabilities, WPIs and $\|\cdot\|_{\mathrm{osc}}^2$-convergence

Definition 52. For a μ-invariant Markov kernel T, and $\epsilon \in (0,1)$ we denote by T_ϵ the μ-invariant kernel $T_\epsilon = \epsilon\mathrm{Id} + (1-\epsilon)T$.

We show in this section that there are close connections between existence of an $\|\cdot\|_{\mathrm{osc}}^2$-WPI for a Markov kernel P, and existence of an $\|\cdot\|_{\mathrm{osc}}^2$-WPI for $P_\epsilon^* P_\epsilon$, where ϵ is any non-trivial *holding probability*. This is also closely connected to $\|\cdot\|_{\mathrm{osc}}^2$-convergence.

Throughout this section, we write $S := (P+P^*)/2$ for the additive reversibilization of P.

Proposition 53. *Let $\epsilon \in (0,1)$. Then P satisfies an $\|\cdot\|_{\mathrm{osc}}^2$-WPI if and only if $P_\epsilon^* P_\epsilon$ satisfies an $\|\cdot\|_{\mathrm{osc}}^2$-WPI.*

Proof. This follows from Lemmas 59 and 60. □

Remark 54. Proposition 53, and some of the results below could also be phrased in terms of the alternative multiplicative reversibilizations of P_ϵ, i.e. $P_\epsilon P_\epsilon^*$.

Proposition 55. *The following hold:*

a). If P is $\|\cdot\|_{\mathrm{osc}}^2$-convergent, then S^2 satisfies an $\|\cdot\|_{\mathrm{osc}}^2$-WPI and S is $\|\cdot\|_{\mathrm{osc}}^2$-convergent.

b). Let $\epsilon \in (0,1)$. If S or P are $\|\cdot\|_{\mathrm{osc}}^2$-convergent then $P_\epsilon^* P_\epsilon$ satisfies an $\|\cdot\|_{\mathrm{osc}}^2$-WPI and P_ϵ is $\|\cdot\|_{\mathrm{osc}}^2$-convergent.

Proof. For the first part, if P is $\|\cdot\|_{\mathrm{osc}}^2$-convergent, then Lemma 50 implies that P^2 is RUPI and satisfies an $\|\cdot\|_{\mathrm{osc}}^2$-WPI. We may then deduce that S^2 satisfies an $\|\cdot\|_{\mathrm{osc}}^2$-WPI because for any $f \in \mathrm{L}_0^2(\mu)$,

$$\mathcal{E}(S^2, f) = \frac{1}{4}\left\{\mathcal{E}(P^2, f) + \mathcal{E}((P^*)^2, f) + \mathcal{E}(PP^*, f) + \mathcal{E}(P^*P, f)\right\} \geq \frac{1}{4}\mathcal{E}(P^2, f).$$

It follows that S is $\|\cdot\|_{\mathrm{osc}}^2$ -convergent by Theorem 7. For the second part, if S or P are $\|\cdot\|_{\mathrm{osc}}^2$ -convergent then Lemma 50 implies that S, or equivalently P, satisfies an $\|\cdot\|_{\mathrm{osc}}^2$-WPI. Hence, by Proposition 53, $P_\epsilon^* P_\epsilon$ satisfies an $\|\cdot\|_{\mathrm{osc}}^2$-WPI, from which we can deduce $\|\cdot\|_{\mathrm{osc}}^2$-convergence by Theorem 7. □

Remark 56. The appearance of $\epsilon \in (0,1)$ in the implication S is $\|\cdot\|_{\mathrm{osc}}^2$-convergent $\Rightarrow$ P_ϵ is $\|\cdot\|_{\mathrm{osc}}^2$-convergent cannot be removed, since it is possible that S is $\|\cdot\|_{\mathrm{osc}}^2$-convergent but P is not; see Example 57. On the other hand, S being $\|\cdot\|_{\mathrm{osc}}^2$-convergent is a necessary condition for P to be $\|\cdot\|_{\mathrm{osc}}^2$-convergent. The appearance of $\epsilon \in (0,1)$ in the implication P is $\|\cdot\|_{\mathrm{osc}}^2$-convergent $\Rightarrow P_\epsilon^* P_\epsilon$ satisfies an $\|\cdot\|_{\mathrm{osc}}^2$-WPI also cannot be removed; see Proposition 67 and note that in that example S^2 is μ-irreducible and so S is $\|\cdot\|_{\mathrm{osc}}^2$-convergent by Corollary 63.

Example 57 (Walks on the circle)**.** For $x, y \in \mathsf{E} = \{1, \ldots, m\}$ let $P(x,y) = \mathbf{1}_{\{1,\ldots,m-1\}}(x)\mathbf{1}_{\{x+1\}}(y) + \mathbf{1}_{\{m\}}(x)\mathbf{1}_{\{1\}}(y)$ so that $P^*(x,y) = \mathbf{1}_{\{2,\ldots,m\}}(x)\mathbf{1}_{\{x-1\}}(y) + \mathbf{1}_{\{1\}}(x)\mathbf{1}_{\{m\}}(y)$. Then the Markov chain associated with P is deterministic and one can deduce that P is not $\|\cdot\|_{\mathrm{osc}}^2$-convergent. On the other hand, $S = (P + P^*)/2$ encodes a random walk on $\{1, \ldots, m\}$ and is $\|\cdot\|_{\mathrm{osc}}^2$-convergent.

In practice, the following result may be useful.

Proposition 58. *Assume P is μ-invariant and satisfies* $\mathrm{ess}_\mu \inf_x P(x, \{x\}) \in (0,1)$*. Then the following are equivalent.*

a). P satisfies an $\|\cdot\|_{\mathrm{osc}}^2$-WPI;

b). P^*P satisfies an $\|\cdot\|_{\mathrm{osc}}^2$-WPI;

c). P is $\|\cdot\|_{\mathrm{osc}}^2$-convergent;

d). PP^* satisfies an $\|\cdot\|_{\mathrm{osc}}^2$-WPI;

e). P^* is $\|\cdot\|_{\mathrm{osc}}^2$-convergent.

Proof. (b. $\Rightarrow$ c.) follows from Theorem 7, and (c. $\Rightarrow$ a.) follows from Lemma 50. We now show (a. $\Rightarrow$ b.). Let $\varepsilon = \mathrm{ess}_\mu \inf_x P(x, \{x\}) \in (0,1)$. Then $T := (P - \varepsilon \mathrm{Id})/(1-\varepsilon)$ is also a μ-invariant Markov kernel and also satisfies an $\|\cdot\|_{\mathrm{osc}}^2$-WPI since $\mathcal{E}(T,f) = (1-\varepsilon)^{-1}\mathcal{E}(P,f)$. Since $P = T_\varepsilon$, we deduce by Proposition 53 that $P^*P = T_\varepsilon^* T_\varepsilon$ satisfies an $\|\cdot\|_{\mathrm{osc}}^2$-WPI.

We now show that the cycle (a. $\Rightarrow$ d. $\Rightarrow$ e. $\Rightarrow$ a.) can also be deduced. Observe that $\mathrm{ess}_\mu \inf_x P^*(x,\{x\}) = \mathrm{ess}_\mu \inf_x P(x,\{x\})$ by Lemma 48, and P satisfying an $\|\cdot\|_{\mathrm{osc}}^2$-WPI is equivalent to P^* satisfying an $\|\cdot\|_{\mathrm{osc}}^2$-WPI, since $\mathcal{E}(P,f) = \mathcal{E}(P^*,f)$. Because $(P^*)^*P^* = PP^*$, we have that (a. $\Rightarrow$ d.) is equivalent to (a. $\Rightarrow$ b.) and (d. $\Rightarrow$ e.) is equivalent to (b. $\Rightarrow$ c.) and (e. $\Rightarrow$ a.) is equivalent to (c. $\Rightarrow$ a.). $\square$

Lemma 59. *Let P be μ-invariant and assume P, or equivalently S, satisfies a (Φ,α)-WPI. For $\epsilon \in (0,1)$, $P_\epsilon^* P_\epsilon$ satisfies a $(\Phi, \frac{1}{2\epsilon(1-\epsilon)}\alpha)$-WPI.*

Proof. It is straightforward to verify that $P_\epsilon^* = \epsilon \mathrm{Id} + (1-\epsilon)P^*$, and therefore

$$P_\epsilon^* P_\epsilon = \epsilon^2 \mathrm{Id} + \epsilon(1-\epsilon)(P^* + P) + (1-\epsilon)^2 P^* P.$$

It follows that

$$\mathcal{E}(P_\epsilon^* P_\epsilon, f) \geq 2\epsilon(1-\epsilon)\mathcal{E}\left(\frac{P^*+P}{2}, f\right) = 2\epsilon(1-\epsilon)\mathcal{E}\left(P,f\right),$$

from which we may conclude. $\square$

Lemma 60. *Let $\epsilon \in (0,1)$. If $P_\epsilon^* P_\epsilon$ satisfies an $\|\cdot\|_{\mathrm{osc}}^2$-WPI then P satisfies an $\|\cdot\|_{\mathrm{osc}}^2$-WPI.*

Proof. First, suppose we have a Markov kernel T such that T^*T satisfies an $\|\cdot\|_{\mathrm{osc}}^2$-WPI. Then for $S_T := (T+T^*)/2$ we have that

$$S_T^2 = \frac{1}{4}\left\{T^2 + (T^*)^2 + TT^* + T^*T\right\},$$

and so $\mathcal{E}(S_T^2, f) \geq \mathcal{E}(T^*T, f)/4$. This implies that S_T^2 also satisfies an $\|\cdot\|_{\mathrm{osc}}^2$-WPI.

Hence, $P_\epsilon^* P_\epsilon$ satisfying an $\|\cdot\|_{\mathrm{osc}}^2$-WPI implies that $(S_\epsilon)^2$ satisfies an $\|\cdot\|_{\mathrm{osc}}^2$-WPI, where $S_\epsilon := (P_\epsilon + P_\epsilon^*)/2$. It follows from Theorem 43 that $(S_\epsilon)^2$ is RUPI, which implies that S_ϵ is RUPI since for any $m \in \mathbb{N}$,

$$\left\langle \mathbf{1}_A, \sum_{k=0}^{m} (S_\epsilon)^{2k} \mathbf{1}_B \right\rangle \leq \left\langle \mathbf{1}_A, \sum_{k=0}^{2m} (S_\epsilon)^{k} \mathbf{1}_B \right\rangle.$$

Since $S_\epsilon = (P_\epsilon + P_\epsilon^*)/2 = \epsilon \mathrm{Id} + (1-\epsilon)S$ with $S = (P+P^*)/2$, we may further deduce that S is RUPI since for any $m \in \mathbb{N}$,

$$\left\langle \mathbf{1}_A, \sum_{k=0}^{m} (S_\epsilon)^k \mathbf{1}_B \right\rangle = \left\langle \mathbf{1}_A, \sum_{k=0}^{m} \sum_{j=0}^{k} a_{k,j} S^j \mathbf{1}_B \right\rangle \le m \left\langle \mathbf{1}_A, \sum_{k=0}^{m} S^k \mathbf{1}_B \right\rangle,$$

where we have used the fact that for each $k \in \{0, \dots, m\}$, $a_{k,0}, a_{k,1}, \dots, a_{k,k} \ge 0$ and $\sum_j a_{k,j} = 1$. It follows that S and therefore P satisfy an $\|\cdot\|_{\mathrm{osc}}^2$-WPI. $\square$

4.2.3 μ-irreducibility implies a WPI

To establish that a given kernel T is RUPI, it is sufficient to show a simple irreducibility condition.

Definition 61. We say that a Markov kernel T on $(\mathsf{E}, \mathscr{E})$ is ν-irreducible for a measure ν on $(\mathsf{E}, \mathscr{E})$ if for any measurable set $A \in \mathscr{E}$ with $\nu(A) > 0$, we have that

$$\sum_{n=0}^{\infty} \lambda^n T^n(x, A) > 0, \quad \forall x \in \mathsf{E},$$

for some (and hence all) $0 < \lambda < 1$.

Proposition 62 ([18, Corollary 4.5])**.** *Suppose that T is μ-irreducible. Then T is RUPI.*

Thus we immediately obtain by Theorem 43 that μ-irreducibility is a sufficient condition for the existence of an $\|\cdot\|_{\mathrm{osc}}^2$-WPI.

Corollary 63. *Suppose the Markov kernel T is μ-irreducible. Then T possesses an $\|\cdot\|_{\mathrm{osc}}^2$-WPI by Proposition 62 and Theorem 43. Moreover, if $\mathrm{ess}_\mu \inf_x T(x, \{x\}) > 0$ then T^*T possesses an $\|\cdot\|_{\mathrm{osc}}^2$-WPI and T is $\|\cdot\|_{\mathrm{osc}}^2$-convergent by Proposition 58.*

It is important to note that P possessing a WPI does not necessarily imply that $\|P^n f\|_2 \to 0$ for all relevant functions. Indeed, a reversible, periodic Markov kernel may satisfy an $\|\cdot\|_{\mathrm{osc}}^2$-WPI yet $\|P^n f\|_2$ cannot converge to 0 for all bounded functions.

Remark 64. When P is reversible, it is possible to deduce the existence of a WPI for $P^2 = P^*P$ from a WPI for P, provided that one has some additional control on the left spectral gap; see [1, Section 2.2.1]. In turn, the existence of a WPI for P can often be straightforwardly deduced from Corollary 63 by establishing irreducibility of P.

Remark 65. Corollary 63 allows us to guarantee the existence of a $\|\cdot\|_{\mathrm{osc}}^2$-WPI for $T = (P^*)^k P^k$, for some $k \in \mathbb{N}$, in many situations. If P is reversible then $P^*P = P^2$ being μ-irreducible implies that a WPI exists for P^*P. Note that if P^2 is not $\mu-$irreducible, then neither is $(P^*)^k P^k = (P^2)^k$ for any $k > 1$. If P is μ-invariant and admits an $\mathrm{L}_0^2(\mu)$-spectral gap then there exists some $k \in \mathbb{N}$ such that $\left\|P^k f\right\|_2 \leq C \left\|f\right\|_2$ for some $C < 1$ and all $f \in \mathrm{L}_0^2(\mu)$ and hence $\left(P^*\right)^k P^k$ admits a strong Poincaré inequality. However, if P is nonreversible then even if $\left\|P^n f\right\|_2$ decays geometrically for bounded functions, it is possible that for all $k \in \mathbb{N}$, $\left(P^*\right)^k P^k$ is not μ-irreducible and does not admit a WPI; see Example 66 and Proposition 67.

The following example demonstrates (in case 1) that for an arbitrary $k \in \mathbb{N}$, there exists nonreversible P such that $\left(P^*\right)^k P^k$ is μ-irreducible while $\left(P^*\right)^i P^i$ is not μ-irreducible for any positive integer $i < k$. It also demonstrates (case 2) that $(P^*)^k P^k$ cannot be μ-irreducible for any $k \in \mathbb{N}$ even though P is μ-irreducible, and in this case it is not clear that one can define an appropriate WPI that provides a vanishing upper bound on $\|P^n f\|_2^2$.

In fact, similar examples have been considered by [19] and [40, Section 6], who are essentially interested in geometrically ergodic Markov chains for which a CLT fails to hold for an L^2 function, or which do not admit an L_0^2 spectral gap. We construct such an example in Proposition 67. Our consideration of the following family of examples is very natural; because there can be arbitrarily long periods of deterministic behaviour, lack of μ-irreducibility is straightforward to deduce.

Example 66. Let $\mathsf{E} = \{1, 2, \ldots\}^2$, and ν a probability mass function on $\{1, 2, \ldots\}$ such that $\nu(1) \in (0, 1)$ and ν has a finite mean. Define

$$P(i, j; i', j') = \begin{cases} 1 & j < i, i' = i, j' = j + 1, \\ \nu(i') & j = i, j' = 1, \\ 1 & \nu(i)\mathbf{1}\{j \leq i\} = 0, (i', j') = (1, 1), \\ 0 & \text{otherwise.} \end{cases}$$

The intuition is that the Markov chain moves to the right along "level" i deterministically until it reaches the point (i, i), at which point it jumps to the start of another level $(K, 1)$ where $K \sim \nu$. The third statement is concerned with initialization of the chain outside the support of the invariant distribution μ, which one can verify directly is given by

$$\mu(i, j) = \frac{\nu(i)\mathbf{1}\{j \leq i\}}{\sum_{k=1}^{\infty} \nu(k) k}.$$

P is μ-irreducible with an accessible, aperiodic atom $(1, 1)$ and its Markov chain converges to μ in total variation from any starting point (by, e.g., [12, Theorem 7.6.4]).

By viewing P^* as the time-reversal of P, and satisfying $\mu(i, j) P^*(i, j; i', j') = \mu(i', j') P(i', j'; i, j)$, we may define

$$P^*(i, j; i', j') = \begin{cases} 1 & j > 1, i' = i, j' = j - 1, \\ \nu(i') & j = 1, j' = i', \\ 1 & \nu(i)\mathbf{1}\{j \leq i\} = 0, (i', j') = (1, 1), \\ 0 & \text{otherwise.} \end{cases}$$

Case 1: Assume that for some $i_0 \in \{2, 3, \ldots\}$, $\nu(i) > 0$ for all $i \le i_0$ and $\nu(i) = 0$ for $i > i_0$. This means there is a maximum level length of i_0. We see that if $k < i_0$ then

$$(P^*)^k P^k(i_0, i_0; i_0, i_0) = 1,$$

since $(P^*)^k(i_0, i_0; i_0 - k, i_0 - k) = 1$ and $P^k(i_0 - k, i_0 - k; i_0, i_0) = 1$. Hence $(P^*)^k P^k$ is reducible for any $k < i_0 - 1$. On the other hand, for $k \ge i_0$, we may deduce that $(P^*)^k P^k$ is μ-irreducible. In particular, since $P^*(1, 1; 1, 1) = P(1, 1; 1, 1) = \nu(1) \in (0, 1)$, we see that $(P^*)^k(i, j; 1, 1) > 0$ for all $(i, j) \in \mathsf{E}$, from which one may deduce that $(P^*)^k P^k(i, j; i', j') > 0$ for all $i, j, i', j' \in \mathsf{E}$ such that $\mu(i', j') > 0$. Note that since $P(i, j; 1, 1) = 1$ for all (i, j) such that $\mu(i, j) = 0$, this is essentially a finite state space Markov chain after 1 step, and hence convergence is geometric.

Case 2: Assume that $\nu(i) > 0$ for all $i \in \{1, 2, \ldots\}$. For any $k \in \mathbb{N}$ we may consider level $i > k$ and we see that $(P^*)^k P^k(i, i; i, i) = 1$ so $(P^*)^k P^k$ is reducible. Hence, there does not exist $k \in \mathbb{N}$ such that $(P^*)^k P^k$ is μ-irreducible.

Our final result in this section shows that P being Φ-convergent does not imply that there exists $k \in \mathbb{N}$ such that $(P^*)^k P^k$ admits a Φ-WPI when P is nonreversible, even in the case where γ decays geometrically. We note that by Proposition 13, geometric convergence can be extended to all functions in $\mathrm{L}^p_0(\mu)$ for any $p > 2$.

Proposition 67. *For the chain in Example 66, let $\nu(i) = (1 - a)a^{i-1}$ for some $a \in (0, 1)$. Then*

a). P is geometrically ergodic and

$$\|P^n f\|_2^2 \le \|f\|_{\mathrm{osc}}^2 C\rho^{2n}, \qquad f \in \mathrm{L}^2_0(\mu),$$

for some $C > 0$ and $\rho \in (0, 1)$;

b). $(P^*)^k P^k$ does not admit an $\|\cdot\|_{\mathrm{osc}}^2$-WPI for any $k \in \mathbb{N}$.

Proof. We will apply [5, Theorem 1.1]. We may consider the state space to be the μ-full set $\mathsf{E} = \{(i, j) \in \{1, 2, \ldots\}^2 : j \le i\}$ for simplicity. We now verify the assumptions (A1)-(A3) in [5]. We first define the set $C = \{(i, j) \in E : i = j\}$. We define the probability measure $\tilde{\nu}(i, j) = \nu(i)\mathbf{1}\{j = 1\}$, and we have that

$$P(x, A) = \tilde{\nu}(A), \qquad x \in C.$$

For any $\lambda \in (\sqrt{a}, 1)$ we may define the Lyapunov function $V(i, j) := \lambda^{j-i}$ and we observe that $V \ge 1$ on E, with $PV(x) \le \lambda V(x)\mathbf{1}_{C^{\complement}}(x) + K\mathbf{1}_C(x)$, where

$$K = \tilde{\nu}(V) = \sum_{i \ge 1} \nu(i)V(i, 1) \propto \sum_{i \ge 1} a^i \lambda^{1-i} < \infty,$$

since $\lambda > \sqrt{a} > a$. Finally, we note that $\tilde{\nu}(C) = \nu(1) > 0$. It then follows by the theorem that there exist $M > 0$, $\rho \in (0, 1)$ such that

$$\sup_{f : |f| \le V, \mu(f) = 0} |P^n f(x)| \le MV(x)\rho^n.$$

Hence, we may deduce that for f such that $|f| \leq V$

$$\|P^n f\|_2^2 \leq M^2 \mu(V^2)\rho^{2n},$$

and since $\lambda > \sqrt{a}$, we have

$$\begin{aligned}
\mu(V^2) &= \sum_{i\geq 1}\sum_{j=1}^{i} \mu(i,j)V(i,j)^2 \\
&\propto \sum_{i\geq 1} a^i \sum_{j=1}^{i} \lambda^{2(j-i)} \\
&\leq \frac{\lambda^2}{1-\lambda^2}\sum_{i\geq 1}\frac{a^i}{\lambda^{2i}} \\
&< \infty.
\end{aligned}$$

The bound on $\|P^n f\|_2^2$ for bounded functions then follows since $V \geq 1$.

For the second part, let $k \in \mathbb{N}$ be arbitrary. Let $A_k = \{(i,i) : i > k\}$ and $f_k = \mathbf{1}_{A_k} - \mu(A_k)$, which satisfies $\mu(f_k) = 0$. Then $\left\|P^k f_k\right\|_2^2 = \left\langle (P^*)^k P^k f_k, f_k\right\rangle = \langle f_k, f_k\rangle = \|f_k\|_2^2$, so $\mathcal{E}((P^*)^k P^k, f_k) = 0$. Since $\|f_k\|_2 > 0$, $(P^*)^k P^k$ cannot satisfy a $\|\cdot\|_{\mathrm{osc}}^2$-WPI. □

4.3 Lyapunov meets Poincaré

A difficulty with functional-analytic approaches to the study of Markov chains is the challenge posed by unbounded supports for μ; in particular, handling the tails of μ. A general strategy consists of splitting the state space E into a distinguished set C on which a form of strong Poincaré inequality is established, while the behaviour of the chain on $C^{\complement}$ is handled with a Lyapunov drift function. Such ideas have been primarily explored for certain classes of continuous-time Markov processes, with [3] establishing quantitative strong Poincaré inequalities for the overdamped Langevin process; these results were later extended to heavy-tailed target distributions in [9] to establish WPIs. It is only recently that some of these ideas were extended to discrete-time Markov chains in [41] where a strategy to establish strong Poincaré inequalities is proposed; we note also the recent contribution of [7]. In this subsection we first briefly review the key results of [41], show how they can be improved in the spirit of [3] by using local Poincaré inequalities (Subsection 4.3.1). In Subsection 4.3.2, we show how these results can be extended to *subgeometric* drift conditions in order to establish WPIs.

We first define precisely the restriction of the μ-invariant kernel P to the set C and the notion of *local* Poincaré inequality.

Definition 68. For some $C \in \mathscr{E}_+$, we define the *restriction of* μ *to* C to be the probability measure μ_C supported on C given by

$$\mu_C(A) := \frac{\mu(A\cap C)}{\mu(C)}, \qquad A \in \mathscr{E},$$

and *the restriction of P to C* is defined to be the kernel P_C defined as: for each $x \in \mathsf{E}$,

$$P_C f(x) := P(f \cdot \mathbf{1}_C)(x) + f(x)P(x, C^{\complement}) .$$

We will say that a *restricted Poincaré inequality holds for P on C* if a strong Poincaré inequality holds for P_C: for some $\mathrm{C_r} > 0$ and all $f \in \mathrm{L}^2_0(\mu_C)$,

$$\begin{aligned}\|f\|^2_{\mu_C,2} &= \mu(C)^{-1} \int f^2(x)\mathbf{1}_C(x)\,\mu(\mathrm{d}x) \\ &\leq \mathrm{C_r}\mathcal{E}(P_C, f),\end{aligned} \tag{18}$$

where $\mathcal{E}(P_C, f) := \int \mu_C(\mathrm{d}x)P_C(x, \mathrm{d}y)\,[f(y) - f(x)]^2$ for any $f \in \mathrm{L}^2(\mu_C)$.

This can equivalently be expressed as requiring: for any $f \in \mathrm{L}^2(\mu)$,

$$\mu_C\big(f^2_m\big) \leq \mathrm{C_r}\mathcal{E}(P_C, f),$$

with $m := \mu(f \cdot \mathbf{1}_C)/\mu(C)$ and $f_m := f - m$.

Finally, we will say that a *local Poincaré inequality holds for P on C* if for some $\mathrm{C}_l > 0$, for all $f \in \mathrm{L}^2(\mu)$, there is some $m \in \mathbb{R}$ such that setting $f_m := f - m$, we have

$$\|f_m\mathbf{1}_C\|^2_2 \leq \mathrm{C}_l\mathcal{E}(P, f). \tag{19}$$

We note that when P is reversible, the restriction P_C is simply a Metropolis–Hastings Markov kernel targeting μ_C and using proposal distribution P. The μ_C-reversibility of such restrictions is well-known; a proof is provided for completeness. For a nonreversible, μ-invariant P it is also well-known that P_C is not necessarily μ_C-invariant.

Lemma 69. *Let P be a $\mu-$reversible Markov kernel. Then the kernel P_C is μ_C-reversible, and furthermore if a restricted Poincaré inequality (18) holds for P_C, then the the following local Poincaré inequality for P on C holds: for any $f \in \mathrm{L}^2(\mu)$,*

$$\|f_m\mathbf{1}_C\|^2_2 \leq \mathrm{C_r}\mathcal{E}(P, f),$$

with $m := \mu(f\mathbf{1}_C)/\mu(C)$ and $f_m := f - m$.

Proof. We first check μ_C-reversibility of P_C. For $f, g \in \mathrm{L}^2(\mu)$, let

$$v = \int f(x)g(x)\mu_C(\mathrm{d}x)P(x, C^{\complement}),$$

and by the $\mu-$reversibility of P, we have

$$\begin{aligned}\int f(x)g(y)\mu_C(\mathrm{d}x)P_C(x, \mathrm{d}y) &= \frac{1}{\mu(C)} \int f(x)\mathbf{1}_C(x)g(y)\mathbf{1}_C(y)\mu(\mathrm{d}x)P(x, \mathrm{d}y) + v \\ &= \frac{1}{\mu(C)} \int f(y)\mathbf{1}_C(y)g(x)\mathbf{1}_C(x)\mu(\mathrm{d}x)P(x, \mathrm{d}y) + v \\ &= \int f(y)g(x)\mu_C(\mathrm{d}x)P_C(x, \mathrm{d}y).\end{aligned}$$

Now from the restricted Poincaré inequality, we have

$$\begin{aligned}\int f_m^2(x)\mathbf{1}_C(x)\mathrm{d}\mu(\mathrm{d}x) &\le \mu(C)\cdot\frac{\mathrm{C_r}}{2}\int \mu_C(\mathrm{d}x)P_C(x,\mathrm{d}y)\,[f(y)-f(x)]^2\\ &= \frac{\mathrm{C_r}}{2}\int \mu(\mathrm{d}x)P\,(x,\mathrm{d}y)\;\mathbf{1}_C(x)\mathbf{1}_C\,(y)\;[f(x)-f(y)]^2\\ &\le \frac{\mathrm{C_r}}{2}\int \mu(\mathrm{d}x)P\,(x,\mathrm{d}y)\;[f(x)-f(y)]^2\\ &= \mathrm{C_r}\mathcal{E}(P,f)\,.\end{aligned}$$

□

In Section 4.3.3, we show how one can deduce local Poincaré inequalities when μ has a strongly log-concave density and a coupling argument

4.3.1 The geometric scenario

The following is the adaption of [3] to the discrete time scenario by [41] where we here replace the minorization condition with a local Poincaré inequality.

Theorem 70 ([41])**.** *Assume the existence of $C\subset \mathsf{X}$, a Lyapunov function $V\colon \mathsf{E}\to[1,\infty)$, and constants $K,b,>0$ and $\lambda\in(0,1]$ such that*

$$PV \leqslant (1-\lambda)\,V + b\mathbf{1}_C, \tag{20}$$

and that we have the following local Poincaré inequality for P on C: for any $f\in \mathrm{L}_0^2(\mu)$, there exists some $m>0$ such that for $f_m := f-m$,

$$\left\langle f_m^2,\mathbf{1}_C\right\rangle \leqslant K\left\langle f,(\mathrm{Id}-P)\,f\right\rangle. \tag{21}$$

Then we have the following (strong) Poincaré inequality for P: for any $f\in \mathrm{L}^2(\mu)$,

$$\frac{\lambda}{1+Kb}\|f-\mu(f)\|_2^2 \leqslant \left\langle f,(\mathrm{Id}-P)\,f\right\rangle.$$

Proof. From $PV\le(1-\lambda)V+b\mathbf{1}_C$ we obtain $\lambda V\le(\mathrm{Id}-P)V+b\mathbf{1}_C$, and so for $f\in\mathrm{L}^2(\mu)$,

$$f_m^2 \le \frac{1}{\lambda}\frac{f_m^2(\mathrm{Id}-P)V}{V} + \frac{b}{\lambda}\frac{f_m^2}{V}\mathbf{1}_C.$$

We observe also that $PV/V\le 1+b<\infty$. Hence, using the variational characterization of the mean, key Lemma 71 (noting that $\sup_{x\in\mathsf{E}} PV/V(x)<\infty$) and (21),

$$\begin{aligned}\|f\|_2^2 &\le \|f_m\|_2^2\\ &\le \frac{1}{\lambda}\left\langle f_m^2, 1-PV/V\right\rangle + \frac{b}{\lambda}\left\langle f_m^2,\mathbf{1}_C\right\rangle\\ &\le \frac{1+Kb}{\lambda}\left\langle f,(\mathrm{Id}-P)f\right\rangle,\end{aligned}$$

and we conclude. □

The proof relies on the important lemma.

Lemma 71 ([41])**.** *Let P be $\mu-$reversible, $V\colon \mathsf{E}\to[1,\infty)$ such that $\|PV/V\|_\infty<\infty$. Then for any $f\in \mathrm{L}^2(\mu)$, $m\in\mathbb{R}$ we have*

$$\left\langle (f-m)^2, 1-PV/V\right\rangle \le \left\langle f, (\mathrm{Id}-P)\, f\right\rangle \,.$$

Proof. We have for any $g\in \mathrm{L}^2(\mu)$,

$$\begin{aligned}0\leqslant \frac{1}{2}\cdot\int \mu\left(\mathrm{d}x\right)\cdot P\left(x,\mathrm{d}y\right)\cdot V\left(x\right)\cdot V\left(y\right)\cdot\left(\frac{g\left(y\right)}{V\left(y\right)}-\frac{g\left(x\right)}{V\left(x\right)}\right)^2\\ =\left\langle g^2, PV/V\right\rangle-\left\langle g, Pg\right\rangle\\ =\left\langle g,g\right\rangle-\left\langle g^2,1-PV/V\right\rangle-\left\langle g,g\right\rangle+\left\langle g,(\mathrm{Id}-P)g\right\rangle\\ =\left\langle g,(\mathrm{Id}-P)g\right\rangle-\left\langle g^2,1-PV/V\right\rangle.\end{aligned}$$

Further we notice that for any $f\in \mathrm{L}^2(\mu)$ and $m\in\mathbb{R}$,

$$\begin{aligned}\left\langle f-m,(\mathrm{Id}-P)\,(f-m)\right\rangle &=\left\langle f-m,(\mathrm{Id}-P)\,f\right\rangle\\ &=\left\langle f,(\mathrm{Id}-P)\,f\right\rangle-m\mu\big((\mathrm{Id}-P)f\big)\\ &=\left\langle f,(\mathrm{Id}-P)\,f\right\rangle\,,\end{aligned}$$

and we conclude. □

4.3.2 The subgeometric scenario

The following is a useful, simple result, which is related to [32, Theorem 14.3.7] and [12, Proposition 4.3.2] but with slightly different conditions and conclusions.

Lemma 72. *Let X be a Markov chain with Markov operator P and unique invariant probability measure μ. Suppose V, f and $\mathfrak{s}$ are nonnegative, finite-valued functions on E such that*

$$PV\le V-f+\mathfrak{s}.$$

Then $\mu(f)\le\mu(\mathfrak{s})$, whether or not $\mu(f)=\infty$.

Proof. We have

$$\begin{aligned}0&\le P^nV(x)\\ &=V(x)+\sum_{i=1}^n\mathbb{E}_x\left[V(X_i)-V(X_{i-1})\right]\\ &=V(x)+\sum_{i=1}^n\mathbb{E}_x\left[PV(X_{i-1})-V(X_{i-1})\right]\\ &\le V(x)+\sum_{i=1}^n\mathbb{E}_x\left[\mathfrak{s}(X_{i-1})-f(X_{i-1})\right],\end{aligned}$$

and hence we find

$$\mathbb{E}_x\left[\sum_{k=0}^{n-1} f(X_k)\right] \le V(x) + \mathbb{E}_x\left[\sum_{k=0}^{n-1} \mathfrak{s}(X_k)\right].$$

Since P has a unique invariant probability measure, we may apply Birkhoff's ergodic theorem; see, e.g., [12, Theorems 5.2.6 and 5.2.1]. First suppose $g \ge 0$ is such that $\mu(g) = \infty$. Then by the ergodic theorem, for μ-almost all x and any $m \ge 0$,

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}_x\left[\sum_{k=0}^{n-1} m \wedge g(X_k)\right] = \mu(m \wedge g),$$

and taking $m \to \infty$ we obtain $\lim_{n\to\infty} \frac{1}{n}\mathbb{E}_x\left[\sum_{k=0}^{n-1} g(X_k)\right] = \infty = \mu(g)$. Next, suppose $g \ge 0$ with $\mu(g) < \infty$. Then by the ergodic theorem, for μ-almost all x,

$$\mu(g) = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}_x\left[\sum_{k=0}^{n-1} g(X_k)\right].$$

Hence, for μ-almost all (and therefore some) x,

$$\mu(f) = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}_x\left[\sum_{k=0}^{n-1} f(X_k)\right] \le \lim_{n\to\infty} \frac{1}{n}\left\{V(x) + \mathbb{E}_x\left[\sum_{k=0}^{n-1} \mathfrak{s}(X_k)\right]\right\} = \mu(\mathfrak{s}),$$

and so we may conclude that $\mu(f) \le \mu(\mathfrak{s})$. ☐

Theorem 73. *Let P be a $\mu-$reversible Markov kernel, μ its unique invariant probability measure, such that:*

a). there exists a set $C \in \mathscr{E}$, a function $V \colon \mathsf{E} \to [1,\infty)$ and $b > 0$ such that

$$PV \le V - \phi \circ V + b\mathbf{1}_C\,,$$

where ϕ: $[1,\infty) \to (0,\infty)$ is a concave, continuous and increasing function;

b). a local Poincaré inequality holds: there exists $K > 0$ such that for any $f \in \mathrm{L}^2(\mu)$,

$$\|f_m \mathbf{1}_C\|_2^2 \le K\mathcal{E}(P,f)\,, \tag{22}$$

with $m = \mu(f \cdot \mathbf{1}_C)/\mu(C)$ and $f_m := f - m$.

Then for any $f \in \mathrm{L}^2_0(\mu)$ and $s > 0$,

$$\|f\|_2^2 \le s\mathcal{E}(P,f) + \beta(s)\|f\|_{\mathrm{osc}}^2,$$

where

$$\beta(s) := \frac{b\mu(C)}{\phi \circ (\mathrm{Id}/\phi)^{-1}\big(s/(1+Kb)\big)}\,.$$

Proof. From $PV \le V - \phi \circ V + b\mathbf{1}_C$ we obtain $\phi \circ V \le (\mathrm{Id} - P)V + b\mathbf{1}_C$ and for $f \in \mathrm{L}^2(\mu)$ we obtain

$$f_m^2 \le \frac{f_m^2(\mathrm{Id} - P)V}{\phi \circ V} + b\frac{f_m^2}{\phi \circ V}\mathbf{1}_C.$$

Now with $A(s) := \{x \in \mathsf{E}\colon s\,\phi \circ V(x) \ge V(x)\}$ for $s > 0$ we have

$$\begin{aligned} f_m^2\mathbf{1}_{A(s)} &\le \frac{f_m^2(\mathrm{Id} - P)V}{\phi \circ V}\mathbf{1}_{A(s)} + b\frac{f_m^2}{\phi \circ V}\mathbf{1}_{C\cap A(s)}\,, \\ &\le s\frac{f_m^2(\mathrm{Id} - P)V}{V}\mathbf{1}_{A(s)} + sb\frac{f_m^2}{V}\mathbf{1}_{C\cap A(s)}\,. \end{aligned}$$

Hence for $s > 0$, we have

$$\begin{aligned} f_m^2 &= f_m^2\mathbf{1}_{A(s)} + f_m^2\mathbf{1}_{A^\complement(s)} \\ &\le s\frac{f_m^2(\mathrm{Id} - P)V}{V} + sb\frac{f_m^2}{V}\mathbf{1}_C + f_m^2\mathbf{1}_{A^\complement(s)}. \end{aligned}$$

We observe also that $PV/V \le 1 + b < \infty$. Consequently, we can take expectations with respect to μ, yielding

$$\begin{aligned} \|f_m\|_2^2 &\le s\left\langle f_m^2, 1 - PV/V\right\rangle + sb\left\|f_m\mathbf{1}_C\right\|_2^2 + \|f_m\|_\infty^2\,\mu(A^\complement(s)) \\ &\le s\mathcal{E}(P, f) + sb\left\|f_m\mathbf{1}_C\right\|_2^2 + \|f\|_{\mathrm{osc}}^2\,\mu(A^\complement(s)), \end{aligned}$$

where we have used $V \ge 1$, Lemma 71 and $\|f_m\|_\infty^2 \le \|f\|_{\mathrm{osc}}^2$ since $\operatorname{ess\,inf}_x |f(x)| \le m \le \operatorname{ess\,sup}_x |f(x)|$. Since ϕ is concave, increasing and continuous, the function $x \mapsto x/\phi(x)$ is increasing and continuous and therefore invertible, and we can write $A^\complement(s) = \{x \in \mathsf{E}\colon \phi \circ (\mathrm{Id}/\phi)^{-1}(s) < \phi \circ V(x)\}$, and therefore

$$\begin{aligned} \mu(A^\complement(s)) &= \mu\left(\phi \circ V > \phi \circ (\mathrm{Id}/\phi)^{-1}(s)\right) \\ &\le \mu\left(\phi \circ V \ge \phi \circ (\mathrm{Id}/\phi)^{-1}(s)\right) \\ &\le \frac{\mu(\phi \circ V)}{\phi \circ (\mathrm{Id}/\phi)^{-1}(s)}, \end{aligned}$$

where $\mu(\phi \circ V) \le b\mu(C)$ by Lemma 72. Using $f \in \mathrm{L}_0^2(\mu)$ and (22),

$$\|f\|_2^2 \le \|f_m\|_2^2 \le (1 + Kb)s\mathcal{E}(P, f) + \frac{b\mu(C)}{\phi \circ (\mathrm{Id}/\phi)^{-1}(s)}\|f\|_{\mathrm{osc}}^2.$$

and we conclude. □

Remark 74. The assumption that P is reversible can be relaxed to some extent. If

$$\frac{1}{2}(P + P^*)V \le V - \phi \circ V + b\mathbf{1}_C\,,$$

then the conclusion also holds for nonreversible P. In particular, this condition allows for the use of Lemma 71, which is the only part of the proof utilizing reversibility.

As pointed out by [41], a standard minorization condition yields a local PI 19.

Lemma 75 ([41, Equation (8)]). *Let P be a $\mu-$invariant Markov kernel satisfying*

$$P(x,A) \geq \epsilon\nu(A)\mathbf{1}\{x \in C\}.$$

Then with $m = \mu(f\mathbf{1}_C)/\mu(C)$,

$$\|f_m\mathbf{1}_C\|_2^2 \leq \frac{2}{\epsilon}\mathcal{E}(P,f).$$

While this provides a relatively straightforward route to establishing a local PI, such an approach may not be sufficiently precise when one is interested in quantitative estimates. In Lemma 69 the minorization condition is replaced with a local PI on C, but P is assumed $\mu-$reversible. This mirrors [3, Proof of Theorem 1.4].

Example 76. When $\phi(v) = cv^\alpha$ for $\alpha \in [0,1)$, we obtain $(\mathrm{Id}/\phi)^{-1}(s) = (cs)^{1/(1-\alpha)}$ and therefore $\phi \circ (\mathrm{Id}/\phi)^{-1}(s) = c(cs)^{\alpha/(1-\alpha)}$. We conclude that $\beta(s) \propto s^{-\alpha/(1-\alpha)}$, and thereby obtain $\gamma(n) \propto n^{-\alpha/(1-\alpha)}$. Drift and minorisation techniques directly lead to a total variation rate of $n^{-\alpha/(1-\alpha)}$, which we do not recover since [1, Remark 12] gives a total variation rate of $\gamma^{1/2}(n)$. On the other hand, Proposition 114 implies a CLT for bounded functions if $\alpha/(1-\alpha) > 1$, i.e. $\alpha > 1/2$. This improves upon the condition $\alpha \geq 2/3$ in [20, Theorem 4.2] and is close to the condition $\alpha \geq 1/2$ obtained when the existence of an atom is assumed [20, Theorem 4.4]. We can straightforwardly obtain rates of convergence and CLTs for functions in $\mathrm{L}_0^p(\mu)$, for $p > 2$ using Proposition 13 and Remark 115, which may be more convenient than considering functions dominated by a power of the Lyapunov function V.

Example 77. If $\phi(v) = cv/\log(v)^\alpha$ then

$$\frac{v}{\phi(v)} = c^{-1}\log(v)^\alpha = s \iff v = \exp\big((cs)^{1/\alpha}\big)$$

and therefore

$$\beta(s) \propto s\exp\big(-(cs)^{1/\alpha}\big) \leq \exp\big(-(c's)^{1/\alpha}\big)$$

which leads to a rate of convergence

$$C'\exp\Big(-\{C(1+\alpha)n\}^{1/(1+\alpha)}\Big)$$

which is similar to what is obtained by [12].

4.3.3 Local Poincaré and isoperimetric inequalities

We use some general results, largely inspired by their recent use in [13].

Lemma 78 ([10, Theorem 4.2], Isoperimetric inequality)**.** *Let μ be a probability measure on $\mathsf{E} \subset \mathbb{R}^d$, whose density $\mu(x) \propto \exp(-U(x))$ w.r.t. Lebesgue is m-strongly log-concave, i.e.*

$$U(x+z) - U(x) - \langle \nabla U(x), z \rangle \geqslant \frac{m}{2} |z|^2 .$$

Then for any (nonempty) $S_1, S_2, S_3 \subset \mathsf{E}$ defining a partition of E we have

$$\mu(S_3) \geqslant \log 2 \cdot \sqrt{m} \cdot d(S_1, S_2) \cdot \mu(S_1) \cdot \mu(S_2) ,$$

where $d(S_1, S_2) := \inf \{|z - z'| : (z, z') \in S_1 \times S_2\}$.

Remark 79. In the original result of [10, Theorem 4.2], the hypothesis on μ is formulated in terms of the log-concavity of the Radon–Nikodym derivative of μ with respect to an appropriate Gaussian measure. We have rephrased the result slightly to emphasize the relationship with the strong convexity of the potential, which is consistent with the presentation of [13, Section 5.4].

Theorem 80 ([28, 6, 13])**.** *Let μ be a probability measure on $\mathsf{E} \subset \mathbb{R}^d$, whose density w.r.t. Lebesgue is m-strongly log-concave, and $C \subseteq \mathsf{E}$ be a convex set. Let P be a $\mu-$invariant Markov kernel and assume that there exist $\delta, \epsilon > 0$ such that for $z, z' \in C$, $|z - z'| \leqslant \delta$ implies*

$$\|P(z, \cdot) - P(z', \cdot)\|_{\mathrm{TV}} < 1 - \varepsilon .$$

Then for any $A \in \mathscr{E}$,

$$\mu \otimes P\big(A \times A^{\complement}\big) \geqslant \frac{\varepsilon}{4} \min \left\{1, \frac{\log 2}{8} \delta \sqrt{m}\right\} \min \left\{\mu(A \cap C), \mu\left(A^{\complement} \cap C\right)\right\} ,$$

and

$$\mu \otimes P\big(A \times A^{\complement}\big) \geqslant \frac{1}{\mu(C)} \frac{\varepsilon}{4} \min \left\{1, \frac{\log 2}{4} \delta \sqrt{m}\right\} \mu(A \cap C) \, \mu\left(A^{\complement} \cap C\right) .$$

Proof. Let $\delta, \epsilon > 0$ be as above. For $A \in \mathscr{E}$ define the sets

$$S_1 := \left\{z \in A \cap C \colon P\left(z, A^{\complement}\right) < \varepsilon/2\right\}$$
$$S_2 := \left\{z \in A^{\complement} \cap C \colon P(z, A) < \varepsilon/2\right\}$$

and $S_3 := C \cap \big(S_1 \cup S_2\big)^{\complement}$. We consider two cases. First we establish that when either $\mu(S_1) \leqslant \frac{1}{2} \mu(A \cap C)$ or $\mu(S_2) \leqslant \frac{1}{2} \mu\left(A^{\complement} \cap C\right)$, then

$$\mu \otimes P\big(A \times A^{\complement}\big) \geqslant \frac{1}{4} \cdot \varepsilon \cdot \min \left\{\mu(A \cap C), \mu\big(A^{\complement} \cap C\big)\right\} .$$

If $\mu(S_1) \leqslant \frac{1}{2}\mu(A \cap C)$ then

$$\begin{aligned}\mu(A \cap C) &= \mu(S_1) + \mu((A \cap C) \setminus S_1) \\ &\leqslant \frac{1}{2}\mu(A \cap C) + \mu((A \cap C) \setminus S_1)\ ,\end{aligned}$$

that is $\frac{1}{2}\mu(A \cap C) \leqslant \mu((A \cap C) \setminus S_1)$. Now,

$$\begin{aligned}\mu \otimes P\big(A \times A^{\complement}\big) &\geqslant \mu \otimes P\left(((A \cap C) \setminus S_1) \times A^{\complement}\right) \\ &\geqslant \frac{1}{2} \cdot \varepsilon \cdot \mu((A \cap C) \setminus S_1) \\ &\geqslant \frac{1}{4} \cdot \varepsilon \cdot \mu(A \cap C)\,.\end{aligned}$$

Similarly if $\mu(S_2) \leqslant \frac{1}{2}\mu\left(A^{\complement} \cap C\right)$ then

$$\begin{aligned}\mu\big(A^{\complement} \cap C\big) &= \mu\big(S_2\big) + \mu\left(\big(A^{\complement} \cap C\big) \setminus S_2\right). \\ &\leqslant \frac{1}{2}\mu\big(A^{\complement} \cap C\big) + \mu\left(\big(A^{\complement} \cap C\big) \setminus S_2\right)\end{aligned}$$

that is $\frac{1}{2}\mu\big(A^{\complement} \cap C\big) \leqslant \mu\left(\big(A^{\complement} \cap C\big) \setminus S_2\right)$ and arguing as before:

$$\begin{aligned}\mu \otimes P\big(A^{\complement} \times A\big) &\geqslant \mu \otimes P\left(\left(\big(A^{\complement} \cap C\big) \setminus S_2\right) \times A\right) \\ &\geqslant \frac{1}{2} \cdot \varepsilon \cdot \mu\left(\big(A^{\complement} \cap C\big) \setminus S_2\right) \\ &\geqslant \frac{1}{4} \cdot \varepsilon \cdot \mu\left(A^{\complement} \cap C\right).\end{aligned}$$

As noticed by [13], reversibility is not required to establish the following

$$\begin{aligned}\mu \otimes P\left(A \times A^{\complement}\right) &= \mu \otimes P\left(\mathsf{X} \times A^{\complement}\right) - \left[\mu \otimes P\left(A^{\complement} \times \mathsf{X}\right) - \mu \otimes P\left(A^{\complement} \times A\right)\right] \\ &= \mu\left(A^{\complement}\right) - \mu\left(A^{\complement}\right) + \mu \otimes P\left(A^{\complement} \times A\right) \\ &= \mu \otimes P\left(A^{\complement} \times A\right)\ ,\end{aligned}$$

and this allows us to establish our first claim. Using the fact that for $B \in \mathscr{E}$,

$$1 \geq \frac{\mu(B \cap C)}{\mu(C)},$$

we may also deduce that if $\mu(S_1) \leqslant \frac{1}{2}\mu(A \cap C)$ or $\mu(S_2) \leqslant \frac{1}{2}\mu\left(A^{\complement} \cap C\right)$ then

$$\mu \otimes P\big(A \times A^{\complement}\big) \geqslant \frac{1}{4\mu(C)} \cdot \varepsilon \cdot \mu(A \cap C) \cdot \mu\big(A^{\complement} \cap C\big)\,.$$

In the second case, $\mu(S_1) > \frac{1}{2}\mu(A\cap C)$ and $\mu\big(S_2\big) > \frac{1}{2}\mu\left(A^{\complement}\cap C\right)$. We then compute

$$\begin{aligned}
\mu\otimes P\left(A\times A^{\complement}\right) &= \frac{1}{2}\mu\otimes P\left(A\times A^{\complement}\right)+\frac{1}{2}\mu\otimes P\left(A^{\complement}\times A\right)\\
&\geqslant \frac{1}{2}\mu\otimes P\left(\left(A\cap C\cap S_1^{\complement}\right)\times A^{\complement}\right)+\frac{1}{2}\mu\otimes P\left(\left(A^{\complement}\cap C\cap S_2^{\complement}\right)\times A\right)\\
&\geqslant \frac{1}{4}\cdot\varepsilon\cdot\mu\left(A\cap C\cap S_1^{\complement}\right)+\frac{1}{4}\cdot\varepsilon\cdot\mu\left(A^{\complement}\cap C\cap S_2^{\complement}\right)\\
&= \frac{1}{4}\cdot\varepsilon\cdot\mu\left(C\cap(S_1\cup S_2)^{\complement}\right)\\
&= \frac{1}{4}\cdot\varepsilon\cdot\mu(S_3)\,.
\end{aligned}$$

Now for $(z,z')\in S_1\times S_2$ we have

$$\begin{aligned}
\|P(z,\cdot)-P(z',\cdot)\|_{\mathrm{TV}} &\geqslant P(z,A)-P(z',A)\\
&= 1-P\left(z,A^{\complement}\right)-P(z',A)\\
&\geqslant 1-\varepsilon.
\end{aligned}$$

This implies that $d(S_1,S_2) = \inf\{|z-z'|:\,(z,z')\in S_1\times S_2\} > \delta$, since for $z,z'\in\mathsf{X}$, $|z-z'|\leqslant\delta$ implies

$$\|P(z,\cdot)-P(z',\cdot)\|_{\mathrm{TV}} < 1-\varepsilon.$$

From Lemma 78 applied to the measure $\mu_C(\cdot) := \mu(\cdot\cap C)/\mu(C)$, we can thus write that

$$\begin{aligned}
\mu(S_3) &\geqslant \mu(C)\cdot\log 2\cdot\sqrt{m}\cdot d(S_1,S_2)\cdot\frac{\mu(S_1)}{\mu(C)}\cdot\frac{\mu(S_2)}{\mu(C)}\\
&\geqslant \frac{\log 2}{\mu(C)}\cdot\sqrt{m}\cdot\delta\cdot\left(\frac{1}{2}\mu(A\cap C)\right)\cdot\left(\frac{1}{2}\mu\big(A^{\complement}\cap C\big)\right)
\end{aligned}$$

and consequently that

$$\mu\otimes P\left(A\times A^{\complement}\right)\geqslant \mu(C)^{-1}\cdot\frac{\log 2}{16}\cdot\varepsilon\cdot\sqrt{m}\cdot\delta\cdot\mu\big(A\cap C\big)\cdot\mu\big(A^{\complement}\cap C\big)\,.$$

The second result then follows. To obtain the first result, since $p\cdot(1-p)\geqslant\frac{1}{2}\cdot\min(p,1-p)$ for $0<p<1$, it holds that

$$\begin{aligned}
\mu(A\cap C)\cdot\mu\big(A^{\complement}\cap C\big) &= \mu(C)^2\cdot\left[\frac{\mu(A\cap C)}{\mu(C)}\right]\cdot\left[1-\frac{\mu(A\cap C)}{\mu(C)}\right]\\
&\geqslant \frac{1}{2}\cdot\mu(C)^2\cdot\min\left\{\frac{\mu(A\cap C)}{\mu(C)},\frac{\mu\left(A^{\complement}\cap C\right)}{\mu(C)}\right\}\\
&= \frac{1}{2}\cdot\mu(C)\cdot\min\left\{\mu(A\cap C),\mu\left(A^{\complement}\cap C\right)\right\}.
\end{aligned}$$

$\square$

Corollary 81. *Under the conditions of Theorem 80, we can deduce that if P is also reversible then a restricted Poincaré inequality holds for P on C, and a local Poincaré inequality holds for P on C.*

Proof. We obtain from the conclusion of Theorem 80 that for any $A \in \mathcal{E}$ with $A \subseteq C$,

$$\mu_C \otimes P_C(A \times A^{\complement}) \geq \frac{\varepsilon}{4} \min\left\{1, \frac{\log 2}{8}\delta\sqrt{m}\right\} \min\left\{\mu_C(A), \mu_C\left(A^{\complement}\right)\right\},$$

and

$$\mu_C \otimes P_C\left(A \times A^{\complement}\right) \geqslant \frac{\varepsilon}{4} \min\left\{1, \frac{\log 2}{4}\delta\sqrt{m}\right\} \mu_C(A)\, \mu_C\left(A^{\complement}\right).$$

Since P is reversible, (11) implies that P_C admits a strong Poincaré inequality. Lemma 69 then implies that a local Poincaré holds for P on C. $\square$

4.4 Restricted Markov chains and vanishing Poincaré constants

In this subsection, we establish a link between the existence of SPIs for restrictions of a Markov chain P to suitable sets and WPIs for the unrestricted chain. Roughly speaking, for subgeometric chains, it is possible that the restriction of the chain to a 'nice' set A exhibits a strong Poincaré inequality, but as $\mu(A)$ grows, the constant in this inequality necessarily degenerates. We will show that the rate at which this constant degenerates as A grows allows one to deduce a quantitative weak Poincaré inequality for P. In what follows, we let $\Phi = \|\cdot\|_{\mathrm{osc}}^2$. Note that $\mathrm{var}_\mu(f) \leq \Phi(f)$.

In the following result we upper and lower bound $\mathcal{E}(P, f)$ by quantities involving Dirichlet forms associated to the restriction of P to a set A.

Lemma 82. *Let P be μ-reversible. Let $A \in \mathscr{E}_+$ and P_A be the μ_A-reversible restriction of P to A. Then*

$$\mu(A)\mathcal{E}(P_A, f) \leq \mathcal{E}(P, f) \leq \mu(A)\mathcal{E}(P_A, f) + \mu(A^{\complement})\Phi(f).$$

Proof. We have

$$\begin{aligned}
\mathcal{E}(P, f) &= \frac{1}{2}\int \mu(\mathrm{d}x)P(x, \mathrm{d}y)\left\{f(x) - f(y)\right\}^2 \\
&\geq \frac{1}{2}\mu(A)\int \mu_A(\mathrm{d}x)P_A(x, \mathrm{d}y)\left\{f(x) - f(y)\right\}^2 \\
&= \mu(A)\mathcal{E}(P_A, f),
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{E}(P,f) &= \frac{1}{2}\int \mu(\mathrm{d}x)P(x,\mathrm{d}y)\left\{f(x)-f(y)\right\}^2 \\
&= \frac{1}{2}\int_{A\times A} \mu(\mathrm{d}x)P(x,\mathrm{d}y)\left\{f(x)-f(y)\right\}^2 \\
&\qquad\qquad + \frac{1}{2}\int_{(A\times A)^{\complement}} \mu(\mathrm{d}x)P(x,\mathrm{d}y)\left\{f(x)-f(y)\right\}^2 \\
&\le \frac{\mu(A)}{2}\int \mu_A(\mathrm{d}x)P_A(x,\mathrm{d}y)\left\{f(x)-f(y)\right\}^2 \\
&\qquad\qquad + \frac{1}{2}\Phi(f)\left\{\mu\otimes P(A^{\complement}\times \mathsf{E}) + \mu\otimes P(A\times A^{\complement})\right\} \\
&\le \mu(A)\mathcal{E}(P_A,f) + \mu(A^{\complement})\Phi(f),
\end{aligned}$$

where we have used the fact that since P is μ-invariant,

$$\mu\otimes P(A\times A^{\complement}) \le \mu\otimes P(E\times A^{\complement}) = \mu(A^{\complement}).$$

□

Now let Π be the Markov kernel such that $\Pi(x,\cdot) = \mu(\cdot)$ for all $x \in \mathsf{E}$, and let Π_A be the corresponding restriction as in Definition 68: for $B \in \mathscr{E}$, $\Pi_A(x,B) = \mu(A\cap B) + \mu(A^{\complement})\mathbf{1}_B(x)$, which is not necessarily equal to $\mu_A(B) = \mu(A)^{-1}\mu(A\cap B)$. In fact, we have the following.

Lemma 83. *For $A \in \mathscr{E}_+$, $\mathcal{E}(\Pi_A, f) = \mu(A)\mathrm{var}_{\mu_A}(f)$.*

Proof. We have

$$\begin{aligned}
\mathcal{E}(\Pi_A,f) &= \frac{1}{2}\int \mu_A(\mathrm{d}x)\Pi_A(x,\mathrm{d}y)\left\{f(x)-f(y)\right\}^2 \\
&= \frac{1}{2}\int \mu_A(\mathrm{d}x)\Pi(x,\mathrm{d}y)\mathbf{1}_A(y)\left\{f(x)-f(y)\right\}^2 \\
&= \frac{1}{2}\mu(A)\int \mu_A(\mathrm{d}x)\mu_A(\mathrm{d}y)\left\{f(x)-f(y)\right\}^2 \\
&= \mu(A)\mathrm{var}_{\mu}(f).
\end{aligned}$$

□

Corollary 84. *Letting $P = \Pi$ in Lemma 82, we obtain*

$$\mu(A)^2\mathrm{var}_{\mu_A}(f) \le \mathrm{var}_{\mu}(f) \le \mu(A)^2\mathrm{var}_{\mu_A}(f) + \mu(A^{\complement})\Phi(f), \qquad A \in \mathscr{E}_+.$$

Theorem 85. *Let P be μ-reversible. For $A \in \mathscr{E}_+$, define $\gamma_P(A)$ to be the (right) "spectral gap"*

$$\gamma_P(A) = \inf_{f\in \mathrm{L}^2(\mu)} \frac{\mathcal{E}(P_A,f)}{\mathrm{var}_{\mu_A}(f)}.$$

Then a P satisfies a (Φ, β)-WPI with

$$\beta(s) = 1 \wedge \inf_{A \in \mathscr{E}_+} \left\{ \mu(A^\complement) : \gamma_P(A) \geq \frac{\mu(A)}{s} \right\}.$$

Proof. Let $s > 0$. If $\mathcal{S} = \{A \in \mathscr{E}_+ : \gamma_P(A) \geq \mu(A)/s\}$ is empty, we may take $\beta(s) = 1$ since $\mathrm{var}_\mu(f) \leq \Phi(f)$. Otherwise, let $A \in \mathcal{S}$. By Corollary 84 and Lemma 82, we obtain that for any $f \in \mathrm{L}^2(\mu)$,

$$\begin{aligned}
\mathrm{var}_\mu(f) &\leq \mu(A)^2 \mathrm{var}_{\mu_A}(f) + \mu(A^\complement)\Phi(f) \\
&\leq \frac{\mu(A)^2}{\gamma_P(A)} \mathcal{E}(P_A, f) + \mu(A^\complement)\Phi(f) \\
&\leq \frac{\mu(A)}{\gamma_P(A)} \mathcal{E}(P, f) + \mu(A^\complement)\Phi(f) \\
&\leq s\mathcal{E}(P, f) + \mu(A^\complement)\Phi(f),
\end{aligned}$$

from which we may deduce that one may take $\beta(s) = \mu(A^\complement)$. The result then follows by taking the infimum over $A \in \mathcal{S}$. ☐

We may revisit the WPI obtained for the IMH in [1] from this perspective as follows; the argument is essentially the same.

Example 86. Consider the IMH with target π and proposal q, and let $w = \mathrm{d}\pi/\mathrm{d}q$. If we define $A = \{x : w(x) \leq s\}$ then we may write

$$\begin{aligned}
\mathcal{E}(P_A, f) &= \frac{1}{2} \int \pi_A(\mathrm{d}x) P_A(x, \mathrm{d}y) \{f(x) - f(y)\}^2 \\
&= \frac{1}{2} \int \pi_A(\mathrm{d}x) q(\mathrm{d}y) \left\{ 1 \wedge \frac{w(y)}{w(x)} \right\} \mathbf{1}_A(y) \{f(x) - f(y)\}^2 \\
&\geq \frac{1}{2s} \pi(A) \int \pi_A(\mathrm{d}x) \pi_A(\mathrm{d}y) \{f(x) - f(y)\}^2 \\
&= \frac{\pi(A)}{s} \mathrm{var}_{\pi_A}(f).
\end{aligned}$$

It follows that $\gamma_P(A) \geq \pi(A)/s$, and so by Theorem 85 we may take $\beta(s) = \pi(A^\complement)$ in a $(\|\cdot\|_{\mathrm{osc}}^2, \beta)$-WPI. This argument is clearly related to the well-known fact that the IMH has a spectral gap if and only if w is upper bounded by a finite constant [31, Theorem 2.1], and we obtain the subgeometric rate here by considering the measures of a sequence of sets on which w is upper bounded by an increasing sequence of constants.

Remark 87. One may equivalently deduce a (Φ, α)-WPI with

$$\alpha(r) = \inf_{A \in \mathscr{E}_+} \left\{ \frac{\mu(A)}{\gamma_P(A)} : \mu(A) \geq 1 - r \right\}.$$

If we define $\gamma_P(t) = \sup_{A \in \mathscr{E}_+} \{\gamma_P(A) : \mu(A) \geq 1 - t\}$ then we see that $\alpha(r) \leq 1/\gamma_P(r)$ and the rate at which $\gamma_P(r) \to 0$ as $r \to 0$ provides an upper bound on the convergence rate.

To our knowledge, the observation that a subgeometric rate of convergence can be related to the rate of decay of the spectral gap on an appropriate sequence of sets is novel. Considering restrictions of μ and P to a set A is reminiscent of the notion of spectral profile introduced by [17], which involves instead considering $\mathcal{E}(P,f)/\mathrm{var}_{\mu}(f)$ when $f \geq 0$ has support restricted to appropriately chosen sets (S_t), and considering the decay as $\mu(S_t) \to 1$. However, it is not clear how to relate the two concepts, and we note that the spectral profile was introduced to obtain bounds on mixing times whereas we are interested here in subgeometric rates of convergence.

Remark 88. Clearly if P has a (right) spectral gap $\gamma_P = \gamma_P(\mathsf{E}) > 0$ then we have $\beta(s) = 0$ for $s \geq \gamma_P^{-1}$.

For some Markov kernels P with state space $\mathsf{E} = \mathbb{R}^d$, the restriction of P to a ball around the origin will have a non-zero right spectral gap. In such cases, the sequence of balls with increasing radius defines a sequence of restrictions and the rate at which the gap decreases together with the rate at which the μ-measure of the balls tends to 1 can be used to deduce a WPI.

Example 89. Assume that for a Markov kernel P there is a family of sets $(A_t)_{t\geq 1}$ constants $C, a, b > 0$ such that for all $t \geq 1$,

$$\gamma_P(A_t) \geq C t^{-a}, \qquad \mu(A_t^{\complement}) \leq D t^{-b}.$$

Then we find that for $\gamma_P(A_t) \geq \frac{1}{s}$ is satisfied by taking $t = (Cs)^{\frac{1}{a}}$, and we then find $\mu(A_t^{\complement}) \leq D(Cs)^{-\frac{b}{a}}$. Hence P satisfies a (Φ, β)-WPI with $\beta(s) = D(Cs)^{-\frac{b}{a}}$. This argument may be valid when P is a random-walk Metropolis kernel on a heavy-tailed target, and A_t is a ball of radius t around the origin, although proving rigorously the lower bounds on $\gamma_P(A_t)$ is not trivial.

5 Examples and applications

5.1 Lower bounds for pseudo-marginal MCMC

We consider a specific and theoretically tractable ABC example covered by positive results from [1]. We show now that there is a quantitative version of the argument in [26] that ABC with local proposals is subgeometric, and that the lower bound on the polynomial rate matches the upper bound given by [1].

In this subsection, we let $\tilde{P}$ be the pseudo-marginal Markov kernel, and in particular we focus on complementing the results in [1]. For any measurable A such that $(x,w) \notin A$, we may write

$$\tilde{P}(x,w;A) = \int q(x,\mathrm{d}y) Q_y(\mathrm{d}u) \left\{1 \wedge r(x,y)\frac{u}{w}\right\} \mathbf{1}_A(y,u),$$

where $\{Q_x : x \in \mathsf{E}\}$ is a family of probability measures such that $\int Q_x(\mathrm{d}w) w = 1$. We focus on the ABC example in [1, Section 4.3], with some prior ν and an approximate, intractable likelihood ℓ_{ABC}. In particular, for some $N \in \mathbb{N}$ and any $x \in \mathsf{E}$ we denote

$$Q_x(A) = Q_{x,N}(A) = \mathbb{P}_x\left(\frac{1}{N}\sum_{i=1}^{N} W_i \in A\right),$$

where under $\mathbb{P}_x$, $W_i = \frac{1}{\ell_{\mathrm{ABC}}(x)} B_i$ and $B_1, \ldots, B_N$ are independent Bernoulli($\ell_{\mathrm{ABC}}(x)$) random variables. The parameter N thereby controls the concentration of $W \sim Q_x$ around 1, and we use the subscript N to emphasize this dependence.

Proposition 90. *Consider the general ABC example in [1, Section 4.3], and take for $a, q \in (0,1)$,*

$$\nu(x) = (1-q)\, q^{x-1} \mathbf{1}_{\{1,2,\ldots\}}(x),$$
$$\ell_{\mathrm{ABC}}(x) = a^{x-1} \mathbf{1}_{\{1,2,\ldots\}}(x),$$

and $q(x, x-1) = q(x, x+1) = 1/2$. Then, for any $N \geqslant 1$, if $\tilde{P}$ admits a (Φ, β)-WPI then $\beta(s) \in \Omega\left(s^{-\frac{\log(aq)}{\log(a)}}\right)$.

Proof. The ABC posterior is $\pi_{\mathrm{ABC}}(x) = (1-aq)(aq)^{x-1}\, \mathbf{1}_{\{1,2,\ldots\}}(x)$, i.e. Geometric $(1-aq)$. We define the pseudo-marginal target distribution on (x,w) to be $\tilde{\pi}$. We may define the set, with $\rho \in \mathbb{N}$,

$$A_\rho = \{(x,w) : x > \rho\},$$

and we obtain $\tilde{\pi}(A_\rho) = \pi_{\mathrm{ABC}}(x > \rho) = (aq)^\rho$.

Let $u < aq/4$, and take $\rho = \lfloor \log(2u) / \log(aq) \rfloor$. Since $x - 1 \leq \lfloor x \rfloor \leq x$, we deduce that $2u \leq \tilde{\pi}(A_\rho) < 1/2$, and hence $\tilde{\pi}(A_\rho)\tilde{\pi}(A_\rho^{\complement}) > 2u \cdot \frac{1}{2} = u$. Now, we find that for $(x,w) \in A_\rho$, and any $N \in \mathbb{N}$, we have the bound

$$\begin{aligned}
\tilde{P}\left(x, w; A_\rho^{\complement}\right) &\leqslant \mathbf{1}_{\{\rho+1\}}(x) q(\rho+1, \rho) \int Q_{\rho,N}(\mathrm{d}u) \left\{1 \wedge \frac{\pi_{\mathrm{ABC}}(\rho)}{\pi_{\mathrm{ABC}}(\rho+1)} \cdot \frac{u}{w}\right\} \\
&\leqslant \mathbf{1}_{\{\rho+1\}}(x) \frac{1}{2} Q_{\rho,N}(u > 0) \\
&\leqslant \mathbf{1}_{\{\rho+1\}}(x) \frac{N}{2} a^{\rho-1},
\end{aligned}$$

where we have used Bernoulli's inequality to deduce that

$$\begin{aligned}
Q_{\rho,N}(u > 0) &= 1 - \left(1 - a^{\rho-1}\right)^N \\
&\leqslant 1 - \left(1 - N a^{\rho-1}\right) \\
&\leqslant N a^{\rho-1}.
\end{aligned}$$

Hence, we obtain that

$$\begin{aligned}
\tilde{\pi} \otimes \tilde{P}\left(A_\rho, A_\rho^{\complement}\right) &\leqslant \tilde{\pi}(A_\rho) \frac{N}{2} a^{\rho-1} \\
&\leqslant \tilde{\pi}(A_\rho) \frac{N}{2} a^{\frac{\log(2u)}{\log(aq)} - 2}.
\end{aligned}$$

It follows that the weak conductance satisfies

$$\begin{aligned}\kappa(u) &\leqslant \frac{\tilde{\pi}\otimes\tilde{P}\left(A_\rho, A_\rho^{\complement}\right)}{\tilde{\pi}\otimes\tilde{\pi}\left(A_\rho, A_\rho^{\complement}\right)}\\ &\leqslant Na^{\frac{\log(2u)}{\log(aq)}-2}\\ &= Na^{-2}\left(2u\right)^{\frac{\log(a)}{\log(aq)}}.\end{aligned}$$

and so we see that $\kappa(u) \in \mathcal{O}\left(u^{\frac{\log(a)}{\log(aq)}}\right)$. This then implies that $\alpha^\star(r) \in \Omega\left(r^{-\frac{\log(a)}{\log(aq)}}\right)$ as $r \downarrow 0$ and $\beta^\star(s) \in \Omega\left(s^{-\frac{\log(aq)}{\log(a)}}\right)$. $\square$

Remark 91. [1] considered the setting where the marginal chain is geometric with strong Poincaré constant C_{P}. They showed that one may take $\beta(s) = \beta'(C_{\mathrm{P}}s)/C_{\mathrm{P}}$ where $\beta'(s) = \tilde{\pi}(w \geqslant s)$. In the case $N = 1$, we see that

$$\begin{aligned}\tilde{\pi}(w \geqslant s) &= \tilde{\pi}\left(\left\{x : \frac{1}{a^{x-1}} \geqslant s\right\}\right)\\ &= \pi\left(\left\{x : x \geqslant \frac{\log(s)}{-\log(a)} + 1\right\}\right)\\ &= (aq)^{\left\lceil \frac{\log(s)}{-\log(a)}\right\rceil}\\ &\sim s^{-\frac{\log(aq)}{\log(a)}}.\end{aligned}$$

Alternatively, [1] showed that for any $N \in \mathbb{N}$ and $p \in \mathbb{N}$, $\beta(s) \in \mathcal{O}(s^{-p})$ if $\int \nu(\mathrm{d}x)\,\ell_{\mathrm{ABC}}(x)^{-(p-1)} < \infty$, which in this case corresponds to $q/a^{p-1} < 1$, or equivalently $p < \log(aq)/\log(q)$, matching the lower bound above.

5.2 Lower bounds for RWM targeting heavy-tailed distributions

In this subsection, we assume $|\cdot|$ is a norm. We consider P a μ-invariant kernel that is local in the sense that

$$b(r) := \inf_{x\in\mathsf{E}} P(x, \mathcal{B}(x,r)),$$

is a real-valued function with $\lim_{r\to\infty} b(r) = 1$. We assume in this subsection that $\Phi(\cdot) = \|\cdot\|_{\mathrm{osc}}^2$.

When μ has polynomial tails, we seek to demonstrate that arguments used to show that $\kappa(0) = 0$, and hence that P does not admit a spectral gap, may also be used to lower bound α or β in a WPI for P. In this sense, such arguments can be made quantitative, although we require more information on the measure of suitable sets to deduce rate information. The following argument is inspired by the approach taken in the proof of [35, Theorem 6.3].

The first lemma upper bounds $\mu\otimes P\left(A\times A^{\complement}\right)$.

Lemma 92. *Let*

$$\phi(\rho, K) := \frac{\mathbb{P}_{\mu}(|X| > \rho + K)}{\mathbb{P}_{\mu}(|X| > \rho)}.$$

Then with $A = \mathcal{B}(0,\rho)^{\complement}$ *and any* $K > 0$*,*

$$\mu \otimes P\left(A \times A^{\complement}\right) \leqslant \mu(A)\left\{1 - \phi(\rho, K) \cdot b(K)\right\}.$$

Proof. Let $(X, Y) \sim \mu_A \otimes P$, where μ_A is as defined in Definition 68, and consider the representation $Y = X + \xi_X$. We bound

$$\begin{aligned}
\mathbb{P}(|Y| > \rho) &\geqslant \mathbb{P}(|X| > \rho + K, |Y| > \rho) \\
&= \mathbb{P}(|X| > \rho + K, |X + \xi_X| > \rho) \\
&\geqslant \mathbb{P}(|X| > \rho + K, |\xi_X| \leqslant K) \\
&\geqslant \phi(\rho, K) \cdot b(K).
\end{aligned}$$

where we have used that the two conditions $|X| > \rho + K, |\xi_X| \leqslant K \implies \rho + K - K \leq |X| - |\xi_X| \leq |X + \xi_X|$ and the fact that $X \sim \mu_A$. It follows that

$$\int \mu_A(\mathrm{d}x)\, P(x, A) \geqslant \phi(\rho, K) \cdot b(K),$$

and hence

$$\begin{aligned}
\mu \otimes P\left(A \times A^{\complement}\right) &= \int_A \mu(\mathrm{d}x)\, P\left(x, A^{\complement}\right) \\
&= \mu(A) \int \mu_A(\mathrm{d}x)\, P\left(x, A^{\complement}\right) \\
&\leqslant \mu(A)\left\{1 - \phi(\rho, K)\, b(K)\right\}.
\end{aligned}$$

□

In the following, the μ considered is a multi-dimensional version of the stylized one-dimensional case considered in [22, Eq. 52]. Although the argument is likely to be useful in other cases, it is necessary to have fairly precise control on both $\mu(\mathcal{B}(0,\rho))$ and $\mu\left(\mathcal{B}(0,\rho)^{\complement}\right)$ in order to quantify how $\phi(\rho, K)$ tends to 1 as ρ and K increase.

Proposition 93. *Assume that for some* $t > 0$*,*

$$\mu\left(\mathcal{B}(0,\rho)^{\complement}\right) = \rho^{-t}, \qquad \rho \geq 1.$$

Assume there exist $D, \eta > 0$ *such that* P *satisfies*

$$b(K) \geqslant 1 - DK^{-\eta}, \qquad K > 0,$$

where $b(\cdot)$ *is as defined in Lemma 92. Then* $\beta^{\star}(s) \in \Omega\left(s^{-t\frac{\eta+1}{\eta}}\right)$.

Proof. Let $\rho_0 = 2^{1/t}$, which satisfies $\mu(\mathcal{B}(0,\rho_0)) = \frac{1}{2}$, from which we may deduce that $\mu(\mathcal{B}(0,\rho)) > 1/2$ for all $\rho > \rho_0$. This will be the smallest ρ which we consider, and it satisfies

$$\mu \otimes \mu \left(B_{\rho_0}(0)^{\complement} \times B_{\rho_0}(0) \right) = \frac{1}{2}\rho_0^{-t} = \frac{1}{4} =: u_0.$$

Given any $u < u_0$, we take $\rho = (2u)^{-\frac{1}{t}} > \rho_0$ and $A = \mathcal{B}(0,\rho)^{\complement}$, which satisfies $\mu\left(A^{\complement}\right) > 1/2$, and so it holds that

$$\mu \otimes \mu \left(A \times A^{\complement} \right) > \frac{1}{2}\rho^{-t} = u.$$

By Lemma 92, we obtain that for any $K > 0$,

$$\begin{aligned} \kappa(u) &\leqslant \frac{\mu \otimes P\left(A \times A^{\complement}\right)}{\mu \otimes \mu\left(A \times A^{\complement}\right)} \\ &= \frac{1 - \phi(\rho, K)\, b(K)}{\mu\left(A^{\complement}\right)} \\ &\leqslant 2\left\{1 - \phi(\rho, K)\, b(K)\right\}. \end{aligned}$$

Letting $v = 2u$, we thus find that

$$\begin{aligned} \phi(\rho, K) &= \frac{\mathbb{P}_\mu(|X| > \rho + K)}{\mathbb{P}_\mu(|X| > \rho)} \\ &= \frac{(\rho + K)^{-t}}{\rho^{-t}} \\ &= \frac{1}{v\left(v^{-\frac{1}{t}} + K\right)^t} \\ &= \frac{1}{\left(1 + v^{\frac{1}{t}} K\right)^t}. \end{aligned}$$

Hence, we have the bound

$$1 - \phi(\rho, K)\, b(K) \leqslant 1 - \frac{1 - DK^{-\eta}}{\left(1 + v^{\frac{1}{t}} K\right)^t},$$

and by taking $K = v^{-\frac{1}{t+\eta t}}$, we may deduce that

$$\lim_{v \downarrow 0} \left\{ v^{-\frac{1}{t}\cdot\frac{\eta}{\eta+1}} \left\{ 1 - \frac{1 - DK^{-\eta}}{\left(1 + v^{\frac{1}{t}} K\right)^t} \right\} \right\} = t + D,$$

from which we may conclude that $\kappa(u) \in \mathcal{O}\left(u^{\frac{1}{t}\cdot\frac{\eta}{\eta+1}}\right)$. Since $\alpha^{\star}(r) \geqslant \frac{1}{2\kappa(2r)}$ by Remark 39, we obtain $\alpha^{\star}(r) \in \Omega\left(r^{-\frac{1}{t}\cdot\frac{\eta}{\eta+1}}\right)$ as $r \downarrow 0$, and so $\beta^{\star}(s) \in \Omega\left(s^{-t\frac{\eta+1}{\eta}}\right)$. $\square$

Remark 94. If P is μ-reversible, one may then deduce that by Proposition 26 if $\epsilon > 0$ then $\|P^n f\|^2$ cannot be in $\mathcal{O}\left(n^{-t\frac{\eta+1}{\eta}-\epsilon}\right)$ for all $f \in \mathrm{L}_0^2(\mu)$ with $\Phi(f) < \infty$. We see that, similar to [21] and [22], the lower bounds suggest that faster rates are possible if η is close to 0, i.e. $P(x,\cdot)$ is heavy-tailed for all x.

5.3 Spectral gap of the RWM in high-dimensions

We let $\mathsf{X} = \mathsf{Z} = \mathbb{R}^d$ throughout. Let P be the Markov transition probability of the Random Walk Metropolis (RWM) with Gaussian proposal, defined for any $(x, A) \in \mathsf{X} \times \mathscr{X}$

$$Q_x(A) = \int \mathbf{1}_A(x + d^{-1/2}z)\, Q(\mathrm{d}z),$$

where $Q = \mathcal{N}\left(0, \sigma^2 \mathrm{Id}\right)$. Then, for any $(x, A) \in \mathsf{X} \times \mathscr{X}$,

$$P(x, A) = \int_A \alpha\left(x, d^{-1/2}z\right) Q(\mathrm{d}z) + \mathbf{1}_A(x)\left[1 - \alpha(x)\right],$$

with for any $(x, z) \in \mathsf{X} \times \mathsf{Z}$, $\alpha(x, z) := \min\{1, \mathsf{r}(x, z)\}$ and

$$\mathsf{r}(x, z) := \frac{\pi(x+z)}{\pi(x)},$$
$$\alpha(x) := \int \alpha\left(x, d^{-1/2}z\right) Q(\mathrm{d}z), \tag{23}$$

and $\pi : \mathsf{X} \to [0, \infty)$ is a target density with respect to Lebesgue measure with $\pi(x) \propto \exp(-U(x))$. In this section and in Section 5.4, we denote by $|\cdot|$ the Euclidean norm in $\mathbb{R}^d$, i.e. $|x| = \left(\sum_{i=1}^d x_i^2\right)^{1/2}$.

Assumption 95. *We assume the following properties of our target distribution:*

a). U is spherically symmetric with $U(x) = u(|x|^2)$, for some increasing function $u : [0, \infty) \to [0, \infty)$. In particular, U attains its minimum at 0.

b). For some $L \geqslant m > 0$, the potential U is m-strongly convex and L-smooth, i.e. for all x, z, one has the bounds

$$\frac{m}{2}|z|^2 \leqslant U(x+z) - U(x) - \langle \nabla U(x), z\rangle \leqslant \frac{L}{2}|z|^2.$$

We impose here spherical symmetry on the potential to make our proof simple, noting that similar results could be expected to hold without this assumption. A very natural example of π satisfying the above is the normal distribution with covariance matrix $\sigma_0^2 \mathrm{Id}$, for which one can take $m = L = \frac{1}{\sigma_0^2}$.

Example 96. Assume π is $\mathcal{N}(0, \sigma_0^2 I_d)$, so $U(x) = \frac{1}{2{\sigma_0}^2} |x|^2$. Then

$$U(x+z) - U(x) - \langle \nabla U(x), z \rangle = \frac{1}{2\sigma_0^2} |z|^2 ,$$

so we have $L = m = 1/\sigma_0^2$.

Another natural class of examples with strongly convex and smooth potentials (but not spherical symmetry) comes from considering Bayesian posterior measures for which the prior is normal, and the log-likelihood is concave with bounded Hessian.

Example 97. Consider the task of Bayesian logistic regression, taking as prior $\pi_0 = \mathcal{N}(0, \sigma_0^2 I_d)$, and observing covariate-response pairs $\{(a_i, y_i)\}_{i=1}^N \subset \mathbb{R}^d \times \{0, 1\}$. The potential corresponding to the posterior measure is then given by

$$U(x) = \frac{1}{2{\sigma_0}^2} |x|^2 + \sum_{i=1}^N \{\log(1 + \exp(-\langle a_i, x \rangle)) - y_i \langle a_i, x \rangle\}$$

Writing A for the $n \times d$ matrix with columns given by the $\{a_i\}$, one can check that U is m-strongly convex and L-smooth with $m \geqslant \frac{1}{\sigma_0^2}$ and $L \leqslant \frac{1}{\sigma_0^2} + \frac{1}{4}\lambda_{\text{Max}}\left(AA^\top\right)$.

The strategy of the proof of the following is to combine two different coupling arguments, in combination with a global application of Theorem 80, which itself rests on the isoperimetric inequality of Lemma 78. Recall that the proposal increments are $\mathcal{N}(0, \sigma d^{-1/2})$. We define "the centre" of the space to be $\{x : |x| \leq b_\kappa \sigma d^{1/2}\}$ for some constant $b_\kappa > 0$, and we always consider points that are close to each other, in that $|x - y| \leq b_\delta \sigma d^{-1/2}$ for some (small) constant b_δ. The proposals Q_x and Q_y can be made close in total variation by Pinsker's inequality for sufficiently small b_δ.

a). When x and y are both in "the centre", we can then ensure that $P(x, \cdot)$ and $P(y, \cdot)$ are close in total variation by additionally ensuring that the acceptance probability is uniformly lower bounded in the centre by a constant strictly above $1/2$ since then

$$\|P(x,\cdot) - P(y,\cdot)\|_{\text{TV}} \leq \|P(x,\cdot) - Q_x\|_{\text{TV}} + \|Q_x - Q_y\|_{\text{TV}} + \|P(y,\cdot) - Q_y\|_{\text{TV}}$$

can be made less than 1 by taking b_δ and $b_\kappa \sigma^2$ sufficiently small. This part of the proof that imposes a maximal value of σ, which is slightly at odds with the common practice of making the acceptance probability close to $1/4$ rather than larger than $1/2$.

b). When at least one of x and y are not in "the centre", we can use a different coupling argument that takes advantage of the fact that the set of points $\{w : |w| \leq |x| \wedge |y|\}$ will be accepted as proposals from both x and y, and is sufficiently large if b_κ is large enough. This overlap allows one to obtain a non-trivial bound on $\|P(x,\cdot) - P(y,\cdot)\|_{\text{TV}}$ with an acceptance rate that is less than $1/2$, which is important because in the tails of the distribution one cannot obtain an acceptance rate larger than $1/2$.

Theorem 98. *Let Assumption 95 hold. Let $\sigma = \varsigma/\sqrt{L}$ with $\varsigma \le \varsigma_\star = 0.073$. Then the conductance (see equation (11)) is lower bounded as follows:*

$$\kappa(0) \ge 8.46 \times 10^{-5} \varsigma \sqrt{\frac{m}{Ld}},$$

and hence

$$\mathrm{Gap}(P) = \mathrm{Gap}_{\mathrm{R}}(P) \ge 8.94 \times 10^{-10} \cdot \varsigma^2 \cdot \frac{m}{Ld}.$$

Proof. Let $\kappa := (4 + 1/16)\sigma$ and $\delta := \sigma/16$. Let $S := \mathcal{B}\left(0, \kappa \cdot d^{1/2}\right)$. Assume $x, y \in \mathsf{E}$ satisfy $|x - y| \le \delta_d := \delta d^{-1/2}$. In either case $(x, y) \in S \times S$ or $(x, y) \notin S \times S$, then $\|P(x, \cdot) - P(y, \cdot)\|_{\mathrm{TV}} \le \frac{31}{32}$ by Lemma 105 Lemma 108 respectively. Hence, we may apply Theorem 80 with $C = \mathsf{E}$ to deduce

$$\begin{aligned}
\kappa(0) &= \inf_{A \in \mathscr{E}} \frac{\mu \otimes P(A \times A^{\complement})}{\mu \otimes \mu(A \times A^{\complement})} \\
&\ge \frac{\varepsilon}{4} \min\left\{1, \frac{\log 2}{4} \frac{\sigma}{16} \sqrt{\frac{m}{d}}\right\} \\
&\ge \frac{1}{4 \cdot 32} \min\left\{1, \frac{\log 2}{4} \frac{\varsigma}{16} \sqrt{\frac{m}{Ld}}\right\} \\
&\ge \frac{1}{128} \min\left\{1, 0.01083\varsigma \sqrt{\frac{m}{Ld}}\right\} \\
&\ge 8.46 \times 10^{-5} \varsigma \sqrt{\frac{m}{Ld}},
\end{aligned}$$

noting that $m \le L$ by Assumption 95, $\varsigma \le \varsigma_\star < 1$ and $d \ge 1$. The bound on $\mathrm{Gap}_{\mathrm{R}}(\mathrm{P})$ follows by (11), and we have $\mathrm{Gap}(P) = \mathrm{Gap}_{\mathrm{R}}(P)$ by [5, Lemma 3.1], since Q is Gaussian. □

Remark 99. If $\pi = \mathcal{N}(0, \sigma_0^2 I_d)$, then $m/L = 1$ and $L = 1/\sigma_0^2$. Hence, we see that σ should scale proportionally with σ_0 as one would expect by a reparametrization argument, and that the bound is then independent of σ_0. The conductance/spectral gap lower bound is maximized by taking $\sigma = 0.073/\sqrt{L}$, and for our argument one cannot take σ larger than this. Theorem 110 below shows that a more specific argument allows for a stronger statement allowing arbitrary $\varsigma > 0$ while retaining the same dimension dependence.

Remark 100. In several places in the proof we have adopted dimension-independent bounds, e.g. by taking $d = 1$, which are certainly sub-optimal for large d. Similarly, for the sake of clarity we have made a few choices of constants that are certainly not optimal. Hence, we can expect that a more refined analysis would produce a larger lower bound on the conductance and a larger maximum value of $\sigma\sqrt{L}$. However, the proof strategy of ensuring a high acceptance rate in the centre does seem to naturally force σ to be artificially small.

Remark 101. We can inspect Assumption 95 when $u \colon \mathbb{R}_+ \to \mathbb{R}$ is continuously differentiable. It is useful to understand conditions on the function u which will guarantee that the desired estimates hold. First, compute explicitly that

$$\nabla U(x) = 2 \cdot \dot{u}\left(|x|^2\right) \cdot x,$$
$$\nabla^2 U(x) = 2 \cdot \left[2 \cdot \ddot{u}\left(|x|^2\right) \cdot x \cdot x^\top + \dot{u}\left(|x|^2\right) \cdot \mathrm{Id}\right].$$

For sufficiently smooth potentials, strong convexity and smoothness can be formulated in terms of the first two derivatives of u. In particular, m-strong convexity requires that for all x, it holds that

$$\begin{aligned}
m &\leqslant \inf_v \left\{ \frac{v^\top \nabla^2 U(x) v}{|v|^2} \right\} \\
&= 2 \cdot \inf_v \left\{ \frac{2\ddot{u}\left(|x|^2\right)\left(v^\top x\right)^2 + \dot{u}\left(|x|^2\right)|v|^2}{|v|^2} \right\} \\
&= 2 \cdot \inf_v \left\{ \frac{2\ddot{u}\left(|x|^2\right)\left(v^\top x\right)^2}{|v|^2} + \dot{u}\left(|x|^2\right) \right\} \\
&= 2 \cdot \left\{ 2 \cdot \min\left(0, \ddot{u}\left(|x|^2\right) \cdot |x|^2\right) + \dot{u}\left(|x|^2\right) \right\},
\end{aligned}$$

i.e. that $\inf_{s \geqslant 0} \{2 \cdot \min(0, \ddot{u}(s) \cdot s) + \dot{u}(s)\} \geqslant \frac{m}{2}$. Similar calculations show that L-smoothness requires that $\sup_{s \geqslant 0} \{2 \cdot \max(0, \ddot{u}(s) \cdot s) + \dot{u}(s)\} \leqslant \frac{L}{2}$. To be more concrete, suppose that u satisfies $0 < m_1 \leqslant \dot{u}(s) \leqslant L_1$ and $|\ddot{u}(s)| \leqslant L_2 s^{-1}$ with $L_2 \leqslant \frac{m_1}{2}$, i.e. it is increasing, essentially sandwiched between two affine functions, and its derivative has slow variation at infinity. It then follows that

$$2 \cdot \min(0, \ddot{u}(s) \cdot s) + \dot{u}(s) \geqslant m_1 - 2 \cdot L_2$$
$$2 \cdot \max(0, \ddot{u}(s) \cdot s) + \dot{u}(s) \leqslant L_1 + 2 \cdot L_2,$$

i.e. that we can take $m = m_1 - 2 \cdot L_2 > 0$, $L = L_1 + 2 \cdot L_2$.

The following two lemmas are known and useful bounds on the total variation distance between two normal distributions, and tail probabilities for χ^2 random variables.

Lemma 102. *For any $\epsilon > 0$ and $x, y \in \mathsf{X}$ such that $|x - y| \leqslant \epsilon \cdot d^{-1/2}$ it holds that*

$$\|Q_x - Q_y\|_{\mathrm{TV}} \leq \frac{\epsilon}{2\sigma}.$$

Proof. This is obtained via Pinsker's inequality. Compute that

$$\begin{aligned}
\mathrm{KL}(Q_x, Q_y) &= \mathbb{E}_{u \sim \mathcal{N}(x, d^{-1}\sigma^2 \mathrm{Id})} \left[\frac{|u-y|^2}{2\sigma^2/d} - \frac{|u-x|^2}{2\sigma^2/d} \right] \\
&= \frac{d}{2 \cdot \sigma^2} \cdot \mathbb{E}_{\xi \sim \mathcal{N}(0, \mathrm{Id})} \left[|x - y + \sigma d^{-1/2} \cdot \xi|^2 - |\sigma d^{-1/2} \cdot \xi|^2 \right] \\
&= \frac{d}{2 \cdot \sigma^2} \cdot |x-y|^2.
\end{aligned}$$

Hence, if $|x-y| \leqslant \epsilon \cdot d^{-1/2}$ then it follows that $\mathrm{KL}\left(Q_x, Q_y\right) \leqslant \frac{\epsilon^2}{2\cdot\sigma^2}\cdot$ Recalling Pinsker's inequality, we deduce that

$$\begin{aligned}\|Q_x - Q_y\|_{\mathrm{TV}} &\leqslant \sqrt{\mathrm{KL}\left(Q_x, Q_y\right)/2} \\ &= \frac{\epsilon}{2\sigma},\end{aligned}$$

as claimed. $\square$

Lemma 103 ([24, Lemma 1]). *If $W \sim \chi_d^2$ then for $u > 0$ we have*

$$\mathbb{P}\left(W \geqslant d + 2\sqrt{du} + 2u\right) \leq \exp\left(-u\right).$$

In particular, for $\epsilon \in (0,1)$, with $\exp(-u) = \epsilon$, $\chi(\epsilon, d) := 1 + 2\sqrt{\frac{\log \epsilon^{-1}}{d}} + 2\frac{\log \epsilon^{-1}}{d}$ and $\chi(\epsilon) := \chi(\epsilon, 1)$, we have

$$\mathbb{P}\left(W \geqslant d \cdot \chi(\epsilon)\right) \leqslant \mathbb{P}\left(W \geqslant d \cdot \chi(\epsilon, d)\right) \leqslant \epsilon.$$

We also have, for $u > 0$,

$$\mathbb{P}\left(W \leq d - 2\sqrt{du}\right) \leq \exp\left(-u\right).$$

Lemma 104. *Assume that U attains its minimum at 0, and is $L-$smooth. For any $\epsilon > 0$, if $\kappa \geq \sigma$ and*

$$\kappa\sigma \leq \frac{1}{L} \cdot \frac{-\log(1-\frac{\epsilon}{2})}{\chi(\frac{\epsilon}{4})} \cdot \frac{2}{3},$$

then for all $x \in \mathcal{B}\left(0, \kappa \cdot d^{1/2}\right)$

$$\|Q_x(\cdot) - P(x,\cdot)\|_{\mathrm{TV}} \leq \epsilon.$$

Proof. First, note that for $(x, A) \in \mathsf{E} \times \mathscr{E}$,

$$|Q_x(A) - P(x,A)| = \left|\int \mathbf{1}\{x + d^{-1/2}z \in A\}[1 - \alpha(x, d^{-1/2}z)]Q(\mathrm{d}z) - [1-\alpha(x)]\mathbf{1}\{x \in A\}\right|,$$

with $\alpha(x) := \int \alpha\left(x, d^{-1/2}z\right) \cdot Q(\mathrm{d}z)$ as in (23), which is maximized for $A = \mathsf{E}\setminus\{x\}$ or $A = \{x\}$ since we are considering the difference of non-negative terms. Therefore

$$\|Q_x(\cdot) - P(x,\cdot)\|_{\mathrm{TV}} = \int |1 - \alpha\left(x, d^{-1/2}z\right)| \cdot Q(\mathrm{d}z) = 1 - \alpha(x).$$

As suggested in [13] we use Markov's inequality, that is for $a \in (0,1]$,

$$\alpha(x) \geqslant a \cdot Q\left(\mathsf{r}\left(x, d^{-1/2}z\right) \geqslant a\right), \tag{24}$$

which motivates seeking a lower bound for

$$\mathsf{r}\left(x, d^{-1/2}z\right) = \frac{\pi\left(x + d^{-1/2}z\right)}{\pi\left(x\right)} = \exp\left(U\left(x\right) - U\left(x + d^{-1/2}z\right)\right).$$

We begin by noting that for $(x, z) \in \mathsf{X} \times \mathsf{Z}$,

$$U\left(x + d^{-1/2}z\right) - U\left(x\right) \leq \left\langle \nabla U\left(x\right), d^{-1/2}z \right\rangle + \frac{L}{2}\left|d^{-1/2}z\right|^2.$$

If $Z \sim Q = \mathcal{N}(0, \sigma^2 I_d)$, then $\langle \nabla U\left(x\right), Z\rangle \sim \mathcal{N}\left(0, \sigma^2 \cdot \left|\nabla U\left(x\right)\right|^2\right)$ and from the equivalent characterization of $L-$smoothness [13, Lemma 9] with $\nabla U\left(0\right) = 0$ we have $\left|\nabla U\left(x\right)\right| \leqslant L \cdot |x|$. Hence $\sup_{x \in \mathcal{B}\left(0, \kappa \cdot d^{1/2}\right)} \left|\nabla U\left(x\right)\right| \leqslant L \cdot \kappa \cdot d^{1/2}$, and from Chernoff's inequality for a normal random variable $\bar{Z} \sim \mathcal{N}\left(0, 1\right)$, that is $\mathbb{P}\left(\bar{Z} \geqslant u\right) \leq \exp\left(-\frac{1}{2}u^2\right)$ for $u > 0$, we can write that

$$\begin{aligned}
Q\left(\left\langle \nabla U\left(x\right), d^{-1/2}z \right\rangle > u\right) &= \mathbb{P}\left(\bar{Z} \cdot \frac{\sigma \cdot \left|\nabla U\left(x\right)\right|}{d^{1/2}} \geqslant u\right) \\
&\leqslant \mathbb{P}\left(\bar{Z} \cdot \frac{\sigma \cdot L \cdot \kappa \cdot d^{1/2}}{d^{1/2}} \geqslant u\right) \\
&= \mathbb{P}\left(\bar{Z} \geqslant \frac{u}{\sigma \cdot L \cdot \kappa}\right) \leqslant \exp\left(-\frac{u^2}{2 \cdot \sigma^2 \cdot L^2 \cdot \kappa^2}\right).
\end{aligned}$$

In particular, taking $u = \sigma \cdot L \cdot \kappa \cdot \sqrt{2 \cdot \log\left(\frac{4}{\epsilon}\right)}$, we see that

$$Q\left(\left\langle \nabla U\left(x\right), d^{-1/2}z \right\rangle > \sigma \cdot L \cdot \kappa \cdot \sqrt{2 \cdot \log\left(\frac{4}{\epsilon}\right)}\right) \leqslant \frac{\epsilon}{4}.$$

From Lemma 103, we have that

$$Q\left(\frac{L}{2} \cdot \left|d^{-1/2} \cdot z\right|^2 > \sigma^2 \cdot \frac{L}{2} \cdot \chi\left(\frac{\epsilon}{4}\right)\right) \leqslant \frac{\epsilon}{4}.$$

Note that for random variables X, Y we have $\mathbb{P}\left(X + Y > a + b\right) \leqslant \mathbb{P}\left(X > a\right) + \mathbb{P}\left(Y > b\right)$ for $a, b \in \mathbb{R}$, because

$$\begin{aligned}
\mathbb{P}\left(X + Y > a + b\right) =& \mathbb{P}\left(X > a, X + Y > a + b\right) + \mathbb{P}\left(X < a, Y > a + b - X\right) \\
\leq& \mathbb{P}\left(X > a\right) + \mathbb{P}\left(Y > b\right).
\end{aligned}$$

Consequently for $x \in \mathcal{B}\left(0, \kappa \cdot d^{1/2}\right)$,

$$
\begin{aligned}
&Q\left(U\left(x+d^{-1/2}z\right)-U\left(x\right)>\sigma\cdot L\cdot\chi\left(\frac{\epsilon}{4}\right)\cdot\left(\kappa+\frac{\sigma}{2}\right)\right)\\
&\quad\leqslant Q\left(\left\langle\nabla U\left(x\right),d^{-1/2}z\right\rangle+\frac{L}{2}\left|d^{-1/2}z\right|^{2}\geqslant\sigma\cdot L\cdot\kappa\cdot\chi\left(\frac{\epsilon}{4}\right)+\frac{\sigma^{2}}{2}\cdot L\cdot\chi\left(\frac{\epsilon}{4}\right)\right)\\
&\quad\leqslant Q\left(\left\langle\nabla U\left(x\right),d^{-1/2}z\right\rangle\geqslant\sigma\cdot L\cdot\kappa\cdot\chi\left(\frac{\epsilon}{4}\right)\right)\\
&\qquad\qquad\qquad\qquad\qquad\qquad+Q\left(\frac{L}{2}\left|d^{-1/2}z\right|^{2}\geqslant\frac{\sigma^{2}}{2}\cdot L\cdot\chi\left(\frac{\epsilon}{4}\right)\right)\\
&\quad\leqslant Q\left(\left\langle\nabla U\left(x\right),d^{-1/2}z\right\rangle\geqslant\sigma\cdot L\cdot\kappa\cdot\sqrt{2\cdot\log\left(\frac{4}{\epsilon}\right)}\right)\\
&\qquad\qquad\qquad\qquad\qquad\qquad+Q\left(\frac{L}{2}\left|d^{-1/2}z\right|^{2}\geqslant\frac{\sigma^{2}}{2}\cdot L\cdot\chi\left(\frac{\epsilon}{4}\right)\right)\\
&\quad\leqslant\frac{\epsilon}{4}+\frac{\epsilon}{4}=\frac{\epsilon}{2},
\end{aligned}
$$

that is,

$$
Q\left(r\left(x,d^{-1/2}z\right)\geqslant\exp\left(-\sigma L\cdot\chi\left(\frac{\epsilon}{4}\right)\cdot\left(\kappa+\frac{\sigma}{2}\right)\right)\right)\geqslant 1-\frac{\epsilon}{2}.
$$

It follows that by taking $a=\exp\left(-\sigma\cdot L\cdot\chi\left(\frac{\epsilon}{4}\right)\cdot\left(\kappa+\frac{\sigma}{2}\right)\right)$ in Markov's inequality (24) and assuming $\kappa\geq\sigma$ we can bound

$$
\begin{aligned}
\alpha\left(x\right)&\geqslant\exp\left(-\sigma\cdot L\cdot\chi\left(\frac{\epsilon}{4}\right)\cdot\left(\kappa+\frac{\sigma}{2}\right)\right)\cdot\left(1-\frac{\epsilon}{2}\right)\\
&\geqslant\exp\left(-L\cdot\chi\left(\frac{\epsilon}{4}\right)\cdot\frac{3}{2}\kappa\sigma\right)\cdot\left(1-\frac{\epsilon}{2}\right).
\end{aligned}
$$

Now if

$$
\kappa\sigma\leq\frac{1}{L}\cdot\frac{-\log(1-\frac{\epsilon}{2})}{\chi(\epsilon/4)}\cdot\frac{2}{3},
$$

then $\exp\left(-L\cdot\chi\left(\frac{\epsilon}{4}\right)\cdot\frac{3}{2}\kappa\sigma\right)\geqslant 1-\frac{\epsilon}{2}$, so that $\alpha\left(x\right)\geqslant\left(1-\frac{\epsilon}{2}\right)^{2}\geqslant 1-\epsilon$, and hence that $1-\alpha\left(x\right)\leqslant\epsilon$, as claimed. $\square$

Lemma 105. *Assume that U attains its minimum at 0, and is $L-$smooth. Let $\sigma\leq\varsigma/\sqrt{L}$ with $\varsigma\leq\varsigma_{\star}:=0.073$, $\kappa:=(4+1/16)\sigma$, and $\delta:=\sigma/16$. Let $S=\mathcal{B}\left(0,\kappa\cdot d^{1/2}\right)$. Then for $(x,y)\in S\times S$ such that $|x-y|\leq\delta d^{-1/2}$ we have*

$$
\|P(x,\cdot)-P(y,\cdot)\|_{\mathrm{TV}}\leq\frac{31}{32}.
$$

Proof. We have

$$
\|P(x,\cdot)-P(y,\cdot)\|_{\mathrm{TV}}\leq\|P(x,\cdot)-Q_{x}\|_{\mathrm{TV}}+\|Q_{x}-Q_{y}\|_{\mathrm{TV}}+\|P(y,\cdot)-Q_{y}\|_{\mathrm{TV}}.
$$

We have $\|Q_x - Q_y\|_{\mathrm{TV}} \le 1/32$ by Lemma 102. We wish to show that for $x \in S$,

$$\|P(x,\cdot) - Q_x\|_{\mathrm{TV}} \le 15/32.$$

This is ensured by Lemma 104: taking $\epsilon = 15/32$ we need to verify that for $b_\kappa := 4 + 1/16$,

$$\kappa\sigma = b_\kappa \sigma^2 \le \frac{1}{L} \cdot \frac{-\log(1-\frac{\epsilon}{2})}{\chi(\frac{\epsilon}{4})} \cdot \frac{2}{3},$$

and so it is sufficient to take

$$\sigma^2 \le \frac{0.073^2}{L} \le \frac{1}{L} \cdot \frac{-\log(\frac{49}{64})}{\chi(15/128)} \cdot \frac{2}{3} \cdot \frac{1}{4+1/16}.$$

□

Lemma 106. *If $|y - x| \le \delta$ and $|y| \ge \delta$ then $|x|^2 \ge (|y| - \delta)^2$.*

Proof. Let $x = y + r$ where $|r| \le \delta \le |y|$. Then by Cauchy–Schwarz,

$$\begin{aligned} |x|^2 &= |y|^2 + 2\langle y, r\rangle + |r|^2 \\ &\ge |y|^2 - 2|y||r| + |r|^2 \\ &= (|y| - |r|)^2, \end{aligned}$$

from which we can conclude. □

Lemma 107. *For any $\beta \in (0,1)$, let z_β denote the β-quantile of the $\mathcal{N}(0,1)$ distribution, namely $\mathbb{P}(Z_1 \ge z_\beta) = 1 - \beta$ for $Z_1 \sim N(0,1)$. For any $\alpha < 1/2$, let $x \in \mathbb{R}^d$ satisfy $|x| \ge c\sigma d^{1/2}$ for some $c > (2\alpha z_{1-\alpha})^{-1}$. Then if $W \sim Q_x$,*

$$\mathbb{P}(|W| \le |x|) \ge \alpha - \frac{1}{2cz_{1-\alpha}} > 0.$$

In particular, if $\alpha = 1/4$ and $c > 3$, $\mathbb{P}(Z \in A) > \frac{1}{4} - \frac{3}{4c} = \frac{1}{4}(1 - \frac{3}{c}) > 0$.

Proof. Without loss of generality, we may assume that $x = (-|x|, 0, \ldots, 0)$. Let $W = x + \sigma d^{-1/2} Z$ where $Z \sim \mathcal{N}(0, I_d)$, and $w = x + \sigma d^{-1/2} z$. Then

$$\begin{aligned} A &= \{z : |w| \le |x|\} \\ &= \{z : \left|x + \sigma d^{-1/2} z\right| \le |x|\} \\ &= \left\{z : 2\frac{\sigma}{d^{1/2}} \sum_{i=1}^d x_i z_i + \frac{\sigma^2}{d} \sum_{i=1}^d z_i^2 \le 0\right\} \\ &= \left\{z : \frac{\sigma}{d^{1/2}} \sum_{i=1}^d z_i^2 \le -2 \sum_{i=1}^d x_i z_i\right\} \\ &= \left\{z : \sum_{i=1}^d z_i^2 \le \frac{2}{\sigma} |x| z_1 d^{1/2}\right\} \\ &\supseteq \left\{z : \sum_{i=1}^d z_i^2 \le 2cv_1 d\right\} \cap \{z : z_1 \ge v_1\}, \end{aligned}$$

for any $v_1 > 0$. Now take $v_1 = z_{1-\alpha}$. Then,

$$\begin{aligned}\mathbb{P}\left(Z \in A\right) &\geq \mathbb{P}\left(|Z|^2 \leq 2cz_{1-\alpha}d, Z_1 \geq z_{1-\alpha}\right) \\ &\geq \mathbb{P}\left(|Z|^2 \leq 2cz_{1-\alpha}d\right) + \mathbb{P}(Z_1 \geq z_{1-\alpha}) - 1 \\ &= \mathbb{P}\left(|Z|^2 \leq 2cz_{1-\alpha}d\right) + \alpha - 1.\end{aligned}$$

By Markov's inequality, we have

$$\mathbb{P}\left(|Z|^2 > 2cz_{1-\alpha}d\right) \leq \frac{1}{2cz_{1-\alpha}},$$

and so

$$\mathbb{P}\left(Z \in A\right) \geq \alpha - \frac{1}{2cz_{1-\alpha}}.$$

For the last part, observe that if $\alpha = 1/4$ then $z_{1-\alpha} > 2/3$, and the conclusion follows. □

Lemma 108. *Assume $U(x) = u(|x|^2)$ with $u : [0,\infty) \to [0,\infty)$ increasing. For $\sigma > 0$, let $\kappa = b_\kappa\sigma$, $\delta = b_\delta\sigma$ for some constants $b_\kappa > b_\delta$. Let $(x,y) \in \left(\mathcal{B}\left(0, \kappa \cdot d^{1/2}\right) \times \mathcal{B}\left(0, \kappa \cdot d^{1/2}\right)\right)^{\complement}$. Then if $|x-y| \leq \delta d^{-1/2}$, we have that*

$$\|P(x,\cdot) - P(y,\cdot)\|_{\mathrm{TV}} \leq \frac{3}{4} + \frac{3}{4(b_\kappa - b_\delta)} + \frac{b_\delta}{2}. \tag{25}$$

In particular, if we take $b_\kappa = 4 + 1/16$ and $b_\delta = 1/16$, then we obtain

$$\|P(x,\cdot) - P(y,\cdot)\|_{\mathrm{TV}} \leq \frac{31}{32}.$$

Proof. For $x, y \in \left(\mathcal{B}\left(0, \kappa \cdot d^{1/2}\right) \times \mathcal{B}\left(0, \kappa \cdot d^{1/2}\right)\right)^{\complement}$, we construct a coupling (X', Y') such that $X' \sim P(x,\cdot)$ and $Y' \sim P(y,\cdot)$, and will show that $\mathbb{P}(X' = Y') \geq 1 - \epsilon$, with ϵ as in the right-hand side of (25). Without loss of generality, assume $|x| \leq |y|$. Hence, we have by Lemma 106 the (crude) bound $|y| \geq |x| \geq \kappa d^{1/2} - \delta d^{-1/2} \geq (\kappa - \delta)d^{1/2}$. Let (W_x, W_y) be distributed according to a maximal coupling of Q_x and Q_y. By Lemma 102,

$$\mathbb{P}(W_x = W_y) = 1 - \|Q_x - Q_y\|_{\mathrm{TV}} \geq 1 - \delta/2\sigma.$$

On the event $W_x = W_y$, we have $X' = Y' = W_x$ if $|W_x| \leq |x|$, since $U(x) = u(|x|)$ so the proposals will be accepted with probability one. Note that $W_x = x + \sigma d^{-1/2} Z$, where $Z \sim \mathcal{N}(0, I_d)$. Hence, by Lemma 107,

$$\mathbb{P}(|W_x| \leq |x|) \geq \alpha - \frac{1}{2cz_{1-\alpha}},$$

for any $\alpha < 1/2$ and $c = (\kappa - \delta)/\sigma$. Hence we have the bound

$$\begin{aligned}\mathbb{P}(X' = Y') &\geq \mathbb{P}(W_x = W_y, |W_x| \leq |X|) \\ &\geq \mathbb{P}(W_x = W_y) + \mathbb{P}(|W_x| \leq |X|) - 1 \\ &\geq \mathbb{P}(|W_x| \leq |X|) - \frac{\delta}{2\sigma} \\ &\geq \alpha - \frac{\sigma}{2(\kappa - \delta) z_{1-\alpha}} - \frac{\delta}{2\sigma}.\end{aligned}$$

Now, taking $\alpha = 1/4$, we obtain

$$\mathbb{P}(X' = Y') \geq \frac{1}{4} - \frac{3}{4(b_\kappa - b_\delta)} - \frac{b_\delta}{2},$$

and we conclude by the coupling inequality $\|P(x, \cdot) - P(y, \cdot)\|_{\mathrm{TV}} \leq \mathbb{P}(X' \neq Y')$. ☐

5.4 Spectral gap for the RWM on a Gaussian target

When π is $\mathcal{N}(0, \sigma_0^2 I_d)$, it is possible to obtain more precise bounds on the conductance and spectral gap, and also for the proposal standard deviation to be an arbitrary multiple of σ_0, when scaled appropriately by $d^{-1/2}$.

Lemma 109. *Assume $U(x) = \frac{1}{2\sigma_0^2} |x|^2$. Let $X' \sim P(x, \cdot)$ with proposal $W = x + \sigma_d Z$, where $\sigma_d = \varsigma d^{-1/2} \sigma_0$ for some $\varsigma > 0$ and $Z \sim \mathcal{N}(0, I_d)$. Then*

$$\mathbb{P}(X' = W) \geq \exp\left\{-\frac{\varsigma^2}{2}\left[1 + 2d^{-1/2} + 2d^{-1}\right]\right\} \cdot \frac{1}{2} \cdot (1 - \mathrm{e}^{-1}).$$

Proof. Since the proposal and target are spherically symmetric, we may assume without loss of generality that $x = (x_1, 0, \ldots, 0)$. Then

$$\begin{aligned}\left|x + \sigma d^{-1/2} z\right|^2 &= |x|^2 + 2\sigma_d \langle x, z\rangle + \frac{\sigma^2}{d} |z|^2 \\ &= |x|^2 + 2\sigma_d x_1 z_1 + \sigma_d^2 |z|^2.\end{aligned}$$

Hence,

$$U(W) - U(x) = \frac{1}{2\sigma_0^2}\left\{2\sigma_d x_1 Z_1 + \sigma_d^2 |Z|^2\right\}. \tag{26}$$

Now, for $r_d > 0$,

$$\mathbb{P}\left(x_1 Z_1 \leq 0, \frac{\sigma_d^2}{2\sigma_0^2} |Z|^2 \leq r_d\right) = \frac{1}{2}\mathbb{P}\left(|Z|^2 \leq d r_d \cdot \frac{2}{\varsigma^2}\right),$$

since $\mathbb{I}(Z_1 > 0)$ is independent of $|Z|^2$. By Lemma 103, we have

$$\mathbb{P}\left(|Z|^2 \leq d\left\{1 + 2\sqrt{\frac{u}{d}} + 2\frac{u}{d}\right\}\right) \geq 1 - \exp(-u).$$

So, taking $u = 1$, we set

$$r_d = \frac{\varsigma^2}{2}(1 + 2d^{-1/2} + 2d^{-1}),$$

which gives

$$\mathbb{P}\left(x_1 Z_1 \leq 0, \frac{\sigma_d^2}{2\sigma_0^2}|Z|^2 \leq r_d\right) \geq \frac{1}{2} \cdot \{1 - \exp(-1)\}.$$

It thus follows from (26) that

$$\mathbb{P}(U(W) - U(x) \leq r_d) \geq \frac{1}{2} \cdot (1 - \mathrm{e}^{-1}),$$

and so

$$\mathbb{P}\left(U(W) - U(x) \leq \frac{\varsigma^2}{2}\left\{1 + 2\sqrt{\frac{1}{d}} + 2\frac{1}{d}\right\}\right) \geq \frac{1}{2} \cdot (1 - \mathrm{e}^{-1}),$$

from which we may conclude, since on this event the proposal is accepted with probability at least $\exp\left\{-\frac{\varsigma^2}{2}\left[1 + 2d^{-1/2} + 2d^{-1}\right]\right\}$. □

Theorem 110. *Assume $U(x) = \frac{1}{2\sigma_0^2}|x|^2$, and let $\sigma = \varsigma\sigma_0$ for any $\varsigma > 0$. Then the conductance*

$$\kappa(0) \geq 0.00216 \exp\left\{-\varsigma^2\left[1 + 2d^{-1/2} + 2d^{-1}\right]\right\} \cdot \varsigma d^{-1/2},$$

and hence

$$\mathrm{Gap}(P) = \mathrm{Gap}_{\mathrm{R}}(P) \geq 5.83 \times 10^{-7} \cdot \exp\left\{-2\varsigma^2\left[1 + 2d^{-1/2} + 2d^{-1}\right]\right\} \cdot \varsigma^2 d^{-1}.$$

Proof. Let $v = \exp\left\{-\frac{\varsigma^2}{2}\left[1 + 2d^{-1/2} + 2d^{-1}\right]\right\} \cdot \frac{1}{2} \cdot (1 - \mathrm{e}^{-1})$. Let $\delta_d = v\sigma d^{-1/2}$, and $x, y \in \mathsf{X}$ such that $|x - y| \leq \delta_d$. Then $\|Q_x - Q_y\|_{\mathrm{TV}} \leq v/2$ by Lemma 102. We construct a specific coupling of (X', Y') such that $X' \sim P(x, \cdot)$ and $Y' \sim P(y, \cdot)$. Without loss of generality, we may assume that $|x| \leq |y|$. First, let (W_x, W_y) be distributed according to a maximal coupling of Q_x and Q_y. Then, with $\mathcal{U} \sim \mathrm{Uniform}(0, 1)$ we define

$$X' \mid \{W_x = w_x, W_y = w_y, \mathcal{U} = u\} = \begin{cases} w_x & u \leq \pi(w_x)/\pi(x), \\ x & u > \pi(w_x)/\pi(x). \end{cases}$$

Similarly we define

$$Y' \mid \{W_x = w_x, W_y = w_y, \mathcal{U} = u\} = \begin{cases} w_y & u \leq \pi(w_y)/\pi(y), \\ y & u > \pi(w_y)/\pi(y). \end{cases}$$

By Lemma 102,

$$\mathbb{P}(W_x = W_y) = 1 - \|Q_x - Q_y\|_{\mathrm{TV}} \geq 1 - \frac{v}{2}.$$

On the event $\{W_x = W_y\} \cap \{X' = W_x\}$, we have $X' = Y' = W_x$ since $\pi(y) \leq \pi(x)$. Hence, using Lemma 109, we have

$$\begin{aligned}
\mathbb{P}(X' = Y') &\geq \mathbb{P}(W_x = W_y, X' = W_x) \\
&\geq \mathbb{P}(W_x = W_y) + \mathbb{P}(X' = W_x) - 1 \\
&= 1 - \|Q_x - Q_y\|_{\mathrm{TV}} - 1 + \mathbb{P}(X' = W_x) \\
&\geq -\frac{v}{2} + v \\
&= \frac{v}{2}.
\end{aligned}$$

Hence, $\|P(x,\cdot) - P(y,\cdot)\|_{\mathrm{TV}} \leq \mathbb{P}(X' \neq Y') \leq 1 - \frac{v}{2}$ by the coupling inequality. We now take $\varepsilon = \frac{v}{2}$ will apply Theorem 80 with $C = \mathsf{E}$. Recall that $m = 1/\sigma_0^2$ from Example 96, and since $\frac{\log 2}{4} v\varsigma d^{-1/2} \leq 1$ for any $d \in \mathbb{N}$ and $\varsigma > 0$, we deduce that

$$\begin{aligned}
\kappa(0) &= \inf_{A \in \mathscr{E}} \frac{\mu \otimes P(A \times A^{\complement})}{\mu \otimes \mu(A \times A^{\complement})} \\
&\geq \frac{\varepsilon}{4} \min\left\{1, \frac{\log 2}{4} v\sigma d^{-1/2} \sqrt{m}\right\} \\
&= \frac{v}{8} \min\left\{1, \frac{\log 2}{4} v\varsigma d^{-1/2}\right\} \\
&= \frac{v^2}{32} \log 2 \cdot \varsigma d^{-1/2} \\
&= \exp\left\{-\varsigma^2 \left[1 + 2d^{-1/2} + 2d^{-1}\right]\right\} \cdot \frac{1}{4} \cdot (1 - \mathrm{e}^{-1})^2 \cdot \frac{\log 2}{32} \cdot \varsigma d^{-1/2} \\
&\geq 0.00216 \cdot \exp\left\{-\varsigma^2 \left[1 + 2d^{-1/2} + 2d^{-1}\right]\right\} \cdot \varsigma d^{-1/2}.
\end{aligned}$$

The bound on $\mathrm{Gap}_{\mathrm{R}}(\mathrm{P})$ follows by (11), and we have $\mathrm{Gap}(P) = \mathrm{Gap}_{\mathrm{R}}(P)$ by [5, Lemma 3.1], since Q is Gaussian. $\square$

Remark 111. The conductance lower bound is in $\Omega(d^{-1/2})$ and the spectral gap lower bound is in $\Omega(d^{-1})$. Fixing ς, we obtain

$$\lim_{d\to\infty} \inf \kappa_d(0) d^{1/2} \geq 0.00216 \cdot \exp\left\{-\varsigma^2\right\} \cdot \varsigma.$$

The maximizing ς for the bound is obtained by $\varsigma^2 = 1/2$, and this value of ς^2 gives

$$\lim_{d\to\infty} \inf \kappa_d(0) d^{1/2} \geq 0.000926.$$

This particular bound-maximizing value of ς is likely an artifact of the proof technique; optimal scaling results suggest that $\varsigma \approx 2.38$ is optimal in high dimensions [36], although they do not provide a bound on the conductance or spectral gap of the associated Markov operator.

To complement this result, we can show that the conductance must decrease at least as $\mathcal{O}(d^{-1/2})$ when the proposal standard deviation scales as $d^{-1/2}$, and that this is the slowest polynomial decay possible. Hence, we may infer that in terms of optimizing conductance and spectral gap, $d^{-1/2}$ is the correct polynomial scaling of the standard deviation.

Proposition 112. *Consider the RWM with Gaussian proposal of standard deviation $\sigma_d = \varsigma\sigma_0 d^{-\beta}$ for some $\beta \in \mathbb{R}$. Then the conductance is bounded as*

$$\kappa(0) \leq 2 \min\left\{2\varsigma d^{-\beta}, \exp\left(-\frac{d}{16}\right) + \exp\left(-\varsigma^2 \frac{d^{1-2\beta}}{8}\right)\right\},$$

and the upper bound is maximized for large d by taking $\beta = 1/2$, giving $\kappa_d(0) \leq 4\varsigma d^{-1/2}$.

Proof. First, let $A = \{x \in \mathsf{X} : x_1 \geq 0\}$, and we observe that $\pi(A) = \frac{1}{2}$. We let $Z \sim \mathcal{N}(0, I_d)$, and by neglecting the acceptance probability and using the Chernoff bound $\mathbb{P}(Z_1 \leq -z) \leq \exp(-z^2/2)$ for $z > 0$, we obtain the bounds

$$\begin{aligned}
\pi \otimes P(A \times A^{\complement}) &= \int_A \pi(\mathrm{d}x) P(x, A^{\complement}) \\
&\leq \int_A \pi(\mathrm{d}x) \mathbb{P}(x + \sigma_d Z \in A^{\complement}) \\
&= \int_A \pi(\mathrm{d}x) \mathbb{P}(x_1 + \sigma_d Z_1 < 0) \\
&= \frac{1}{\sqrt{2\pi\sigma_0^2}} \int_0^\infty \exp\left\{-\frac{x_1^2}{2\sigma_0^2}\right\} \mathbb{P}\left(Z_1 < -\frac{x_1}{\sigma_d}\right) \mathrm{d}x_1 \\
&\leq \frac{1}{\sqrt{2\pi\sigma_0^2}} \int_0^\infty \exp\left\{-\frac{x_1^2}{2\sigma_0^2} - \frac{x_1^2}{2\sigma_d^2}\right\} \mathrm{d}x \\
&= \left(\frac{\sigma_d^2}{\sigma_d^2 + \sigma_0^2}\right)^{1/2} \\
&\leq \frac{\sigma_d}{\sigma_0} = \varsigma d^{-\beta},
\end{aligned}$$

and it follows that $\kappa_d(0) \leq \varsigma d^{-\beta} / \pi \otimes \pi(A \times A^{\complement}) = 4\varsigma d^{-\beta}$, giving the first upper bound.

Now let $B := \{x : |x| \leq \delta_d\}$, where $\delta_d := \frac{\sigma_d \sqrt{d}}{4\sqrt{2}} \wedge c_d = \frac{\sigma d^{-\beta+1/2}}{4\sqrt{2}} \wedge c_d$, where c_d is chosen so that $\pi(|x| \leq c_d) = 1/2$. Hence, $\pi(B) \leq \frac{1}{2}$ and $\pi(B^{\complement}) \geq \frac{1}{2}$. We observe that for $x \in B$,

$$\begin{aligned}
|x + \sigma_d z|^2 - |x|^2 &= |x|^2 + 2\sigma_d \langle x, z\rangle + \sigma_d^2 |z|^2 - |x|^2 \\
&\geq -2\sigma_d |x| |z| + \sigma_d^2 |z|^2 \\
&\geq -2\sigma_d \delta_d |z| + \sigma_d^2 |z|^2 \\
&= \sigma_d |z| (\sigma_d |z| - 2\delta_d) . \qquad (27)
\end{aligned}$$

By Lemma 103,

$$\mathbb{P}(|Z|^2 \leq d - 2\sqrt{du}) \leq \exp(-u),$$

and taking $u = d/16$ and $C_d = d\sigma_d^2/2$ we obtain

$$\mathbb{P}(\sigma_d^2 |Z|^2 \leq C_d) = \mathbb{P}(|Z|^2 \leq d/2) \leq \exp(-d/16).$$

Since $2\delta_d\sqrt{C_d} \leq C_d/2$, on the event $\sigma_d^2 |Z|^2 \geq C_d$ we have from (27) that

$$|x + \sigma_d Z|^2 - |x|^2 \geq \sqrt{C_d}\left(\sqrt{C_d} - 2\delta_d\right) = C_d - 2\delta_d\sqrt{C_d} \geq C_d/2.$$

It follows that for $x \in B$, the acceptance probability satisfies

$$\begin{aligned}
\mathbb{E}\left[1 \wedge r(x, \sigma_d Z)\right] &= \mathbb{E}\left[1 \wedge \exp\left\{-\frac{1}{2\sigma_0^2}\left(|x + \sigma_d Z|^2 - |x|^2\right)\right\}\right] \\
&\leq 1 \cdot \mathbb{P}\left(\sigma_d^2 |Z|^2 \leq C_d\right) + \exp\left\{-\frac{1}{2\sigma_0^2}\left(C_d - 2\delta_d\sqrt{C_d}\right)\right\} \cdot \mathbb{P}\left(\sigma_d^2 |Z|^2 > C_d\right) \\
&\leq \exp\left(-\frac{d}{16}\right) + \exp\left\{-\frac{1}{2\sigma_0^2}\left(C_d - 2\delta_d\sqrt{C_d}\right)\right\} \\
&\leq \exp\left(-\frac{d}{16}\right) + \exp\left(-\frac{C_d}{4\sigma_0^2}\right) \\
&= \exp\left(-\frac{d}{16}\right) + \exp\left(-\frac{d\sigma_d^2}{8\sigma_0^2}\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{\pi \otimes P(B \times B^{\complement})}{\pi \otimes \pi(B \times B^{\complement})} &= \frac{\int \pi_B(\mathrm{d}x) P(x, B^{\complement})}{\pi(B^{\complement})} \\
&\leq 2\int \pi_B(\mathrm{d}x) P(x, \{x\}^{\complement}) \\
&\leq 2\left\{\exp\left(-\frac{d}{16}\right) + \exp\left(-\varsigma^2 \frac{d^{1-2\beta}}{8}\right)\right\},
\end{aligned}$$

and we conclude. □

A natural question is whether the lower bound for the spectral gap is of the correct order when the proposal standard deviation scales as $d^{-1/2}$, i.e. whether indeed $\mathrm{Gap}(P)$ scales as d^{-1}. In this case, we can verify directly that this is the case.

Proposition 113. *Let π be such that $\mathbb{E}_\pi[X_1] = 0$ and $\mathbb{E}_\pi[X_1^2] = \sigma_0^2$, and the proposal satisfy $Q_x(A) = \int_A \mathcal{N}(y; x, \sigma_d^2 I_d)\mathrm{d}y$ for $A \in \mathscr{X}$. Then*

$$\mathrm{Gap}(P) \leq \frac{\sigma_d^2}{2\sigma_0^2}.$$

Proof. We use the fact that $\mathrm{Gap}_\mathrm{R}(P) = \inf_{f\in \mathrm{L}_0^2(\pi)} \mathcal{E}(P, f)/ \|f\|_2^2$. Let $f(x) = x_1$. Then we compute

$$\begin{aligned}\mathcal{E}(P, f) &= \frac{1}{2}\int \pi(\mathrm{d}x)P(x, \mathrm{d}y)(y_1 - x_1)^2 \\ &\leq \frac{1}{2}\int \pi(\mathrm{d}x)Q_x(\mathrm{d}y)(y_1 - x_1)^2 \\ &= \frac{1}{2}\sigma_d^2,\end{aligned}$$

while $\|f\|_2^2 = \sigma_0^2$, and we conclude from $\mathrm{Gap}(P) \leq \mathcal{E}(P, f)/ \|f\|_2^2$. □

5.5 Central limit theorems

Obtaining a central limit theorem follows in a relatively straightforward manner when $\|P^n f\|_2^2$ decays quickly enough.

Proposition 114. *Let $f \in \mathrm{L}_0^2(\mu)$ with $\Phi(f) < \infty$. Let (X_n) be a Markov chain with Markov kernel P. Assume $\|P^n f\|_2^2 \leq \Phi(f)\gamma(n)$ with $\gamma(n) \in \mathcal{O}(n^{-a})$ for some $a > 1$. Then for μ-almost all X_0,*

$$\frac{1}{\sqrt{n}}\sum_{i=0}^{n-1} f(X_i) \xrightarrow{L} \mathcal{N}(0, \sigma^2),$$

where $\sigma^2 = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}_\mu\left[\left\{\sum_{i=0}^{n-1} f(X_i)\right\}^2\right] < \infty$.

Proof. We will verify the Maxwell–Woodroofe condition:

$$\sum_{n=1}^{\infty} n^{-3/2}\|V_n f\|_2 < \infty, \tag{28}$$

where $V_n f = \sum_{k=0}^{n-1} P^k f$. The central limit theorem then follows from [30, Corollary 1]. Minkowski's inequality gives

$$\|V_n f\|_2 = \left\|\sum_{k=0}^{n-1} P^k f\right\|_2 \leq \sum_{k=0}^{n-1}\left\|P^k f\right\|_2 \leq \Phi^{1/2}(f)\sum_{k=0}^{n-1}\gamma^{1/2}(k).$$

For $a > 1$ then we may write $\gamma^{1/2}(k) \le C(k+1)^{-a/2}$ for some $C > 0$ and note that $\gamma(0) < \infty$. Then

$$\begin{aligned}\frac{1}{C}\sum_{k=1}^{n-1}\gamma^{1/2}(k) &\le \sum_{k=1}^{n-1}(k+1)^{-a/2}\\ &= \sum_{k=1}^{n-1}(k+1)^{-a/2}\\ &\le \int_1^n x^{-a/2}\mathrm{d}x\\ &\le \frac{2}{2-a}n^{1-\frac{a}{2}},\end{aligned}$$

from which we may deduce that if $a \in (1,2)$ then $\sum_{k=0}^{n-1}\gamma^{1/2}(k) \in \mathcal{O}(n^{1/2-\epsilon})$ for some $\epsilon > 0$ and (28) holds. If $a \ge 2$ then $\gamma(n) \in \mathcal{O}(n^{-b})$ for any $b \in (1,2)$ and we can also conclude that (28) holds. □

Remark 115. One may verify that $\gamma(n) \in \mathcal{O}(n^{-a})$ by verifying a (Φ,β)-WPI with $\beta \in \mathcal{O}(s^{-a})$; see [1, Lemma 15]. We observe that if $\gamma(n) \in \mathcal{O}(n^{-b})$ with $\Phi = \|\cdot\|_{\mathrm{osc}}^2$ and $b > 1$, then by Proposition 13 we may deduce that for $p > 2$, $\|P^n f\|_2^2 \le \|f\|_{\mathrm{L}^p(\mu)}^2 \gamma_p(n)$ for $f \in \mathrm{L}_0^p(\mu)$ with $\gamma_p(n) \in \mathcal{O}(n^{-b(1-\frac{p}{2})})$. It then follows that a CLT holds for all $f \in \mathrm{L}_0^p(\mu)$ if $p > 2b/(b-1)$. If $\gamma(n)$ decays faster than polynomially, then a CLT holds for all $f \in \mathrm{L}_0^p(\mu)$ with $p > 2$ arbitrary.

A Miscellaneous results and proofs

Proof of Proposition 13. We follow the proof of [8, Lemma 5.1]. So we choose some $g \in \mathrm{L}_0^p(\mu)$ with $\|g\|_p = 1$, and for $R > 1$ to be chosen later, define $g_R := g \wedge R \vee (-R)$, and set $m_R := \int g_R \,\mathrm{d}\mu$. So we also obtain

$$|m_R| \le \|g\|_p^p / R^{p-1}$$

and

$$\|g - g_R\|_2^2 \le \|g\|_p^p / R^{p-2}.$$

Then we bound using the fact that P^n is a contraction on $\mathrm{L}^2(\mu)$,

$$\begin{aligned}\|P^n g\|_2 &\le \|P^n g - P^n g_R\|_2 + \|P^n(g_R - m_R)\|_2 + |m_R|\\ &\le \|g - g_R\|_2 + \|P^n(g_R - m_R)\|_2 + |m_R|\\ &\le \|g\|_p^p / R^{\frac{p-2}{2}} + \gamma^{1/2}(n)\|g_R - m_R\|_{\mathrm{osc}} + \|g\|_p^p / R^{p-1}\\ &\le 1/R^{(p-2)/2} + 2R\gamma^{1/2}(n) + 1/R^{p-1}\\ &\le 2R\gamma^{1/2}(n) + 2/R^{(p-2)/2}.\end{aligned}$$

Finally this can be optimized by choosing $R = 2^{2/p}\gamma^{-1/p}(n)$. The result then follows. □

Lemma 116. *Assume Φ defines a subspace of $\mathrm{L}_0^2(\mu)$, $\mathcal{F} = \{f \in \mathrm{L}_0^2(\mu) : \Phi(f) < \infty\}$. Let T be self-adjoint and assume that $f \in \mathcal{F} \Rightarrow Tf \in \mathcal{F}$. Let S denote the restriction of T to the Hilbert space $\bar{\mathcal{F}}$, the closure of $\mathcal{F}$. Then ψ in Remark 22 satisfies*

$$\psi(0;\Phi) = \mathrm{Gap}_{\mathrm{R}}(S).$$

If $\Phi = \|\cdot\|_{\mathrm{osc}}^2$, then $\bar{\mathcal{F}} = \mathrm{L}_0^2(\mu)$ and $\psi(0;\Phi)$ is the $\mathrm{L}_0^2(\mu)$ spectral gap of T.

Proof. $\mathcal{F}$ is a normed vector space with norm $\|\cdot\|_2$, and hence $\bar{\mathcal{F}}$ is a Hilbert space. We may deduce that the restriction of T to $\mathcal{F}$ is an operator from $\mathcal{F}$ to $\bar{\mathcal{F}}$, and that S is its unique extension as a bounded linear operator from $\bar{\mathcal{F}}$ to $\bar{\mathcal{F}}$. By [12, Theorem 22.A.19] we have $\sup_{f \in \bar{\mathcal{F}}, \|f\|_2 = 1} \langle Sf, f \rangle = \sup \sigma(S)$ so that $\inf_{f \in \bar{\mathcal{F}}} \mathcal{E}(S,f)/\|f\|_2^2 = \mathrm{Gap}_{\mathrm{R}}(S)$.

Now assume that $\Phi = \|\cdot\|_{\mathrm{osc}}^2$. For any $f \in \mathrm{L}_0^2(\mu)$ we may define $f_n = \mathbf{1}_{A_n} \cdot f$ and $g_n = f_n - \mu(f_n)$, where $A_n = \{x : -n \le f_n(x) \le n\}$. Then (g_n) is a sequence of bounded functions in $\mathrm{L}_0^2(\mu)$ with $g_n \to f$ pointwise and $|g_n| \le |f|$. We have

$$|\mu(f_n)| = |\mu(f_n) - \mu(f)| \le \mu(|f_n - f|) = \|f_n - f\|_{L^1(\mu)} \le \|f_n - f\|_2 \,,$$

from which we obtain that $\|g_n - f\|_2 \le 2\|f_n - f\|_2 \to 0$ by dominated convergence, and hence $\bar{\mathcal{F}} = \mathrm{L}_0^2(\mu)$. ☐

Remark 117. If $T = P^*P$, then T is self-adjoint and positive, and by [12, Theorem 22.A.17 and Corollary 22.A.18] we may further deduce that

$$\|S\|_{\bar{\mathcal{F}} \to \bar{\mathcal{F}}} = \|R\|_{\bar{\mathcal{F}} \to \bar{\mathcal{F}}}^2 = 1 - \psi(0;\Phi),$$

where R is the restriction of P to $\bar{\mathcal{F}}$.

Lemma 118. *Let P be a μ-reversible Markov transition kernel P on $(\mathsf{E}, \mathscr{E})$. Then for any $A \in \mathcal{E}$*

$$\mathcal{E}(P, \mathbf{1}_A) = \mu \otimes P\big(A \times A^{\complement}\big) \text{ and } \mathrm{var}\big(\mathbf{1}_A\big) = \mu \otimes \mu\big(A \times A^{\complement}\big) \,.$$

Proof. Let $A \in \mathcal{E}$. By polarization, considering when $|\mathbf{1}_A(x) - \mathbf{1}_A(y)| = 1 \neq 0$ and using the symmetry of $\mu \otimes P$ we have

$$\begin{aligned}
\mathcal{E}(P, \mathbf{1}_A) &= \frac{1}{2} \int \big[\mathbf{1}_A(x) - \mathbf{1}_A(y)\big]^2 \mu \otimes P(\mathrm{d}x, \mathrm{d}y) \\
&= \frac{1}{2} \int \big[\mathbf{1}_A(x)\mathbf{1}_{A^{\complement}}(y) + \mathbf{1}_{A^{\complement}}(x)\mathbf{1}_A(y)\big] \mu \otimes P(\mathrm{d}x, \mathrm{d}y) \\
&= \int \mathbf{1}_A(x)\mathbf{1}_{A^{\complement}}(y) \mu \otimes P(\mathrm{d}x, \mathrm{d}y) \,.
\end{aligned}$$

The result on the variance follows by considering $P(x, A) = \mu(A)$ for $(x, A) \in \mathsf{E} \times \mathscr{E}$ and the classical identity $\mathrm{var}\big(\mathbf{1}_A\big) = \frac{1}{2}\mathbb{E}_{\mu \otimes \mu}\Big[\big(\mathbf{1}_A(X) - \mathbf{1}_A(Y)\big)^2\Big]$. ☐

Lemma 119 ([25]). *Let ν be a symmetric probability measure on $(\mathsf{E}\times\mathsf{E}, \mathscr{E}\otimes\mathscr{E})$. Then for any $h\colon \mathsf{E}\times\mathsf{E} \to \mathbb{R}_+$ such that $h \in \mathrm{L}^1(\nu)$ and for any $x \in \mathsf{E}$, $y \to h(x,y)$ is constant. Writing $h(x) := h(x,y)$ for notational simplicity, define $A_u := \{x \in \mathsf{E}\colon h(x) \leq u\}$ for $u \geq 0$. Then we have*

$$\mathbb{E}_\nu\left[|h(X) - h(Y)|\right] = 2\int \nu(A_t, A_t^{\complement})\,\mathrm{d}t\,. \tag{29}$$

Proof. We have by symmetry of ν and Fubini,

$$\begin{aligned}\mathbb{E}_\nu\left[|h(X) - h(Y)|\right] &= 2\int\int \nu(\mathrm{d}x, \mathrm{d}y)\mathbf{1}\{h(x) \leq t < h(y)\}\,\mathrm{d}t \\ &= 2\int \nu(A_t, A_t^{\complement})\,\mathrm{d}t\,\cdot\end{aligned}$$

□

References

[1] Christophe Andrieu, Anthony Lee, Sam Power, and Andi Q. Wang. Comparison of Markov chains via weak Poincaré inequalities with application to pseudo-marginal MCMC. https://arxiv.org/abs/2112.05605v2, 2021.

[2] Yves F. Atchadé. Approximate spectral gaps for Markov chain mixing times in high dimensions. *SIAM Journal on Mathematics of Data Science*, 3(3):854–872, 2021.

[3] Dominique Bakry, Franck Barthe, Patrick Cattiaux, and Arnaud Guillin. A simple proof of the Poincaré inequality for a large class of probability measures. *Electronic Communications in Probability*, 13:60–66, 2008.

[4] Dominique Bakry, François Bolley, and Ivan Gentil. Around Nash inequalities. *Journées Équations aux dérivées partielles*, 2:1–16, 2010.

[5] Peter H. Baxendale. Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *The Annals of Applied Probability*, 15(1B):700–738, 2005.

[6] Alexandre Belloni and Victor Chernozhukov. On the computational complexity of MCMC-based estimators in large samples. *The Annals of Statistics*, 37(4):2011–2055, 2009.

[7] José A Cañizo and Stéphane Mischler. Harris-type results on geometric and subgeometric convergence to equilibrium for stochastic semigroups. *arXiv preprint arXiv:2110.09650*, 2021.

[8] Patrick Cattiaux, Djalil Chafai, and Arnaud Guillin. Central limit theorems for additive functionals of ergodic Markov diffusions processes. *ALEA, Lat. Am. J. Probab. Math. Stat*, 9(2):337–382, 2012.

[9] Patrick Cattiaux, Nathael Gozlan, Arnaud Guillin, and Cyril Roberto. Functional inequalities for heavy tailed distributions and application to isoperimetry. *Electronic Journal of Probability*, 15:346–385, 2010.

[10] Ben Cousins and Santosh Vempala. A cubic algorithm for computing Gaussian volume. In *Proceedings of the twenty-fifth annual ACM-SIAM symposium on discrete algorithms*, pages 1215–1228. SIAM, 2014.

[11] Persi Diaconis and Laurent Saloff-Coste. Nash inequalities for finite Markov chains. *Journal of Theoretical Probability*, 9(2):459–510, 1996.

[12] Randal Douc, Eric Moulines, Pierre Priouret, and Philippe Soulier. *Markov Chains*. Springer, 2018.

[13] Raaz Dwivedi, Yuansi Chen, Martin J. Wainwright, and Bin Yu. Log-concave sampling: Metropolis–Hastings algorithms are fast. *Journal of Machine Learning Research*, 20(183):1–42, 2019.

[14] Paul Embrechts and Marius Hofert. A note on generalized inverses. *Mathematical Methods of Operations Research*, 77(3):423–432, 2013.

[15] James Allen Fill. Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process. *The Annals of Applied Probability*, pages 62–87, 1991.

[16] Jørund Gåsemyr. The spectrum of the independent Metropolis–Hastings algorithm. *Journal of Theoretical Probability*, 19(1):152–165, 2006.

[17] Sharad Goel, Ravi Montenegro, and Prasad Tetali. Mixing time bounds via the spectral profile. *Electronic Journal of Probability*, 11:1–26, 2006.

[18] Fuzhou Gong and Liming Wu. Spectral gap of positive operators and applications. *Journal de Mathématiques Pures et Appliquées*, 85(2):151–191, 2006.

[19] Olle Häggström. On the central limit theorem for geometrically ergodic Markov chains. *Probability Theory and Related Fields*, 132(1):74–82, 2005.

[20] Søren F. Jarner and Gareth O. Roberts. Polynomial convergence rates of Markov chains. *The Annals of Applied Probability*, 12(1):224–247, 2002.

[21] Søren F. Jarner and Gareth O. Roberts. Convergence of heavy-tailed Monte Carlo Markov chain algorithms. *Scandinavian Journal of Statistics*, 34(4):781–815, 2007.

[22] Søren F. Jarner and Richard L. Tweedie. Necessary conditions for geometric and polynomial ergodicity of random-walk-type. *Bernoulli*, 9(4):559 – 578, 2003.

[23] Mark Jerrum and Alistair Sinclair. Conductance and the rapid mixing property for Markov chains: the approximation of permanent resolved. In *Proceedings of the twentieth annual ACM symposium on Theory of computing*, pages 235–244, 1988.

[24] B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, 28(5):1302 – 1338, 2000.

[25] Gregory F. Lawler and Alan D. Sokal. Bounds on the L^2 spectrum for Markov chains and Markov processes: a generalization of Cheeger's inequality. *Transactions of the American Mathematical Society*, 309(2):557–580, 1988.

[26] Anthony Lee and Krzysztof Łatuszyński. Variance bounding and geometric ergodicity of Markov chain Monte Carlo kernels for approximate Bayesian computation. *Biometrika*, 101(3):655–671, 08 2014.

[27] Thomas M Liggett. L_2 rates of convergence for attractive reversible nearest particle systems: the critical case. *The Annals of Probability*, 19(3):935–959, 1991.

[28] László Lovász. Hit-and-run mixes fast. *Mathematical Programming*, 86(3):443–461, 1999.

[29] László Lovász and Miklós Simonovits. Random walks in a convex body and an improved volume algorithm. *Random Structures & Algorithms*, 4(4):359–412, 1993.

[30] Michael Maxwell and Michael Woodroofe. Central limit theorems for additive functionals of Markov chains. *The Annals of Probability*, pages 713–724, 2000.

[31] Kerrie L. Mengersen and Richard L. Tweedie. Rates of convergence of the Hastings and Metropolis algorithms. *The Annals of Statistics*, 24(1):101–121, 1996.

[32] S.P. Meyn and R.L. Tweedie. *Markov chains and stochastic stability*. Cambridge University Press, 2 edition, 2009.

[33] Laurent Miclo and Cyril Roberto. Trous spectraux pour certains algorithmes de Métropolis sur $\mathbb{R}$. *Séminaire de Probabilités XXXIV*, pages 336–352, 2000.

[34] Ravi Montenegro and Prasad Tetali. Mathematical aspects of mixing times in Markov chains. *Foundations and Trends® in Theoretical Computer Science*, 1(3):237–354, 2006.

[35] Omiros Papaspiliopoulos and Gareth Roberts. Stability of the Gibbs sampler for Bayesian hierarchical models. *The Annals of Statistics*, 36(1):95 – 117, 2008.

[36] Gareth O. Roberts and Jeffrey S. Rosenthal. Optimal scaling for various Metropolis–Hastings algorithms. *Statistical Science*, 16(4):351–367, 2001.

[37] Gareth O. Roberts and Richard L. Tweedie. Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, 83(1):95–110, 1996.

[38] Michael Röckner and Feng-Yu Wang. Weak Poincaré inequalities and L2 convergence rates of Markov semigroups. *Journal of Functional Analysis*, 185:564–603, 2001.

[39] Chris Sherlock. Reversible Markov chains: variational representations and ordering. http://arxiv.org/abs/1809.01903, 2018.

[40] Wolfgang Stadje and Achim Wübker. Three kinds of geometric convergence for Markov chains and the spectral gap property. *Electronic Journal of Probability*, 16:1001–1019, 2011.

[41] Amirhossein Taghvaei and Prashant G. Mehta. On the Lyapunov Foster criterion and Poincaré inequality for reversible Markov chains. *IEEE Transactions on Automatic Control*, 2021.

[42] Luke Tierney. A note on Metropolis–Hastings kernels for general state spaces. *The Annals of Applied Probability*, 8(1):1–9, 1998.

[43] Feng Yu Wang. Criteria of spectral gap for Markov operators. *Journal of Functional Analysis*, 266(4):2137–2152, 2014.