The Annals of Statistics
2022, Vol. 50, No. 6, 3592–3618
https://doi.org/10.1214/22-AOS2241

COMPARISON OF MARKOV CHAINS VIA WEAK POINCARÉ INEQUALITIES WITH APPLICATION TO PSEUDO-MARGINAL MCMC

BY CHRISTOPHE ANDRIEU[a], ANTHONY LEE[b], SAM POWER[c] AND ANDI Q. WANG[d]

School of Mathematics, University of Bristol, [a]*C.Andrieu@bristol.ac.uk,* [b]*anthony.lee@bristol.ac.uk,* [c]*sam.power@bristol.ac.uk,* [d]*andi.wang@bristol.ac.uk*

We investigate the use of a certain class of functional inequalities known as weak Poincaré inequalities to bound convergence of Markov chains to equilibrium. We show that this enables the straightforward and transparent derivation of subgeometric convergence bounds for methods such as the Independent Metropolis–Hastings sampler and pseudo-marginal methods for intractable likelihoods, the latter being subgeometric in many practical settings. These results rely on novel quantitative comparison theorems between Markov chains. Associated proofs are simpler than those relying on drift/minorisation conditions and the tools developed allow us to recover and further extend known results as particular cases. We are then able to provide new insights into the practical use of pseudo-marginal algorithms, analyse the effect of averaging in Approximate Bayesian Computation (ABC) and the use of products of independent averages and also to study the case of log-normal weights relevant to particle marginal Metropolis–Hastings (PMMH).

1. Introduction.

1.1. *Motivation.* The theoretical analysis of Markov chain Monte Carlo (MCMC) algorithms can provide twofold benefits for users. On the one hand, it provides fundamental reassurance and theoretical guarantees for the correctness of algorithms, and on the other hand, can also offer guidance on parameter tuning to maximise efficacy.

Aside from high-dimensional scaling limit arguments [28], two approaches have proved particularly successful for characterising the properties of MCMC algorithms [6]: Lyapunov drift/minorisation conditions [15, 26, 30] and functional-analytic tools on Hilbert spaces, in particular in the reversible setup [23], Chapter 22 in [15]. The former have been the most successful for the study of stability and convergence rates, despite the inherent difficulty of constructing an appropriate Lyapunov function. A particular success has been the development of tools to analyse the scenario where the Markov transition kernel does not possess a spectral gap, and hence converges at a subgeometric rate (see [15] for a book-length treatment). In contrast, functional-analytic tools have been particularly successful at characterising the resulting asymptotic variance, but their application to characterising convergence rates has been limited to the scenario where a spectral gap exists (see [22] for example). This is despite the existence of functional-analytic tools such as weak Poincaré or Nash inequalities, which have been successfully applied to continuous-time Markov processes in the absence of a spectral gap [29].

The aim of this paper is to fill this gap, and show how weak Poincaré inequalities can be particularly useful for analysing certain MCMC algorithms and answering pertinent practical questions. Our main focus here will be on pseudo-marginal algorithms [3], a particular type of MCMC method for which pointwise unbiased estimates of the target density are sufficient

Received December 2021; revised August 2022.
MSC2020 subject classifications. Primary 65C40, 65C05; secondary 62J10.
Key words and phrases. Markov chain Monte Carlo, weak Poincaré inequality, subgeometric convergence, pseudo-marginal.

for their implementation. We show that weak Poincaré inequalities allow us to significantly expand and greatly simplify the results of [4], characterising precisely the degradation in performance incurred when using noisy estimates of the target density. This is particularly appealing because pseudo-marginal Markov kernels often do not possess a spectral gap on general state spaces, either because the noise is unbounded [3, 4] or because the noise is not uniformly bounded and "local proposals" are used [24], which is fairly common in practice.

To the best of our knowledge, while Nash inequalities for finite state space Markov chains have been considered in [13], weak Poincaré inequalities have not received the same attention in this context and it is not possible to point to a suitable reference for background. In Section 2, we provide a comprehensive overview of the theory tailored to the Markov chain scenario; some of the results given therein are new to the best of our knowledge. In Section 3, we develop a series of new comparison results between Markov chains sharing a common invariant distribution. In Section 4, we apply our results to pseudo-marginal algorithms, providing a simple and comprehensive theory of the impact of using noisy densities on the convergence properties of pseudo-marginal algorithms, which we leverage to clarify implementational considerations. We consider the effect of averaging, with applications to Approximate Bayesian Computation (ABC) and when using products of independent averages, and finally provide an analysis when the weights are log normal, relevant to the Particle Marginal MH (PMMH).

The proofs not appearing in the main text can be found in the Supplementary Material [2], along with a list of notation used throughout the article.

2. Weak Poincaré inequalities.

2.1. *General case.*

2.1.1. *Definitions and basic results.* Throughout this work, in analogue with the existing notions for continuous-time Markov processes [29], we will call a *weak Poincaré inequality* an inequality of the following form:

DEFINITION 1 (Weak Poincaré inequality, α-parameterisation). Given a Markov transition operator P on E, we will say that *P^*P satisfies a weak Poincaré inequality* if, for any $f \in \mathrm{L}^2_0(\mu)$,

$$\|f\|_2^2 \le \alpha(r)\mathcal{E}\big(P^*P, f\big) + r\Phi(f) \quad \forall r > 0,$$

where $\alpha : (0,\infty) \to [0,\infty)$ is a decreasing function, and $\Phi : \mathrm{L}^2(\mu) \to [0,\infty]$ is a functional satisfying for any $f \in \mathrm{L}^2(\mu)$, $c > 0$ and $n \in \mathbb{N}$,

$$\Phi(cf) = c^2\Phi(f), \qquad \Phi\big(P^n f\big) \le \Phi(f), \qquad \|f - \mu(f)\|_2^2 \le a\Phi\big(f - \mu(f)\big), \tag{1}$$

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

REMARK 2. A popular choice of Φ is $\Phi = \|\cdot\|_{\mathrm{osc}}^2$, for which $a \le 1$, but we will also later consider $\Phi = \|\cdot\|_{2p}^2$ for $p \ge 1$, which also has $a \le 1$ by Lyapunov's inequality.

REMARK 3. Note that $\alpha(r)$ typically diverges as $r \to 0$. By contrast, a *strong Poincaré inequality* refers to the situation when α is uniformly bounded above by $\alpha(r) \le 1/C_{\mathrm{P}}$ for some $C_{\mathrm{P}} > 0$; in this case, we may take $r \to 0$ and recover the standard strong Poincaré inequality $C_{\mathrm{P}}\|f\|_2^2 \le \mathcal{E}(P^*P, f)$ for $f \in \mathrm{L}^2_0(\mu)$, from which one can immediately deduce geometric convergence [17], that is, for any $f \in \mathrm{L}^2_0(\mu)$, $n \in \mathbb{N}$,

$$\|P^n f\|_2^2 \le (1 - C_{\mathrm{P}})^n \|f\|_2^2. \tag{2}$$

In what follows, we show that a weak Poincaré inequality implies the existence of a function $n \mapsto \gamma(n)$, which is decreasing to 0 such that, for any $f \in \mathrm{L}_0^2(\mu)$ and $\Phi(f) < \infty$,

$$\|P^n f\|_2^2 \le \gamma(n)\Phi(f). \tag{3}$$

A very useful equivalent formulation of the weak Poincaré inequality, which bears some resemblance to the "super-Poincaré inequality" of [29], is the following.

DEFINITION 4 (Weak Poincaré inequality, β-parameterisation). Given a Markov transition operator P on E, we will say that P^*P *satisfies a weak Poincaré inequality* if, for any $f \in \mathrm{L}_0^2(\mu)$,

$$\|f\|_2^2 \le s\mathcal{E}(P^*P, f) + \beta(s)\Phi(f) \quad \forall s > 0,$$

where $\beta : (0,\infty) \to [0,\infty)$ is a decreasing function with $\beta(s) \downarrow 0$ as $s \to \infty$, and $\Phi : \mathrm{L}^2(\mu) \to [0,\infty]$ is a functional satisfying, (1) for any $f \in \mathrm{L}^2(\mu)$, $c > 0$ and $n \in \mathbb{N}$.

These two formulations are equivalent; see our Remark 5 below, and we will typically refer to a "weak Poincaré inequality" without specifying the parameterisation. If there is ambiguity, we will write α- or β-weak Poincaré inequality to specify the parameterisation. Because a is such that $\|f\|_2^2 \le a\Phi(f)$ for all $f \in \mathrm{L}_0^2(\mu)$, one can always take $\beta \le a$ in Definition 4 and $\alpha(r) = 0$ for $r \ge a$ in Definition 1.

REMARK 5. Suppose an α-weak Poincaré inequality holds for a function α with $\alpha(r) = 0$ for $r \ge a$. Then a β-weak Poincaré inequality holds with $\beta(s) := \inf\{r > 0 : \alpha(r) \le s\}$. Conversely, suppose a β-weak Poincaré inequality holds for a function β with $\beta \le a$. Then an α-weak Poincaré inequality holds with $\alpha(r) := \inf\{s > 0 : \beta(s) \le r\}$. This procedure always returns a right-continuous function, so for a given α (or β) satisfying a weak Poincaré inequality, iterating this procedure will return the right-continuous version of α (or β).

While in practice establishing a weak Poincaré inequality is often the most tractable option, a third (essentially) equivalent formulation plays an important rôle to establish (3) with optimal rate function γ. We need the following functions.

DEFINITION 6. For β as in Definition 4, we let:

1. $K : [0,\infty) \to [0,\infty)$ be such that $K(u) := u\beta(1/u)$ for $u > 0$ and $K(0) := 0$,
2. $K^* : [0,\infty) \to [0,\infty]$ be such that $K^*(v) := \sup_{u \ge 0}\{uv - K(u)\}$ is the convex conjugate of K.

Then for $f \in \mathrm{L}_0^2(\mu)$ such that $0 < \Phi(f) < \infty$, the weak Poincaré inequality can be formulated as follows with $u = 1/s > 0$:

$$u\|f\|_2^2 \le \mathcal{E}(P^*P, f) + K(u)\Phi(f),$$

which by rearranging terms and optimising leads to

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

Relevant properties of K^* can be found in Lemma 1 of the Supplementary Material. The rate function γ in (3) is the inverse function of F_a given below, which is well defined.

LEMMA 7. *Let* $F_a(\cdot)\colon (0,a] \to \mathbb{R}$, *where* $(0,a] \subset \mathrm{D}$, *be given by*

$$F_a(x) := \int_x^a \frac{\mathrm{d}v}{K^*(v)},$$

where K^* *is given in Definition* 6 *and* $\mathrm{D} := \{v \geq 0 : K^*(v) < \infty\}$. *Then* $F_a(\cdot)$:

1. *is well defined, convex, continuous and strictly decreasing*;
2. *is such that* $\lim_{x\downarrow 0} F_a(x) = \infty$;
3. *has a well-defined inverse function* $F_a^{-1} : (0,\infty) \to (0,a)$, *with* $F_a^{-1}(x) \to 0$ *as* $x \to \infty$.

The main result of this section is as follows.

THEOREM 8. *Assume that* μ *and* P^*P *satisfy a weak Poincaré inequality as in Definition* 4. *Then for* $f \in \mathrm{L}_0^2(\mu)$ *such that* $0 < \Phi(f) < \infty$ *and any* $n \in \mathbb{N}$,

$$\|P^n f\|_2^2 \leq \Phi(f) F_a^{-1}(n),$$

where $F_a\colon (0,a] \to \mathbb{R}$ *is the decreasing convex and invertible function as in Lemma* 7.

REMARK 9. When $\int_a^\infty \frac{\mathrm{d}v}{K^*(v)} < \infty$, one can define $F_\infty(x) := \int_x^\infty \frac{\mathrm{d}v}{K^*(v)}$ for each $x > 0$, and since $F_a(x) \leq F_\infty(x)$, one can similarly derive a bound $\|P^n f\|_2^2 \leq \Phi(f) F_\infty^{-1}(n)$.

REMARK 10. A different proof relying on an alternative use of the Poincaré inequality is given in the Supplementary Material for completeness, which corresponds to the formulation of [29], Theorem 2.1, advocated by the authors for its tractability, but leads to suboptimal results. We have found the formulation of Theorem 8 sufficiently flexible for our applications. This general approach was in fact suggested in the continuous-time setting; see, for example, [29], equation (1.4), but only later utilised in [5], where improved rates were obtained. Our approach here can be seen as the natural discrete-time analogue; however, we further generalise the approach to allow for general Φ and $a \neq \infty$, and make explicit the connection with convex conjugates.

REMARK 11. If P^*P satisfies a strong Poincaré inequality with constant C_{P}, one may take the corresponding β to be $\beta(s) = a\mathbb{I}\{s \leq C_{\mathrm{P}}^{-1}\}$. Conversely, if $\beta(s) = a\mathbb{I}\{s \leq C_{\mathrm{P}}^{-1}\}$ then one can deduce that a strong Poincaré inequality holds. A simple calculation shows that $K^*(v) = C_{\mathrm{P}} v$, for $0 \leq v \leq a$, and

$$F_a(x) = \int_x^a \frac{\mathrm{d}v}{C_{\mathrm{P}} v}\,\mathrm{d}v = C_{\mathrm{P}}^{-1} \log\left(\frac{a}{x}\right),$$

from which we recover an exponential rate. However, $F_a^{-1}(n) = a\exp(-C_{\mathrm{P}} n)$ and since $\exp(-C_{\mathrm{P}} n) \geq (1 - C_{\mathrm{P}})^n$ because $-x/\sqrt{1-x} \leq \log(1-x) \leq -x$ for $x \in [0,1)$, this suggests a loss compared to a more direct method leading to (2). In this setting, we may also take $\Phi(f) = \|f\|_2^2$ and $a = 1$.

REMARK 12. By following the proof of Theorem 8 and stopping early, one can obtain bounds which are tighter but sometimes less convenient to work with.

For example, writing $T^{\circ n}$ for the n-fold composition of the map T with itself, one can obtain the bound

$$\frac{\|P^n f\|_2^2}{\Phi(f)} \leq \left(\mathrm{Id} - K^*\right)^{\circ n}\left(\frac{\|f\|_2^2}{\Phi(f)}\right), \tag{4}$$

and indeed a decay estimate of this form is equivalent to the original WPI holding with no loss of information; take $n = 1$. Going one step further in the proof, one can obtain the bound

$$\frac{\|P^n f\|_2^2}{\Phi(f)} \le F_a^{-1}\left(n + F_a\left(\frac{\|f\|_2^2}{\Phi(f)}\right)\right),$$

which is weaker than (4) due to the integral approximation, but stronger than the separable bound which is stated in the theorem.

A useful lemma we will make use of later concerning linear rescalings is the following.

LEMMA 13. *Let* $\tilde{\beta}(s) := c_1\beta(c_2 s)$ *for* $c_1, c_2 > 0$. *Then* $\tilde{K}^*(v) := \sup_{u\in\mathbb{R}_+} u[v - \tilde{\beta}(1/u)] = c_1 c_2 K^*(v/c_1)$ *and the corresponding function* $\tilde{F}_a(w) = c_2^{-1} F_{a/c_1}(w/c_1)$. *Furthermore, when* $c_1 \ge 1$, $\tilde{F}_a(w) \le c_2^{-1} F_a(w/c_1)$, *and we can conclude* $\tilde{F}_a^{-1}(x) \le c_1 F_a^{-1}(c_2 x)$.

2.1.2. *Examples of* $\beta(s)$ *and* $\gamma = F_a^{-1}$. Throughout the following examples (which coincide with those of [29], Corollary 2.4), we use the notation of Theorem 8.

LEMMA 14. *For* $\beta(s) = c_0 s^{-c_1}$, $K^*(v) = C(c_0, c_1) v^{1+c_1^{-1}}$. *Then with* F_∞ *as in Remark* 9, *the convergence rate is bounded by*

$$F_\infty^{-1}(n) \le c_0(1 + c_1)^{1+c_1} n^{-c_1}.$$

LEMMA 15. *Assume* $\beta(s) = \eta_0 \exp(-\eta_1 s^{\eta_2})$ *for* $\eta_0, \eta_1, \eta_2 > 0$ *and choose* $a > 0$. *Then there exist* $C > 0$, $0 < v_0 < 1 \wedge a$ *such that, for* $v \in [0, v_0]$,

$$K^*(v) \ge C v\left(\log\left(\frac{1}{v}\right)\right)^{-1/\eta_2}.$$

In addition, there exists $C' > 0$ *such that, for all* $n \in \mathbb{N}$,

$$F_a^{-1}(n) \le C' \exp\left(-\left(C\frac{1+\eta_2}{\eta_2} n\right)^{\eta_2/(1+\eta_2)}\right).$$

LEMMA 16. *Assume* $\beta(s) = c_0 \cdot (\log \max(c_1, s))^{-p}$ *for* $c_0 > 0$, $c_1 > 1$, $p > 0$. *Then there exist* $v_0 > 0$, $C > 0$ *such that, for* $v \in [0, v_0]$,

$$K^*(v) \ge C \cdot v^{1+1/p} \cdot \exp(-(v/c_0)^{-1/p}).$$

In addition, there exists $C' > 0$ *such that, for all* $n \in \mathbb{N}$,

$$F_a^{-1}(n) \le C' \cdot (\log \max(n, 2))^{-p}.$$

2.2. *Reversible case.* When the kernel P is reversible with respect to μ, we can derive a simplified weak Poincaré inequality in terms of P directly, rather than P^*P, making the approach much more practical. This kind of result seems to be new to the best of our knowledge, and indeed the need to handle P^*P is one of the key subtleties of our present discrete-time setting as opposed to the continuous-time setting. Furthermore, we can also derive a converse result; a weak Poincaré inequality is necessary for subgeometric convergence.

2.2.1. *Simplified weak Poincaré inequality.*

DEFINITION 17 (Weak Poincaré inequality; reversible case). Given a reversible Markov transition operator P on E, we will say that P *satisfies a weak Poincaré inequality* if, for any $f \in \mathrm{L}_0^2(\mu)$,

$$\|f\|_2^2 \le \alpha(r)\mathcal{E}(P, f) + r\Phi(f) \quad \forall r > 0,$$

where $\alpha : (0, \infty) \to [0, \infty)$ is a decreasing function, and $\Phi : \mathrm{L}^2(\mu) \to [0, \infty]$ is a functional satisfying: for any $f \in \mathrm{L}^2(\mu)$, $c > 0$ and $n \in \mathbb{N}$,

$$\Phi(cf) = c^2\Phi(f), \qquad \Phi\big(P^n f\big) \le \Phi(f), \qquad \|f - \mu(f)\|_2^2 \le a\Phi\big(f - \mu(f)\big),$$

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

A β*-weak Poincaré inequality for* P is analogously defined: for a function $\beta : (0, \infty) \to [0, \infty)$ decreasing with $\beta(s) \downarrow 0$ as $s \to \infty$, for any $f \in \mathrm{L}_0^2(\mu)$,

$$\|f\|_2^2 \le s\mathcal{E}(P, f) + \beta(s)\Phi(f) \quad \forall s > 0.$$

We are interested now to obtain an appropriate Poincaré inequality for $P^*P = P^2$ from a corresponding Poincaré inequality for P. The key complication is the left-hand side of the spectrum, around -1. In order to rule out periodic behaviour (which will prevent convergence), some assumptions on the spectrum in a neighbourhood of -1 are required.

LEMMA 18. *Suppose that the reversible kernel P possesses a left spectral gap: there exists some $0 < c_{\mathrm{gap}} \le 1$ such that the spectrum of P is bounded below:*

$$\inf \sigma(P) \ge -1 + c_{\mathrm{gap}}.$$

Then we obtain the bound on the Dirichlet forms given $f \in \mathrm{L}^2(\mu)$ by

$$\mathcal{E}\big(P^2, f\big) \ge c_{\mathrm{gap}}\mathcal{E}(P, f).$$

COROLLARY 19. *In the setting of Lemma 18, it immediately follows that if P satisfies a weak Poincaré inequality with function β as in Definition 17, P^2 satisfies a weak Poincaré inequality with $\tilde{\beta}$ given by $\tilde{\beta}(s) := \beta(c_{\mathrm{gap}}s)$:*

$$\|f\|_2^2 \le s\mathcal{E}\big(P^2, f\big) + \tilde{\beta}(s)\Phi(f) \quad \forall s > 0.$$

Thus the convergence rate $\tilde{F}_a^{-1}$ for $\|P^n f\|_2^2$ can be immediately deduced from Theorem 8 and Lemma 13.

When there is no left spectral gap, we can generalise the above results using a weak Poincaré inequality for $-P$.

THEOREM 20. *Suppose P is μ-reversible. Assume the following two weak-Poincaré inequalities hold: for all $s > 0$, $f \in \mathrm{L}_0^2(\mu)$:*

$$\|f\|_2^2 \le s\mathcal{E}(-P, f) + \beta_-(s)\Phi(f), \tag{5}$$

$$\|f\|_2^2 \le s\mathcal{E}(P, f) + \beta_+(s)\Phi(f). \tag{6}$$

Then the following weak Poincaré inequality for P^2 holds:

$$\|f\|_2^2 \le s\mathcal{E}\big(P^2, f\big) + \beta(s)\tilde{\Phi}(f), \tag{7}$$

for all $s > 0$, $f \in \mathrm{L}_0^2(\mu)$, *where*

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

Recall that a μ-reversible Markov kernel P is *positive* if for any $f \in \mathrm{L}^2(\mu)$, $\langle Pf, f\rangle \geq 0$, and a positive reversible kernel P has spectrum contained in the nonnegative interval $\sigma(P) \subset [0, 1]$. When P is reversible and positive, convergence of P^n can be straightforwardly derived as then $c_{\text{gap}} = 1$.

THEOREM 21. *Assume that the kernel P is reversible and positive and satisfies a weak Poincaré inequality as in Definition 17. Then Theorem 8 applies, so for $f \in \mathrm{L}_0^2(\mu)$ such that $0 < \Phi(f) < \infty$ and any $n \in \mathbb{N}$,*

$$\|P^n f\|_2^2 \leq \Phi(f) F_a^{-1}(n).$$

PROOF. Since P is reversible and positive, we can apply Corollary 19 with $c_{\text{gap}} = 1$ to see that P^2 satisfies a weak Poincaré inequality with the same function β. We can then immediately apply Theorem 8 to conclude. □

REMARK 22. Realistic MCMC kernels will all possess a nonzero left spectral gap. Indeed, popular methods such as the Independent Metropolis–Hastings sampler, many random walk Metropolis algorithms, and the resulting pseudo-marginal chains we will consider later are even positive [4], Proposition 16. Furthermore, a given reversible kernel P can be straightforwardly modified to possess a positive left spectral gap by considering the so-called lazy chain $Q := \epsilon\mathrm{Id} + (1 - \epsilon)P$ for $\epsilon \in [0, 1)$. Indeed, one of our contributions in Section 3 is to generalise this construction and give versions of Lemma 18 and Corollary 19 holding under weaker assumptions; see Theorem 40.

2.2.2. *Necessity of weak Poincaré inequalities.* We can also derive a converse to Theorem 8 in the reversible setting. For our explicit examples of β in Section 2.1.2 as well as in the geometric case (Remark 3), it turns out that we can derive a converse result, thus demonstrating, at least in the reversible setting, that our approach is able to recover the best possible rates of convergence for a given β when β is polynomial or polylogarithmic; see Remark 24. In the continuous-time setting, similar converse results have also been derived; see, for instance, [29], Theorem 2.3.

PROPOSITION 23. *Let P be μ-self-adjoint Markov transition operator and assume for any $n \in \mathbb{N}$, and $f \in \mathrm{L}_0^2(\mu)$,*

$$\|P^n f\|_2^2 \leq \gamma(n)\Phi(f), \tag{8}$$

for some functional $\Phi : \mathrm{L}^2(\mu) \to [0, \infty]$ satisfying (1) and a decreasing function $\gamma : \mathbb{R}_+ \to (0, \infty)$, with $\gamma(s) \to 0$ as $s \to \infty$. Then P^2 satisfies a weak Poincaré inequality (Definition 4) with

$$\beta(s) \leq \beta_1(s) := \sup_{t \geq s} \inf_{n \geq 2} \left\{ \frac{t^n}{(t-1)^{n-1}} \cdot \frac{(n-1)^{n-1}}{n^n} \cdot \gamma(n) \right\} \quad \forall s > 1,$$

which is decreasing and decreases to 0.

Similarly, suppose that under the same assumptions on (μ, f, Φ), *and only assuming* P *to be* μ*-invariant, it holds that*

$$\|P^n f\|_2^2 \le \Phi(f) \cdot F^{-1}\left(n + F\left(\frac{\|f\|_2^2}{\Phi(f)}\right)\right)$$

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* $\log(-\mathrm{D}F^{-1})$ *is convex.*

Then P^*P *satisfies a weak Poincaré inequality with*

$$K^*(v) \le K_1^*(v) = v - F^{-1}\big(1 + F(v)\big),$$

from which a corresponding β *can be deduced via convex duality.*

Finally, suppose that under the same assumptions on (μ, f, Φ), *and again only assuming* P *to be* μ*-invariant, it holds that*

$$\frac{\|P^n f\|_2^2}{\Phi(f)} \le \big(\mathrm{Id} - \tilde{K}^*\big)^{\circ n}\left(\frac{\|f\|_2^2}{\Phi(f)}\right),$$

for some function $\tilde{K}^* : [0, a] \to [0, a]$ *which is increasing, convex, and vanishes at* 0. *Then* P^*P *satisfies a weak Poincaré inequality with*

$$K^*(v) \le \tilde{K}^*(v),$$

from which a corresponding β *can again be deduced via convex duality.*

REMARK 24. For explicit computations, it can be useful to apply the elementary bounds

$$\frac{s^n}{(s-1)^{n-1}} = (s-1) \cdot \left(\frac{s}{s-1}\right)^n \le (s-1) \cdot \exp\left(\frac{n}{s-1}\right),$$

and

$$\frac{(n-1)^{n-1}}{n^n} = \frac{1}{n} \cdot \left(\frac{n-1}{n}\right)^{n-1} \le \frac{1}{2n},$$

to bound

$$\begin{aligned}
&\inf_{n \ge 2}\left\{\frac{s^n}{(s-1)^{n-1}} \cdot \frac{(n-1)^{n-1}}{n^n} \cdot \gamma(n)\right\} \\
&\quad \le \frac{1}{2} \cdot \inf_{n \ge 2}\left\{\gamma(n)\left(\frac{n}{s-1}\right)^{-1} \cdot \exp\left(\frac{n}{s-1}\right)\right\} \\
&\quad \le \frac{1}{2} \cdot \sup_{t \ge s} \inf_{n \ge 2}\left\{\gamma(n)\left(\frac{n}{t-1}\right)^{-1} \cdot \exp\left(\frac{n}{t-1}\right)\right\} =: \beta_2(s),
\end{aligned}$$

which is often more convenient to work with, and is again decreasing and decreases to 0.

We may consider the following procedure. Given

$$\|f\|_2^2 \le s\mathcal{E}\big(P^*P, f\big) + \beta(s)\Phi(f) \quad \forall s > 0,$$

apply our Theorem 8 to deduce that $\|P^n f\|_2^2 \le \Phi(f) \cdot \gamma(n)$. Then apply the above construction to show that P^*P satisfies

$$\|f\|_2^2 \le s\mathcal{E}\big(P^*P, f\big) + \beta_2(s)\Phi(f) \quad \forall s > 0,$$

with β_2 as above. In the examples considered in Section 2.1.2, we find that $\beta_2(s) \le c_1 \cdot \beta(c_2 \cdot s)$ for positive constants c_1, c_2; see the Supplementary Material for explicit calculations. We do not address whether this will hold in general. In particular, if β is polynomial or polylogarithmic then β_2 and β have the same asymptotic behaviour up to a multiplicative constant.

2.3. *Illustration: Independent MH sampler.* As a concrete illustration of the results of Section 2.2 we consider the Independent Metropolis–Hastings (IMH) algorithm. This has been studied previously in [20, 21] using drift/minorisation conditions, and we show in this subsection that we recover comparable subgeometric rates of convergence using weak Poincaré inequalities. We fix a target density π and a positive proposal density q on E and define

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

Then the IMH chain has reversible transition kernel P given by

$$P(x, \mathrm{d}y) = a(x, y)q(y)\,\mathrm{d}y + \rho(x)\delta_x(\mathrm{d}y),$$

where $a(x, y) = [1 \wedge \frac{w(y)}{w(x)}]$, and $\rho(x) = \int [1 - a(x, y)]q(\mathrm{d}y)$. In this case, it follows from reversibility that we have the following well-known representation.

LEMMA 25. *We can express*

$$\mathcal{E}(P, f) = \frac{1}{2} \int \pi(x)\pi(y)\big(w^{-1}(x) \wedge w^{-1}(y)\big)\big[f(y) - f(x)\big]^2 \,\mathrm{d}x\,\mathrm{d}y,$$

and

$$\|f\|_2^2 = \frac{1}{2} \int \pi(x)\pi(y)\big[f(y) - f(x)\big]^2 \,\mathrm{d}x\,\mathrm{d}y.$$

For the IMH, the following is known [20, 21].

PROPOSITION 26. *If w is uniformly bounded from above, then the IMH sampler is uniformly ergodic. However, if w is not uniformly bounded above, then the chain is not even geometrically ergodic.*

Thus since we are interested in the case of subgeometric convergence, we assume that w is not bounded from above, or equivalently, w^{-1} is not bounded from below by any positive constant. Thus given any $s > 0$, we define the following sets:

$$A(s) := \big\{(x, y) \in \mathsf{E} \times \mathsf{E} : w^{-1}(x) \wedge w^{-1}(y) \geq 1/s\big\}.$$

Since we are assuming the subgeometric case, there is no $s > 0$ for which $A(s) = \mathsf{E} \times \mathsf{E}$.

PROPOSITION 27. *For the IMH, we have the following weak Poincaré inequality: given any $f \in \mathrm{L}_0^2(\pi)$ and $s > 0$:*

$$\|f\|_2^2 \leq s\mathcal{E}(P, f) + \frac{\pi \otimes \pi(A(s)^{\complement})}{2}\|f\|_{\mathrm{osc}}^2.$$

Our bound in Proposition 27 allows us to directly link the tail properties of the weights $w(x)$ under π and the resulting rates of subgeometric convergence. We can apply Theorem 21, since the IMH kernel is always positive [18].

We turn now to some concrete examples inspired by [20, 21] where $\beta(s)$ can be evaluated directly.

2.3.1. *Exponential target and proposal case.* We work on $\mathsf{E} = (0, \infty) \subset \mathbb{R}$, and have target and proposal densities

$$\pi(x) = a_1 \exp(-a_1 x), \qquad q(x) = a_2 \exp(-a_2 x).$$

Since we are interested in the subgeometric case, we assume that $a_2 > a_1$. For this example, it was shown in [21], Proposition 9(b), that there is polynomial convergence, with rate at least $a_1/(a_2 - a_1)$.

LEMMA 28. *We have that for $s \geq 1$,*

$$\frac{\pi \otimes \pi(A(s)^{\complement})}{2} = \frac{1}{2}\big[1 - \big(1 - s^{-\frac{a_1}{a_2 - a_1}}\big)^2\big] \leq s^{-\frac{a_1}{a_2 - a_1}}.$$

In this case, we can make use of Lemma 14 to conclude the following, consistent with [21], Proposition 9(b).

PROPOSITION 29. *For our exponential example, we recover the convergence rate for some $C > 0$,*

$$\|P^n f\|_2^2 \leq C \|f\|_{\mathrm{osc}}^2 n^{-\frac{a_1}{a_2 - a_1}}.$$

2.3.2. *Polynomial target and proposal case.* We take $\mathsf{E} = [1, \infty) \subset \mathbb{R}$, and target and proposal densities

$$\pi(x) = \frac{b_1}{x^{1+b_1}}, \qquad q(x) = \frac{b_2}{x^{1+b_2}}.$$

We are interested in subgeometric convergence, so assume that $b_2 > b_1$. It was shown in [21], Proposition 9(a), that for this example there is polynomial convergence with rate at least $b_1/(b_2 - b_1)$. An entirely analogous calculation to the exponential example above allows us to conclude the following.

PROPOSITION 30. *For our polynomial example, we obtain the convergence rate, for some $C > 0$,*

$$\|P^n f\|_2^2 \leq C \|f\|_{\mathrm{osc}}^2 n^{-\frac{b_1}{b_2 - b_1}}.$$

3. Chaining Poincaré inequalities. In this section, we show how comparison of Dirichlet forms can be used to deduce convergence properties of a given Markov chain from another one. These results extend existing quantitative comparison results.

PROPOSITION 31. *Let P_1 and P_2 be two μ-invariant Markov kernels. Let $T_i = P_i$ or $T_i = P_i^* P_i$. Assume that for all $s > 0$ and $f \in \mathrm{L}_0^2(\mu)$,*

$$C_{\mathrm{P}} \|f\|_2^2 \leq \mathcal{E}(T_1, f),$$
$$\mathcal{E}(T_1, f) \leq s \mathcal{E}(T_2, f) + \beta'(s) \Phi(f),$$

where:

1. $C_{\mathrm{P}} > 0$ *and* $\beta' \colon (0, \infty) \to (0, \infty)$ *is decreasing and* $\beta'(s) \downarrow 0$ *as* $s \to \infty$,
2. $\Phi \colon \mathrm{L}^2(\mu) \to [0, \infty]$ *is such that* (1) *holds.*

Then for any $s > 0$,

$$\|f\|_2^2 \le s\mathcal{E}(T_2, f) + \beta(s)\Phi(f),$$

with $\beta(s) = \beta'(C_{\mathrm{P}}s)/C_{\mathrm{P}}$.

The proof is immediate.

REMARK 32. This generalises the comparison of Dirichlet forms used in [10], which corresponds to $\beta(s) = 0$ for all $s > \bar{s}$ for some $\bar{s} > 0$. Further, assume that for any $(x, A) \in \mathsf{E} \times \mathcal{F}$, $P_2(x, A \setminus \{x\}) \ge \varepsilon(x)P_1(x, A \setminus \{x\})$ for some $\varepsilon \colon \mathsf{E} \to (0, 1]$, then with $\{(x, y) \in \mathsf{E}^2 \colon \varepsilon(x)s > 1\}$ and $s > 0$ we have

$$\begin{aligned}\mathcal{E}(P_1, f) &\le \frac{1}{2}\int_{A(s)} \varepsilon(x)s\mu(\mathrm{d}x)P_1(x, \mathrm{d}y)\big[f(y) - f(x)\big]^2 \\ &\quad + \frac{1}{2}\int_{A(s)^\complement} \mu(\mathrm{d}x)P_1(x, \mathrm{d}y)\big[f(y) - f(x)\big]^2 \\ &\le s\mathcal{E}(P_2, f) + \frac{1}{2}\mu\big(\varepsilon^{-1}(X) \ge s\big)\|f\|_{\mathrm{osc}}^2,\end{aligned}$$

which is a generalisation of [10], Theorem A3, and together with Theorem 21 leads to a counterpart of [10], Theorem A1, for rates of convergence. However, we have not found an elegant generalisation of [10], Theorem A2, concerned with asymptotic variances. Theorem 36 further generalises these comparison ideas.

This can be further extended to the scenario where T_1 satisfies a weak Poincaré inequality.

THEOREM 33. *Let P_1 and P_2 be two μ-invariant Markov kernels. Let $T_i = P_i$ or $T_i = P_i^* P_i$. Assume that for all $s > 0$ and $f \in \mathrm{L}_0^2(\mu)$,*

$$\begin{aligned}\|f\|_2^2 &\le s\mathcal{E}(T_1, f) + \beta_1(s)\Phi_1(f), \\ \mathcal{E}(T_1, f) &\le s\mathcal{E}(T_2, f) + \beta_2(s)\Phi_2(f),\end{aligned} \tag{9}$$

where:

1. $\beta_1, \beta_2 \colon (0, \infty) \to (0, \infty)$ *are decreasing and* $\beta_1(s), \beta_2(s) \downarrow 0$ *as* $s \to \infty$,
2. $\Phi_1, \Phi_2 \colon \mathrm{L}^2(\mu) \to [0, \infty]$ *are such that* (1) *hold for* P_1 *and* P_2, *respectively*,
3. *for any* $n \in \mathbb{N}$ *and* $f \in \mathrm{L}_0^2(\mu)$, $\Phi_1(P_2^n f) \le \Phi_1(f)$.

Then for any $s > 0$,

$$\|f\|_2^2 \le s\mathcal{E}(T_2, f) + \beta(s)\Phi(f), \tag{10}$$

where $\Phi := \Phi_1 \vee \Phi_2$ *and*

$$\beta(s) := \inf\big\{s_1\beta_2(s_2) + \beta_1(s_1) | s_1 > 0, s_2 > 0, s_1 s_2 = s\big\}.$$

Furthermore, $\beta : (0, \infty) \to (0, \infty)$ is monotone decreasing and satisfies $\beta(s) \downarrow 0$ as $s \to \infty$, $\Phi(cf) = c^2\Phi(f)$ for $c > 0$ and $\Phi(P_2^n f) \le \Phi(f)$ for any $n \in \mathbb{N}$ and $f \in \mathrm{L}_0^2(\mu)$.

Additionally, writing $K_i(u) = u \cdot \beta_i(1/u)$, $K(u) = u \cdot \beta(1/u)$, it holds that

$$\begin{aligned}K(u) &= \inf\big\{K_2(u_2) + u_2 K_1(u_1) | u_1 > 0, u_2 > 0, u_1 u_2 = u\big\}, \\ K^*(v) &= K_2^* \circ K_1^*(v).\end{aligned}$$

PROOF. Fix $s > 0$. Given any $s_1, s_2 > 0$ with $s_1 s_2 = s$, by direct substitution in (9), we can arrive at

$$\begin{aligned}\|f\|_2^2 &\le s\mathcal{E}(T_2, f) + \beta_1(s_1)\Phi_1(f) + s_1\beta_2(s_2)\Phi_2(f)\\ &\le s\mathcal{E}(T_2, f) + [\beta_1(s_1) + s_1\beta_2(s_2)][\Phi_1(f) \vee \Phi_2(f)].\end{aligned}$$

Taking an infimum, we arrive at (10).

Now we prove the monotonicity of β. Fix some $s > 0$ and any $s_1, s_2 > 0$ with $s_1 s_2 = s$. Given any $s' \ge s$, note that

$$\begin{aligned}\beta(s') &\le s_1\beta_2(s'/s_1) + \beta_1(s_1)\\ &\le s_1\beta_2(s/s_1) + \beta_1(s_1)\\ &= s_1\beta_2(s_2) + \beta_1(s_1).\end{aligned}$$

Here, we made use of the fact that β_2 is a decreasing function. Taking an infimum over s_1, s_2, we conclude that $\beta(s') \le \beta(s)$.

We now show that given $\epsilon > 0$, we can find $s > 0$ such that $\beta(s) \le \epsilon$. Combined with monotonicity, this proves that $\beta(s) \downarrow 0$ as $s \to \infty$. So fix $\epsilon > 0$. Choose $s_1 > 0$ such that $\beta_1(s_1) \le \epsilon/2$, which can be done since $\beta_1(s) \downarrow 0$ as $s \to \infty$. Given such an s_1, now choose $s_2 > 0$ large enough so that $s_1\beta_2(s_2) \le \epsilon/2$. Thus for $s := s_1 s_2$ for these choices of s_1, s_2, we have shown that $\beta(s) \le \epsilon/2 + \epsilon/2 = \epsilon$.

To complete the proof, write

$$\begin{aligned}K(u) &= u \cdot \beta(1/u)\\ &= u \cdot \inf\{s_1\beta_2(s_2) + \beta_1(s_1) | s_1 > 0, s_2 > 0, s_1 s_2 = 1/u\}\\ &= u \cdot \inf\{(1/u_1) \cdot \beta_2(1/u_2) + \beta_1(1/u_1) | 1/u_1 > 0, 1/u_2 > 0, (1/u_1) \cdot (1/u_2) = 1/u\}\\ &= \inf\{(u/u_1) \cdot \beta_2(1/u_2) + (u_1 u_2) \cdot \beta_1(1/u_1) | u_1 > 0, u_2 > 0, u_1 u_2 = u\}\\ &= \inf\{u_2\beta_2(1/u_2) + u_2 \cdot u_1\beta_1(1/u_1) | u_1 > 0, u_2 > 0, u_1 u_2 = u\}\\ &= \inf\{K_2(u_2) + u_2 \cdot K_1(u_1) | u_1 > 0, u_2 > 0, u_1 u_2 = u\},\end{aligned}$$

as claimed. Finally,

$$\begin{aligned}K^*(v) &:= \sup_{u \ge 0}\{uv - K(u)\}\\ &= \sup_{u \ge 0}\Big\{uv - \inf_{u_1, u_2}\{K_2(u_2) + u_2 \cdot K_1(u_1)\}\Big\}\\ &= \sup_{u, u_1, u_2}\{uv - K_2(u_2) - u_2 \cdot K_1(u_1)\},\end{aligned}$$

where u_1, u_2 are again constrained to be nonnegative and have their product equal to u. Now, rewrite $u = u_1 u_2$ and eliminate the variable u to write

$$\begin{aligned}K^*(v) &= \sup_{u_1, u_2 > 0}\{u_1 u_2 v - K_2(u_2) - u_2 \cdot K_1(u_1)\}\\ &= \sup_{u_1, u_2 > 0}\{u_2 \cdot \{u_1 v - K_1(u_1)\} v - K_2(u_2)\}\\ &= \sup_{u_2 > 0}\Big\{u_2 \cdot \sup_{u_1 > 0}\{u_1 v - K_1(u_1)\} - K_2(u_2)\Big\}.\end{aligned}$$

Taking the inner supremum simplifies this expression to

$$K^*(v) = \sup_{u_2 > 0}\{u_2 \cdot K_1^*(v) - K_2(u_2)\},$$

and taking the remaining supremum allows us to conclude that $K^*(v) = K_2^*(w)$ with $w = K_1^*(v)$ as claimed, that is, $K^* = K_2^* \circ K_1^*$. $\square$

EXAMPLE 34. If one has $\beta_i(s) = c_i s^{-\alpha_i}$ for $i \in \{1, 2\}$, then $\beta(s) \propto s^{-\alpha_*}$ with $\alpha_* = \frac{\alpha_1\alpha_2}{1+\alpha_1+\alpha_2}$. To see this, write

$$\begin{aligned}
\beta(s) &:= \inf\{s_1\beta_2(s_2) + \beta_1(s_1) | s_1 > 0, s_2 > 0, s_1 s_2 = s\} \\
&= \inf\{c_2 s_1 s_2^{-\alpha_2} + c_1 s_1^{-\alpha_1} | s_1 > 0, s_2 > 0, s_1 s_2 = s\} \\
&= \inf\left\{c_2 s_1 \left(\frac{s_1}{s}\right)^{\alpha_2} + c_1 s_1^{-\alpha_1} \middle| s_1 > 0\right\} \\
&= \inf\{c_2 s^{-\alpha_2} s_1^{1+\alpha_2} + c_1 s_1^{-\alpha_1} | s_1 > 0\}.
\end{aligned}$$

Taking derivatives and solving for a stationary point gives $\hat{s}_1 = \frac{c_1\alpha_1}{c_2(1+\alpha_2)} s^{\frac{\alpha_2}{1+\alpha_1+\alpha_2}}$, from which point routine algebraic manipulations confirm the conclusion.

Our next main result is Theorem 36, which provides us with a practical way of establishing (9) for $T_i = P_i$. We first establish an intermediate result.

PROPOSITION 35. *Let P be a μ-invariant Markov kernel, and let $A \in \mathcal{F} \otimes \mathcal{F}$. Let $p \in (1, \infty]$, $q \geq 1$ satisfy $p^{-1} + q^{-1} = 1$. Then one can bound for $f \in \mathrm{L}^2(\mu)$,*

$$\int_A \mu(\mathrm{d}x) P(x, \mathrm{d}y) \big(f(x) - f(y)\big)^2 \leq \mu \otimes P\big(A \cap \{X \neq Y\}\big)^{1/q} \cdot \Phi_p(f),$$

with Φ_p given by

$$\Phi_p(f) := \begin{cases} 4\|f\|_{2p}^2 & p \in (1, \infty), \\ \|f\|_{\mathrm{osc}}^2 & p = \infty. \end{cases} \tag{11}$$

Moreover, it holds that for all $f \in \mathrm{L}^2(\mu)$ and $p \in [1, \infty]$, $\Phi_p(Pf) \leq \Phi_p(f)$.

PROOF. For $p \in (1, \infty)$, we use Hölder's inequality to write

$$\begin{aligned}
&\int_A \mu(\mathrm{d}x) P(x, \mathrm{d}y) \big(f(x) - f(y)\big)^2 \\
&\quad = \int \mu(\mathrm{d}x) P(x, \mathrm{d}y) \{\mathbb{I}_{A \cap \{x \neq y\}}(x, y) \cdot \big(f(x) - f(y)\big)^2\} \\
&\quad \leq \left(\int \mu(\mathrm{d}x) P(x, \mathrm{d}y) \mathbb{I}_{A \cap \{x \neq y\}}(x, y)\right)^{1/q} \cdot \left(\int \mu(\mathrm{d}x) P(x, \mathrm{d}y) \big|f(x) - f(y)\big|^{2p}\right)^{1/p}.
\end{aligned}$$

By Jensen's inequality, one can check that $|f(x) - f(y)|^{2p} \leq 2^{2p-1} \cdot (|f(x)|^{2p} + |f(y)|^{2p})$. Because $\mu P = \mu$,

$$\begin{aligned}
\int \mu(\mathrm{d}x) P(x, \mathrm{d}y) \big|f(x) - f(y)\big|^{2p} &\leq 2^{2p-1} \cdot \int \mu(\mathrm{d}x) P(x, \mathrm{d}y) \cdot \big(\big|f(x)\big|^{2p} + \big|f(y)\big|^{2p}\big) \\
&= 2^{2p} \cdot \|f\|_{2p}^{2p}.
\end{aligned}$$

One then concludes that

$$\begin{aligned}
\int_A \mu(\mathrm{d}x) P(x, \mathrm{d}y) \big(f(x) - f(y)\big)^2 &\leq \mu \otimes P(A)^{1/q} \cdot \big(2^{2p} \cdot \|f\|_{2p}^{2p}\big)^{1/p} \\
&= \mu \otimes P(A)^{1/q} \cdot \Phi(f),
\end{aligned}$$

as desired. The nonexpansivity of Φ under the action of P can be deduced by writing

$$\begin{aligned}\|Pf\|_{2p}^{2p} &= \int \mu(\mathrm{d}x)|Pf(x)|^{2p} \\ &= \int \mu(\mathrm{d}x)\left|\int P(x,\mathrm{d}y)f(y)\right|^{2p} \\ &\le \int \mu(\mathrm{d}x)P(x,\mathrm{d}y)|f(y)|^{2p} \\ &= \int \mu(\mathrm{d}y)|f(y)|^{2p} \\ &= \|f\|_{2p}^{2p},\end{aligned}$$

where the inequality uses Jensen's inequality against the probability measure $P(x,\cdot)$, and the penultimate equality uses the μ-invariance of P.

When $p=\infty$, we use an analogous argument, noting that $(f(x)-f(y))^2 \le \|f\|_{\mathrm{osc}}^2$ almost everywhere. $\square$

THEOREM 36. *Let P_1 and P_2 be two μ-invariant Markov kernels. Assume that for any $(x,A) \in \mathsf{E}\times\mathcal{F}$, $P_2(x, A\setminus\{x\}) \ge \int_{A\setminus\{x\}} \varepsilon(x,y)P_1(x,\mathrm{d}y)$ for some $\varepsilon\colon \mathsf{E}^2 \to (0,\infty)$. Then for any $p\in(1,\infty]$, $q\ge 1$ such that $p^{-1}+q^{-1}=1$, any $s>0$, and any $f\in \mathrm{L}_0^{2p}(\mu)\subset \mathrm{L}_0^2(\mu)$,*

$$\mathcal{E}(P_1,f) \le s\mathcal{E}(P_2,f) + \frac{1}{2}\cdot \mu\otimes P_1\big(A(s)^{\complement}\cap\{X\neq Y\}\big)^{1/q}\Phi_p(f),$$

with $A(s) := \{(x,y)\in\mathsf{E}^2\colon s\varepsilon(x,y)>1\}$ and $\Phi_p(f)$ as in (11), which satisfies (1).

PROOF. For any $s>0$, we have

$$\begin{aligned}\mathcal{E}(P_1,f) &\le \frac{1}{2}\int_{A(s)} s\varepsilon(x,y)\mu(\mathrm{d}x)P_1(x,\mathrm{d}y)[f(y)-f(x)]^2 \\ &\quad + \frac{1}{2}\int_{A(s)^{\complement}} \mu(\mathrm{d}x)P_1(x,\mathrm{d}y)[f(y)-f(x)]^2 \\ &\le \frac{s}{2}\int_{A(s)} \mu(\mathrm{d}x)P_2(x,\mathrm{d}y)[f(y)-f(x)]^2 \\ &\quad + \frac{1}{2}\cdot\mu\otimes P_1\big(A(s)^{\complement}\cap\{X\neq Y\}\big)^{1/q}\Phi_p(f),\end{aligned}$$

where we have used the assumed inequality between P_1 and P_2 and Proposition 35. $\square$

REMARK 37. Assume for simplicity that for μ-almost all x, $P_1(x,\cdot)\equiv P_2(x,\cdot)$, that is, $P_1(x,\cdot)$ and $P_2(x,\cdot)$ are equivalent measures. This implies $\mu\otimes P_1 \equiv \mu\otimes P_2$ and we may take $\varepsilon(x,y) = \frac{\mathrm{d}P_2(x,\cdot)}{\mathrm{d}P_1(x,\cdot)}(y) = \frac{\mathrm{d}\mu\otimes P_2}{\mathrm{d}\mu\otimes P_1}(x,y)$ to be positive $\mu\otimes P_1$-almost everywhere, and we can write

$$A(s)^{\complement}\cap\{(x,y): x\neq y\} = \left\{(x,y)\in\mathsf{E}^2 : \mathbb{I}\{x\neq y\}\frac{\mathrm{d}\mu\otimes P_1}{\mathrm{d}\mu\otimes P_2}(x,y)\ge s\right\}.$$

Hence, $\lim_{s\to\infty}\mu\otimes P_1(A(s)^{\complement}\cap\{X\neq Y\}) = 0$. Therefore, Theorem 36 covers many cases where P_2 places mass on the same sets as P_1. In fact, it covers slightly more general cases in which we only have $P_1(x, A\setminus\{x\})>0 \Rightarrow P_2(x,A\setminus\{x\})>0$ for μ-almost all x and all $A\in\mathcal{F}$.

REMARK 38. Our result concerned with the IMH algorithm in Section 2.3 is a particular case where $P_{\mathrm{IMH}}(x, A \setminus \{x\}) \geq \int_{A\setminus\{x\}}[w^{-1}(x) \wedge w^{-1}(y)]\pi(\mathrm{d}y)$, which combined with Proposition 31 with $C_{\mathrm{P}} = 1$ leads to Proposition 27.

REMARK 39. We note that this approach to identifying weak Poincaré inequalities can also be generalised to the setting of continuous-time Markov processes. To this end, consider a continuous-time Markov process with infinitesimal generator $\mathcal{L}$, and recall the definition of the carré du champ operator

$$\Gamma(f, g)(x) := \frac{1}{2}\{\mathcal{L}(fg) - (\mathcal{L}f) \cdot g - f \cdot (\mathcal{L}g)\}.$$

Note that Γ is bilinear and that for all suitable functions f, the function $\Gamma(f) := \Gamma(f, f)$ is pointwise nonnegative.

Suppose now that for two processes with the same invariant measure π and infinitesimal generators given by $\mathcal{L}_1$ and $\mathcal{L}_2$, respectively, their carré du champ operators can be ordered pointwise as

$$\Gamma_1(f)(x) \geq w(x) \cdot \Gamma_2(f)(x)$$

for some nonnegative function w (note that the subscripts here simply index the processes, and have no relation to so-called “iterated carré du champ” operators).

Defining $A(s) = \{x : s \cdot w(x) \geq 1\}$, one can then compute that

$$\begin{aligned}
\mathcal{E}(\mathcal{L}_2, f) &= \int \mu(\mathrm{d}x) \cdot \Gamma_2(f)(x) \\
&\leq s \cdot \int_{A(s)} \mu(\mathrm{d}x) \cdot w(x) \cdot \Gamma_2(f)(x) + \int_{A(s)^\complement} \mu(\mathrm{d}x) \cdot \Gamma_2(f)(x) \\
&\leq s \cdot \int \mu(\mathrm{d}x) \cdot \Gamma_1(f)(x) + \pi(A(s)^\complement) \cdot \sup_{x\in\mathsf{E}}\{\Gamma_2(f)(x)\} \\
&= s \cdot \mathcal{E}(\mathcal{L}_1, f) + \beta(s) \cdot \Phi(f),
\end{aligned}$$

where we have defined

$$\begin{aligned}
\beta(s) &:= \mu(A(s)^\complement), \\
\Phi(f) &:= \sup_{x\in\mathsf{E}}\{\Gamma_2(f)(x)\}.
\end{aligned}$$

Note that in many applications, $\Gamma(f)$ has the character of a squared gradient, and hence $\Phi(f)$ will behave much like a squared Lipschitz constant for the function f.

Comparisons of this form have been used implicitly in the study of so-called weighted and converse weighted Poincaré inequalities [8, 11], which are known to imply weak Poincaré inequalities. Such comparison inequalities can then be applied to compare the convergence of continuous-time processes in much the same fashion as in this work.

The following generalises the criterion $P(x, \{x\}) \geq \varepsilon$ for some $\varepsilon > 0$ and all $x \in \mathsf{E}$ often used to establish the existence of a left spectral gap for reversible Markov chains.

THEOREM 40. *Assume that P is μ-invariant and that for any $(x, A) \in \mathsf{E} \times \mathcal{F}$ it satisfies $P(x, A) \geq \varepsilon(x) \int_A \delta_x(\mathrm{d}y)$ for some $\varepsilon\colon \mathsf{E} \to [0, 1]$. Then*:

1. *for any $(x, A) \in \mathsf{E} \times \mathcal{F}$, $P^2(x, A) \geq \varepsilon(x)P(x, A)$,*

2. *for any* $p \in (1, \infty]$, $1/q = 1 - 1/p$, *any* $f \in \mathrm{L}_0^{2p}(\mu) \subset \mathrm{L}_0^2(\mu)$ *and* $s > 0$

$$\mathcal{E}(P, f) \le s\mathcal{E}\big(P^2, f\big) + \frac{1}{2} \cdot \mu\big(\varepsilon(X)^{-1} \ge s\big)^{1/q} \Phi_p(f).$$

COROLLARY 41. *Proposition 31 or Theorem 33 can be applied with $T_1 = P_1 = P$ and $T_2 = P_2 = P^2$. This can be applied to the Metropolis–Hastings (MH) algorithm (see (12)) as soon as $\mu(\rho(X) > 0) = 1$ and also means that weakly lazy chains can be defined as $\varepsilon(x)\mathrm{Id} + (1 - \varepsilon(x))\check{P}$ where $\check{P}$ is a MH using proposal P.*

PROOF OF THEOREM 40. For $(x, A) \in \mathsf{E} \times \mathcal{F}$,

$$P^2(x, A) = \int P(x, \mathrm{d}y)P(y, A) \ge \varepsilon(x)P(x, A)$$

and we apply Theorem 36. Now $\mu \otimes P(\{\varepsilon(X)^{-1} \ge s\} \cap \{X \ne Y\}) \le \mu(\varepsilon(X)^{-1} \ge s)$ and we conclude. □

Our final result in this section concerns the situation when one has a sequence of weak Poincaré inequalities given by functions $\{\beta_{2,\iota}\}_{\iota > 0}$, which converge pointwise to an appropriate function β_1. We give conditions under which the corresponding convergence rates $F_{2,\iota}^{-1}$ will also converge to the corresponding F_1^{-1}.

PROPOSITION 42. *Let P_1 and $(P_{2,\iota})_{\iota > 0}$ be μ-invariant Markov kernels. Assume P_1 satisfies a weak Poincaré inequality with function β_1 and that for any $\iota > 0$,*

$$\|f\|_2^2 \le s\mathcal{E}\big(P_{2,\iota}^* P_{2,\iota}, f\big) + \beta_{2,\iota}(s)\Phi(f) \quad \forall s > 0,$$

where each $\beta_{2,\iota}$ satisfies the conditions in Definition 4. Let $F_1, F_{2,\iota} : (0, a] \to [0, \infty)$ for each $\iota > 0$ be as defined in Section 2.

Assume that for any $\iota > 0$, $\beta_{2,\iota} \ge \beta_1$ pointwise and for any $s > 0$, $\lim_{\iota \to 0} \beta_{2,\iota}(s) = \beta_1(s)$. Then for any $\iota > 0$ and $n \in \mathbb{N}$, $F_{2,\iota}^{-1}(n) \ge F_1^{-1}(n)$ and

$$\lim_{\iota \to 0} \sup_{n \ge 0} \big\{F_{2,\iota}^{-1}(n) - F_1^{-1}(n)\big\} = 0.$$

PROOF. Let $v > 0$ and (u_n) be such that $K_1^*(v) = \lim_{n \to \infty} u_n[v - \beta_1(1/u_n)]$. Then for any $\iota > 0$ and any $n \ge 1$, $K_{2,\iota}^*(v) \ge u_n[v - \beta_{2,\iota}(1/u_n)]$ and, therefore,

$$\liminf_{\iota \to 0} K_{2,\iota}^*(v) \ge \lim_{\iota \to 0} u_n\big[v - \beta_{2,\iota}(1/u_n)\big] = u_n\big[v - \beta_1(1/u_n)\big].$$

Consequently,

$$\liminf_{\iota \to 0} K_{2,\iota}^*(v) \ge \lim_{n \to \infty} u_n\big[v - \beta_1(1/u_n)\big] = K_1^*(v).$$

Since for any $\iota > 0$, $\beta_{2,\iota} \ge \beta_1$ implies $K_{2,\iota}^* \le K_1^*$, we have $\limsup_{\iota \to 0} K_{2,\iota}^*(v) \le K_1^*(v)$. We therefore conclude that $\lim_{\iota \to 0} K_{2,\iota}^*(v) = K_1^*(v)$. Now let $0 < x \le a$, and choose $\epsilon > 0$ such that $K_1(x) - \epsilon > 0$. Then there exists $\iota_0 > 0$ such that, for any $0 < \iota \le \iota_0$, $K_{2,\iota}^*(x) \ge K_1(x) - \epsilon > 0$, and since $v \mapsto K_{2,\iota}^*(v)$ is increasing, we deduce $0 < \sup_{v \in [x,a]} (K_{2,\iota}^*(v))^{-1} \le (K_{2,\iota}^*(x))^{-1} \le (K_1^*(x) - \epsilon)^{-1} < \infty$. We can therefore apply the bounded convergence theorem and conclude

$$\lim_{\iota \to 0} \int_x^a \frac{\mathrm{d}v}{K_{2,\iota}^*(v)} = F_1(x).$$

For any $\iota > 0$, $F_1, F_{2,\iota}\colon (0,a] \to [0,\infty)$ are decreasing and continuous and so are the inverse functions $F_1^{-1}, F_{2,\iota}^{-1}\colon [0,\infty) \to (0,a]$, and consequently for any $x \in [0,\infty)$, $\lim_{\iota\to 0} F_{2,\iota}^{-1}(x) = F_1^{-1}(x)$ (note that $F_{2,\iota}^{-1}(0) = F_1^{-1}(0) = a$). Since $K_{2,\iota}^* \le K_1^*$, we immediately have the ordering $F_{2,\iota}^{-1}(x) \ge F_1^{-1}(x)$ for any $x \in [0,\infty)$.

Now let $\epsilon > 0$, then there exists $n_0 \in \mathbb{N}$ such that for any $n \ge n_0$, $F_1^{-1}(n) \le \epsilon/2$. From the convergence above, there exists $\iota_0 > 0$ such that, for any $0 < \iota \le \iota_0$,

$$0 \le F_{2,\iota}^{-1}(n_0) - F_1^{-1}(n_0) \le \epsilon/2,$$

and since $n \mapsto F_{2,\iota}^{-1}(n)$ is decreasing, for any $n \ge n_0$,

$$F_{2,\iota}^{-1}(n) - F_1^{-1}(n) \le F_{2,\iota}^{-1}(n) \le \epsilon.$$

Now, there exists $\iota_0' > 0$ such that, for any $0 < \iota < \iota_0'$,

$$\max_{0\le n<n_0} \{F_{2,\iota}^{-1}(n) - F_1^{-1}(n)\} \le \epsilon.$$

Therefore,

$$\lim_{\iota\to 0} \sup_{n\ge 0} \{F_{2,\iota}^{-1}(n) - F_1^{-1}(n)\} = 0.$$ □

EXAMPLE 43. Let $\beta_1(s) = a\mathbb{I}\{s \le C_{\mathrm{P}}^{-1}\}$, which corresponds to P_1 satisfying a strong Poincaré inequality as in Remark 11. Let $P_{2,\iota}$ satisfy a weak Poincaré inequality with $\beta_{2,\iota}(s) = a \wedge \{\beta_\iota'(C_{\mathrm{P}}s)/C_{\mathrm{P}}\}$ (e.g., by Proposition 31) where $\beta_\iota'(s) \ge \mathbb{I}\{s \le 1\}$ so that $\beta_{2,\iota} \ge \beta_1$. If $\lim_{\iota\to 0}\beta_\iota'(s) = \mathbb{I}\{s \le 1\}$ for all $s > 0$, then Proposition 42 can be applied and we recover exponential convergence as $\iota \to 0$, and for any $\epsilon > 0$ the existence of $\iota > 0$, such that $F_{2,\iota}^{-1}(n) - F_1^{-1}(n) < \epsilon$ for all n. In other words, one may obtain convergence for P_2 arbitrarily close to that given by β_1 by taking ι sufficiently small.

4. Application to pseudo-marginal methods. We now present our main application. Fix a probability distribution π on a measure space X, with a density function also denoted π. Pseudo-marginal algorithms [3] extend the scope of the Metropolis–Hastings algorithms to the scenario where the density π is intractable, but for any $x \in \mathsf{X}$, nonnegative estimators $\hat{\pi}(x)$ such that $\mathbb{E}[\hat{\pi}(x)] = C\pi(x)$ for some constant $C > 0$ are available. This can be conveniently formulated as $\tilde{\pi}(\mathrm{d}x, \mathrm{d}w) = \pi(\mathrm{d}x)Q_x(\mathrm{d}w)w = \pi(\mathrm{d}x)\tilde{\pi}_x(\mathrm{d}w)$ with $\int_{\mathbb{R}_+} wQ_x(\mathrm{d}w) = 1$ on an extended space $\mathsf{E} := \mathsf{X} \times \mathbb{R}_+$. We will refer to these auxiliary w random variables as *weights* or *perturbations*.

4.1. *A weak Poincaré inequality for pseudo-marginal chains.* A question of interest is to characterise the degradation in performance, compared to the *marginal algorithm*, which uses the exact density π. More specifically, for $\{q(x,\cdot), x \in \mathsf{X}\}$ a family of proposal distributions, the marginal algorithm is described by the kernel

$$\begin{aligned} P(x,\mathrm{d}y) &= [1 \wedge \mathfrak{r}(x,y)]q(x,\mathrm{d}y) + \delta_x(\mathrm{d}y)\rho(x), \\ &\text{where } \mathfrak{r}(x,y) := \frac{\pi(y)q(y,\mathrm{d}x)}{\pi(x)q(x,\mathrm{d}y)}, \end{aligned} \tag{12}$$

and ρ is the rejection probability given by $\rho(x) := 1 - \int [1 \wedge \mathfrak{r}(x,y)]q(x,\mathrm{d}y)$ for each $x \in \mathsf{X}$. For brevity, we will also define the acceptance probability as $a(x,y) := [1 \wedge \mathfrak{r}(x,y)]$.

The pseudo-marginal Metropolis–Hastings kernel is given by

$$\tilde{P}(x,w;\mathrm{d}y,\mathrm{d}u) = \left[1 \wedge \left\{\mathfrak{r}(x,y)\frac{u}{w}\right\}\right]q(x,\mathrm{d}y)Q_y(\mathrm{d}u) + \delta_{x,w}(\mathrm{d}y,\mathrm{d}u)\tilde{\rho}(x,w),$$

where the (joint) rejection probability $\tilde{\rho}(x, w)$ is analogously defined. It is known in this context that perturbing the acceptance ratio of the marginal algorithm leads to a loss in performance, in particular in terms of convergence rates to equilibrium. More specifically, if P is geometrically ergodic, then $\tilde{P}$ is geometrically ergodic if the perturbations are bounded uniformly in x, and cannot be geometrically ergodic if the perturbations are unbounded on a set of positive π-probability, which addressed in [4] in specific scenarios using Foster–Lyapunov and minorisation conditions by linking the existence of moments of the perturbations to the subgeometric rate of convergence of the algorithm. When the perturbations are bounded for each x but not bounded uniformly in x, the situation is more complicated: if "local proposals" are used then [24] proves that $\tilde{P}$ cannot be geometric under fairly weak assumptions in statistical applications whereas if global proposals are used $\tilde{P}$ may still be geometric (consider, for instance, the setting of [12], Remark 5). We show here that convergence results can be made completely general using weak Poincaré inequalities, with much simpler and considerably more transparent arguments.

We will be assuming throughout this section that the pseudo-marginal kernel $\tilde{P}$ is positive, in order to utilise our results from Section 2.2.1. We note that this positivity assumption is not restrictive; as established in [4], Proposition 16, $\tilde{P}$ will be positive if the marginal chain P is an Independent MH sampler, or a random walk Metropolis kernel with multivariate Gaussian or student-t increments.

The following comparison theorem plays a central role.

THEOREM 44. *Let $\bar{P}$ be the embedding of P in the joint space $\mathsf{E} = \mathsf{X} \times \mathbb{R}_+$,*

$$\bar{P}(x, w; \mathrm{d}y, \mathrm{d}u) := a(x, y)q(x, \mathrm{d}y)\tilde{\pi}_y(\mathrm{d}u) + \delta_{x,w}(\mathrm{d}y, \mathrm{d}u)\rho(x).$$

Then for any $p \in (1, \infty]$, $q \geq 1$ such that $p^{-1} + q^{-1} = 1$, any $s > 0$, and any $f \in \mathrm{L}^{2p}(\tilde{\pi}) \subset \mathrm{L}^2_0(\tilde{\pi})$,

$$\mathcal{E}(\bar{P}, f) \leq s\mathcal{E}(\tilde{P}, f) + \frac{1}{2}\cdot\Phi_p(f)\left(2\int_{\mathsf{X}} \tilde{\pi}_x(w \geq s)\pi(\mathrm{d}x)\right)^{1/q},$$

with $\Phi_p(f)$ given in (11).

PROOF. We apply Theorem 36. Let $\varepsilon(w, u) := w^{-1} \wedge u^{-1}$, then for any $(x, w) \in \mathsf{E}$ and $B \in \mathcal{F}$,

$$\begin{aligned}
&\int_{B\setminus\{x,w\}} \varepsilon(w, u)\bar{P}(x, w; \mathrm{d}y, \mathrm{d}u) \\
&\quad = \int_B q(x, \mathrm{d}y)\tilde{\pi}_y(\mathrm{d}u)a(x, y)\big(w^{-1} \wedge u^{-1}\big) \\
&\quad = \int_B q(x, \mathrm{d}y)Q_y(\mathrm{d}u)ua(x, y)\big(w^{-1} \wedge u^{-1}\big) \\
&\quad = \int_B q(x, \mathrm{d}y)Q_y(\mathrm{d}u)a(x, y)\left(1 \wedge \frac{u}{w}\right) \\
&\quad \leq \int_B q(x, \mathrm{d}y)Q_y(\mathrm{d}u)\left[1 \wedge \left(\mathfrak{r}(x, y)\frac{u}{w}\right)\right], \\
&\quad = \tilde{P}\big(x, w; B \setminus \{x, w\}\big),
\end{aligned}$$

where we have used that $1 \wedge (ab) \geq (1 \wedge a)(1 \wedge b)$ for $a, b \geq 0$. Now for $s > 0$ let

$$A(s) := \big\{(w, u) \in \mathbb{R}^2_+ \colon w^{-1} \wedge u^{-1} > 1/s\big\},$$

$$\bar{A}(s) := \big\{(x, w, y, u) \in \mathsf{E} \times \mathsf{E} \colon w^{-1} \wedge u^{-1} > 1/s\big\}.$$

Then

$$\mu \otimes \bar{P}\big(\bar{A}^{\complement}(s) \cap \{(X,W) \neq (Y,U)\}\big) \leq \int_{\bar{A}(s)^{\complement}} a(x,y)\pi(\mathrm{d}x)q(x,\mathrm{d}y)\tilde{\pi}_x(\mathrm{d}w)\tilde{\pi}_y(\mathrm{d}u)$$
$$= \int_{\mathsf{X}^2}\bigg[a(x,y)\int_{A(s)^{\complement}} \tilde{\pi}_x(\mathrm{d}w)\tilde{\pi}_y(\mathrm{d}u)\bigg]\pi(\mathrm{d}x)q(x,\mathrm{d}y),$$

and

$$\begin{aligned}\int_{A(s)^{\complement}} \tilde{\pi}_x(\mathrm{d}w)\tilde{\pi}_y(\mathrm{d}u) &= 1 - \tilde{\pi}_x(w \leq s)\tilde{\pi}_y(u \leq s)\\ &= 1 - \big[1 - \tilde{\pi}_x(w > s)\big]\big[1 - \tilde{\pi}_y(u > s)\big]\\ &\leq \tilde{\pi}_x(w \geq s) + \tilde{\pi}_y(u \geq s).\end{aligned}$$

Therefore,

$$\mu \otimes \bar{P}\big(\bar{A}(s)^{\complement} \cap \{(X,W) \neq (Y,U)\}\big) \leq 2\int \tilde{\pi}_x(w \geq s)\pi(\mathrm{d}x).$$

We conclude. □

We are now in a position to apply Proposition 31 or Theorem 33. We will see that the tail behaviour of the perturbations governs the rate at which our bound on $\|\tilde{P}^n f\|_2$ vanishes as $n \to \infty$. For simplicity, for the remainder of this section we focus on the case where $\Phi(f) = \|f\|_{\mathrm{osc}}^2$.

COROLLARY 45. *When $\bar{P}$ satisfies a strong Poincaré inequality with constant C_{P} as in Remark 3, Proposition 31 establishes that $\tilde{P}$ satisfies Definition 17 with $\beta(s) = \beta'(C_{\mathrm{P}}s)/C_P$ where $\beta'(s) = \int \tilde{\pi}_x(w \geq s)\pi(\mathrm{d}x)$ and $\Phi(f) = \|f\|_{\mathrm{osc}}^2$. Consequently, Theorem 21 applies to $\tilde{P}$ with a rate determined by $\beta(s)$.*

Furthermore, using Markov's inequality, the existence of moments of W under $\tilde{\pi}_x$ of order $k \in \mathbb{N}_*$ implies

$$\beta'(s) \leq s^{-k}\int_{\mathsf{X}} \mathbb{E}_{\tilde{\pi}_x}\big[|W|^k\big]\pi(\mathrm{d}x).$$

Provided the integral is finite, this leads to a polynomial rate of convergence $O(n^{-k})$ by Lemma 14.

Similarly, if $\bar{P}$ satisfies a weak Poincaré inequality, one can apply Theorem 33 and deduce the new rate of convergence as in Example 34.

REMARK 46. Notice that when the perturbations are uniformly bounded, that is, there exists $\bar{w}$ such that, for all $x \in \mathsf{X}$, $\tilde{\pi}_x(w \geq \bar{w}) = 0$, and $\bar{P}$ satisfies a strong Poincaré inequality, then $C_{\mathrm{P}}\|f\|^2 \leq \mathcal{E}(\bar{P}, f) \leq \bar{w}\mathcal{E}(\tilde{P}, f)$ and we recover the known results of [3, 4].

Examples of chains for which $\bar{P}$ satisfies a strong Poincaré inequality are numerous; the IMH and random walk Metropolis algorithms often possess a spectral gap; see [4] where these examples are considered in the context of pseudo-marginal algorithms.

We provide a general result demonstrating that under very weak conditions pseudo-marginal convergence can be made arbitrarily close to marginal convergence, strengthening the result of [3], Section 4.

REMARK 47. Assume that there is a parameter $\iota > 0$ controlling the quality of the perturbations $W \sim \tilde{\pi}_{x,\iota}$ such that, for each $x \in \mathsf{X}$, W converges in probability to 1 as $\iota \to 0$:

$$\lim_{\iota \to 0} \tilde{\pi}_{x,\iota}(w \geq s) = \mathbb{I}\{s \leq 1\}, \quad x \in \mathsf{X}, s > 0.$$

Let

$$\beta_\iota(s) := \int_{\mathsf{X}} \tilde{\pi}_{x,\iota}(w \geq s)\pi(\mathrm{d}x),$$

then the bounded convergence theorem implies that

$$\lim_{\iota \to 0} \beta_\iota(s) = \mathbb{I}\{s \leq 1\}, \quad s > 0.$$

Assume now that $\bar{P}$ satisfies a weak Poincaré inequality with function $\bar{\beta}$. Similar to Example 43, one can compare the convergence bounds for $\tilde{P}_\iota$ and $\bar{P}$ via their respective functions $\tilde{\beta}_\iota$ and $\bar{\beta}$. Indeed, Theorem 44 and Theorem 33 imply that $\tilde{P}_\iota$ satisfies a weak Poincaré inequality with function

$$\tilde{\beta}_{\iota,\epsilon}(s) = \frac{s}{1+\epsilon}\beta_\iota(1+\epsilon) + \bar{\beta}\left(\frac{s}{1+\epsilon}\right),$$

where $\epsilon > 0$ is arbitrary. Note that for $s > 0$, $\tilde{\beta}_{\iota,\epsilon}(s) \geq \bar{\beta}(s/(1+\epsilon)) \geq \bar{\beta}(s)$ and since $\lim_{\iota \to 0} \tilde{\beta}_\iota(s) = \bar{\beta}(s/(1+\epsilon))$ we can apply Proposition 42 to obtain convergence bounds arbitrarily close to those of $\bar{P}$ with rate function $\bar{\beta}$.

4.2. *The effect of averaging*. A natural idea to reduce the variability of pseudo-marginal chains is to average several estimators $\hat{\pi}$ of the target density at each iteration. As pointed out in [4], this is unlikely to affect asymptotic rates of convergence. Furthermore, it was established in [9, 31] that when considering asymptotic variance, it is preferable to combine the output of N independent chains each using 1 estimator, rather than running 1 chain averaging N estimators at each iteration. The following, motivated by the application of Markov's inequality, adds nuance to these conclusions by showing how bias can be reduced by averaging, particularly in situations where higher order moments of the perturbations are large.

LEMMA 48. *Let $\{W_i\}$ be i.i.d., of expectation 1 and such that, for a given $p \in \mathbb{N}$ with $p \geq 2$, $\mathbb{E}(|W_1|^p) < \infty$. Then there are some constants $\{C_{p,k}\}$, such that, for any $N \in \mathbb{N}$,*

$$\mathbb{E}\left[\left|\frac{1}{N}\sum_{i=1}^{N} W_i\right|^p\right] \leq 1 + \sum_{k=2}^{p} N^{-k/2} C_{p,k}\mathbb{E}[|W_1 - 1|^k]. \tag{13}$$

For large N, this bound is $1 + O(N^{-1})$.

As an illustration, we focus here on the scenario where the marginal chain satisfies a strong Poincaré inequality (Remark 3) and the moments are uniformly bounded in $x \in \mathsf{X}$. Let $\mathcal{W}_N := N^{-1}\sum_{i=1}^{N} W_i$, then Markov's inequality implies that for the pseudo-marginal algorithm which averages N estimators,

$$\beta'_N(s) \leq \left[\sup_{x \in \mathsf{X}} \mathbb{E}_{\tilde{\pi}_x}(\mathcal{W}_N^p)\right] s^{-p},$$

and while the rate of convergence in s is independent of N, the multiplicative constant in square brackets does depend on N. Indeed, by Lemma 48, averaging by choosing $N > 1$ can reduce its magnitude and reduce our convergence upper bounds in Theorem 21, thanks to Lemma 13. The bound obtained in (13) suggests that while the asymptotic rate of decay for large N is governed by the term $\mathbb{E}[|W_1 - 1|^2]N^{-1}$, inversely proportional to the increased computational cost at each MCMC iteration, higher order moments may play an important role for small to moderate values of N.

This is expected for heavy-tailed distributions, for example, consider an inverse gamma distribution of expectation 1 and shape parameter $\mathsf{s} > 1$. Its raw (polynomial) moments grow very rapidly as $\mathbb{E}(W_1^k) = (\mathsf{s}-1)^k / \prod_{i=1}^k (\mathsf{s}-i)$ for $k \in \mathbb{N}$, $k < \mathsf{s}$ and s large, and for small and moderate values of N, summands other than $k = 2$ in (13) will be most prominent.

4.3. *ABC example*. We consider an Approximate Bayesian Computation (ABC) setting, using notation inspired by [24]. We assume we have a true posterior density $\pi_0(x) \propto \nu(x)\ell_y(x)$ on a space $\mathsf{X} \subset \mathbb{R}^{d_x}$, where $\nu(\cdot)$ represents the prior and $x \mapsto \ell_y(x)$ is an intractable likelihood corresponding to a probability density $f_x(y) = \ell_y(x)$ for some fixed observations $y \in \mathsf{Y} \subset \mathbb{R}^{\mathrm{d}}$. It is known that ABC Markov chains of the type considered here cannot be geometrically ergodic under fairly weak conditions when a "local proposal" is used [24], Theorem 2.

Fix an $\epsilon > 0$ and $x \in \mathsf{X}$, and for $j = 1, \dots, N$, let $z_j \overset{\text{i.i.d.}}{\sim} f_x(\cdot)$ be auxiliary random variables and define the random variables W_j, where $|\cdot|$ denotes the Euclidean norm,

$$W_j = \begin{cases} 1/\ell_{\mathrm{ABC}}(x) & \text{if } |z_j - y| < \epsilon, \\ 0 & \text{else,} \end{cases}$$

with $\ell_{\mathrm{ABC}}(x) := \mathbb{P}_x(|z_1 - y| < \epsilon)$. In an ABC setup, the intractable π_0 is replaced with the ABC posterior $\pi(x) \propto \nu(x)\ell_{\mathrm{ABC}}(x)$, which is typically also intractable and itself approximated using a pseudo-marginal approach: for fixed $N \in \mathbb{N}$, define

$$\tilde{\pi}(x, z_1, \dots, z_N) \propto \nu(x)\ell_{\mathrm{ABC}}(x)\left[\prod_{j=1}^N f_x(z_j)\right] \cdot \frac{1}{N}\sum_{j=1}^N W_j.$$

It is easily seen that $\mathcal{W}_N := \frac{1}{N}\sum_{j=1}^N W_j$ has expectation 1 under $[\prod_{j=1}^N f_x(z_j)]\,\mathrm{d}z_1 \times \cdots \times \mathrm{d}z_N$ for a fixed $x \in \mathsf{X}$. In our previous notation, $Q_x(\mathrm{d}w)$ is then the law of $\mathcal{W}_N$ when the $(z_1, \dots, z_N)$ are drawn from $[\prod_{j=1}^N f_x(z_j)]\,\mathrm{d}z_1 \times \cdots \times \mathrm{d}z_N$, and $\tilde{\pi}_x(\mathrm{d}w) = wQ_x(\mathrm{d}w)$. Given $x \in \mathsf{X}$, it is clear that under Q_x, we have that $\ell_{\mathrm{ABC}}(x)\sum_{j=1}^N W_j \sim \mathrm{Bin}(N, \ell_{\mathrm{ABC}}(x))$.

Thus from our previous result, Corollary 45, in order to bound the rate of convergence of the resulting pseudo-marginal algorithm, we need to bound for $s > 0$,

$$\int_{\mathsf{X}} \pi(\mathrm{d}x)\tilde{\pi}_x(\mathcal{W}_N \geq s).$$

So given $x \in \mathsf{X}$, $s > 0$, we first consider $\tilde{\pi}_x(\mathcal{W}_N \geq s)$. Using Markov's inequality, for any $p \in \mathbb{N}$, we can bound

$$\tilde{\pi}_x(\mathcal{W}_N \geq s) \leq \frac{\tilde{\pi}_x[\mathcal{W}_N^p]}{s^p} = \frac{Q_x[\mathcal{W}_N^{p+1}]}{s^p}.$$

This seems to suggest that if the marginal algorithm is geometrically ergodic, then its ABC approximation converges at any polynomial rate. The following result tells us that this may not be the case.

PROPOSITION 49. *For a given $p \in \mathbb{N}$, suppose that $\int_{\mathsf{X}} \nu(x)\ell_{\mathrm{ABC}}^{-(p-1)}(x)\,\mathrm{d}x < \infty$. Then there is $C_{N,p} > 0$ such that, for all $s > 0$,*

$$\int_{\mathsf{X}} \pi(\mathrm{d}x)\tilde{\pi}_x(\mathcal{W}_N \geq s) \leq C_{N,p}s^{-p},$$

and as $N \to \infty$, $C_{N,p} = 1 + O(1/N)$. In particular, we may always take $p = 1$. The resulting convergence rate for the pseudo-marginal chain is then also $O(n^{-p})$ as in Lemma 14.

4.4. *Products of averages*. The results in Sections 4.2–4.3 suggest that N may not need to be taken too large in the case of simple averaging. We consider here a scenario where the perturbation is instead a product of T independent averages, which gives different conclusions, and can be seen as a simple version of the perturbation involved in a particle marginal Metropolis–Hastings (PMMH) algorithm [1], a special case of a pseudo-marginal algorithm. Such scenarios can arise in random effects and latent variable models. For example, [34], Section 4.1, uses a random effects model from [14], Section 6.1, to analyse the data from [19], while [25], Section 4.2, considers an ABC example with i.i.d. data and [25], Section 4.3, considers a single-cell gene expression model proposed by [27] and employed, for example, by [33].

The following bound can be used in Corollary 45, and indicates that it is sufficient to take N proportional to T to obtain T-independent bounds on the relevant tail probabilities as long as π is sufficiently concentrated.

PROPOSITION 50. *Assume $W \sim Q_x$ can be written as $W = \prod_{t=1}^{T} W_t$, where each W_t is independent and nonnegative, and for each $t \in \{1, \dots, T\}$,*

$$W_t = \frac{1}{N} \sum_{i=1}^{N} W_{t,i},$$

is an average of nonnegative, identically distributed random variables with mean 1. *Assume that for some $p \in \mathbb{N}$ with $p \geq 2$, and any $x \in \mathsf{X}$,*

$$\max_{t \in \{1,\dots,T\}} \mathbb{E}[W_{t,1}^{p}] < \infty.$$

Then there exists a function $M_p : \mathsf{X} \to \mathbb{R}_+$ such that if we take

$$N \geq \alpha T + \frac{1}{2} + \sqrt{\alpha T},$$

for some $\alpha > 0$, then

$$\int \pi(\mathrm{d}x) \tilde{\pi}_x(W \geq s) \leq s^{-p+1} \int \pi(\mathrm{d}x) \exp\left(\frac{M_p(x)}{\alpha}\right),$$

where the right-hand side may be finite or infinite depending on π.

In particular, we can see that if the function M_p grows quickly in the tails of π, then the bound is finite only if π has sufficiently light tails.

EXAMPLE 51. Assume $M_p(x) = bx^k$ and $\pi(\mathrm{d}x) \propto \mathbb{I}_{\mathbb{R}_+}(x) \exp(-cx^{\ell})\mathrm{d}x$ for some $k, \ell \geq 0$. If $\ell < k$ then the integral $\int \pi(\mathrm{d}x) \exp(M_p(x)/\alpha)$ in Proposition 50 is infinite. If $\ell > k$, then the integral is finite. If $\ell = k$, then the integral is finite if and only if $\alpha > b/c$.

4.5. *Log-normal example*. We consider now a limiting case of the perturbations in a PMMH algorithm, motivated by [7], Theorem 1.1, which has also been analysed using other techniques [16, 32]. The result of [7] concerns a particular mean 1 perturbation $W_{T,N}$ that is also a product of T averages, with N random variables involved in each average, but where the random variables are not independent. They show that, under regularity conditions, if $N = \alpha T$ there is a σ_0^2 such that with $\sigma^2 = \sigma_0^2/\alpha$, $\log(W_{T,N})$ converges in distribution to $N(-\frac{1}{2}\sigma^2, \sigma^2)$ as $T \to \infty$.

We consider here the setting where for some large T, the log-perturbation is exactly $N(-\frac{1}{2}\sigma^2, \sigma^2)$, in which case one can think of $\sigma^2 = \sigma_0^2 T/N$, and since the precise value of T is not relevant we suppress it in the sequel. To be explicit, we have that W has law

$$Q_x(\mathrm{d}w) = \frac{1}{w\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log w + \sigma^2/2)^2}{2\sigma^2}\right) \mathrm{d}w,$$

where $\mathrm{d}w$ is Lebesgue measure on $\mathbb{R}_+$ and $\sigma > 0$ is a variance parameter, which we assume for simplicity is independent of x. We note that a pseudo-marginal kernel with log-normal perturbations can only converge subgeometrically, since the perturbations are not bounded.

4.5.1. *Tail probabilities and convergence bound.*

LEMMA 52. *We have the bound, for $s > 0$, $\tilde{\pi}_x(W \geq s) \leq \beta(s)$, where*

$$\beta(s) := \exp\left(-\frac{((\log s - \sigma^2/2)_+)^2}{2\sigma^2}\right). \tag{14}$$

REMARK 53. Note that despite (14) being an upper bound on the quantity of interest, it satisfies the conditions in Example 43 and Remark 47, implying that the rate of convergence of the marginal algorithm is recovered in the limit, as $\sigma \to 0$.

Although it is theoretically possible to work directly with β as in (14), in order to derive clean and practically useful tuning guidelines, we now derive some tractable bounds on the corresponding convergence rate.

LEMMA 54. *We have a lower bound on the convex conjugate, for $0 < v < 1$, $K^*(v) \geq \frac{v}{2}\exp(-\sigma\sqrt{-2\log\frac{v}{2}} - \sigma^2/2)$.*

PROOF. This is immediate from choosing $u = \exp(-\sigma\sqrt{-2\log\frac{v}{2}} - \sigma^2/2)$ in the definition of the convex conjugate, $K^*(v) = \sup_{u>0}\{uv - u\beta(1/u)\}$. $\square$

As before, we define $F(w) := \int_w^1 \frac{\mathrm{d}v}{K^*(v)}$. We are able to deduce the following convergence bound.

LEMMA 55. *We have the upper bound for $x > 0$,*

$$F^{-1}(x) \leq 2\exp\left\{-\frac{1}{2\sigma^2}\mathsf{W}^2\left(\frac{x\sigma^2}{2\exp(\sigma^2/2)}\right)\right\}, \tag{15}$$

where W^2 denotes the Lambert function squared.

PROPOSITION 56. *Assume that the marginal chain $\bar{P}$ satisfies a strong Poincaré inequality with constant C_{P} as in Remark 3. Then the final convergence bound for the pseudo-marginal chain is given by*

$$F_{\mathrm{PM}}^{-1}(n) \leq \frac{2}{C_{\mathrm{P}}}\exp\left\{-\frac{1}{2\sigma^2}\mathsf{W}^2\left(\frac{C_{\mathrm{P}}n\sigma^2}{2\exp(\sigma^2/2)}\right)\right\}. \tag{16}$$

PROOF. This is immediate from Corollary 45, Lemma 55. $\square$

4.5.2. *Mixing times.* It is possible to obtain mixing time type results.

PROPOSITION 57. *Let $\epsilon \in (0, 1]$ and $\sigma^2 > 0$. Then, to obtain $F_{\mathrm{PM}}^{-1}(n) \le \epsilon^2$, it is sufficient to take*

$$n \ge \frac{2\sqrt{H(\epsilon)}}{C_{\mathrm{P}}\sigma} \exp\left(\frac{\sigma^2}{2} + \sqrt{H(\epsilon)}\sigma\right), \tag{17}$$

where $H(\epsilon) = 2\log(2/(\epsilon^2 C_{\mathrm{P}}))$.

PROOF. One can calculate directly that the bound of F_{PM}^{-1} in (15) evaluated at the right-hand side of (17) is equal to $2\exp(-H(\epsilon)/2)/C_{\mathrm{P}}$, which combined with the definition of $H(\epsilon)$ and the monotonicity of F_{PM}^{-1} gives the result. $\square$

Now we consider the minimum computational budget required to achieve a given precision of ϵ, and the corresponding split between the number of MCMC iterations n and the number of particles N. The budget required is significantly lower than the result in Proposition 57 would imply for a fixed N and, therefore, σ^2.

PROPOSITION 58. *Let $\epsilon \in (0, 1]$. For simplicity, let $\bar{n}$ and $\bar{N}$ be real-valued counterparts of n and N, respectively. The "budget" function $(\bar{n}, \bar{\sigma}) \mapsto \mathsf{B}(\bar{n}, \bar{\sigma}) = \bar{n}\bar{N} = \bar{n}\sigma_0^2/\bar{\sigma}^2$ on $\mathbb{R}_+^2$ is minimised subject to the constraint $F_{\mathrm{PM}}^{-1}(\bar{n}; \bar{\sigma}) = \epsilon^2$ (with F_{PM}^{-1} as in (15)) when*

$$\bar{\sigma} = \bar{\sigma}_\star(\epsilon) := \frac{\sqrt{H(\epsilon) + 12} - \sqrt{H(\epsilon)}}{2},$$

where $H(\epsilon) := 2\log(\frac{2}{C_{\mathrm{P}}\epsilon^2}) > 0$ and $\lim_{H(\epsilon)\to\infty} \sqrt{H(\epsilon)}\bar{\sigma}_\star(\epsilon) = 3$. Moreover, for $\epsilon > 0$ such that $H(\epsilon) \ge 1$, we obtain $F_{\mathrm{PM}}^{-1}(\bar{n}; \bar{\sigma}) = \epsilon^2$ with $\bar{\sigma}(\epsilon) = 3/\sqrt{H(\epsilon)}$,

$$\bar{N}(\epsilon) = \frac{2}{9}\sigma_0^2 \log\left(\frac{2}{C_{\mathrm{P}}\epsilon^2}\right),$$

$$\bar{n}(\epsilon) \le \frac{4\exp(15/2)}{3C_{\mathrm{P}}} \log\left(\frac{2}{C_{\mathrm{P}}\epsilon^2}\right),$$

$$\mathsf{B}(\epsilon) \le \frac{8\sigma_0^2 \exp(15/2)}{27C_{\mathrm{P}}} \log\left(\frac{2}{C_{\mathrm{P}}\epsilon^2}\right)^2,$$

which is asymptotically accurate and optimal as $H(\epsilon) \to \infty$, that is, if $\epsilon \downarrow 0$ or $C_{\mathrm{P}} \downarrow 0$, except that the constant factors $\exp(15/2)$ *will tend to* $\exp(3)$.

These nonasymptotic results take into account both C_{P} and σ_0^2 in a natural manner and are easily interpretable. We note that they also indicate how a given computational budget B should be split between N and n in order to achieve best precision: in particular N should increase as B increases. This is to be contrasted with results (see [16, 32] and below) concerned with the asymptotic variance, which recommend a fixed number of particles for any B sufficiently large and allocation of the remaining resources to iterating the MCMC algorithm for this fixed number of particles.

4.5.3. *Asymptotic variance.* We now show that our bounds lead to recommendations for N similar to those of [16, 32] when considering the asymptotic variance as a criterion. We can use the bound (16) to give an upper bound on the resulting asymptotic variance.

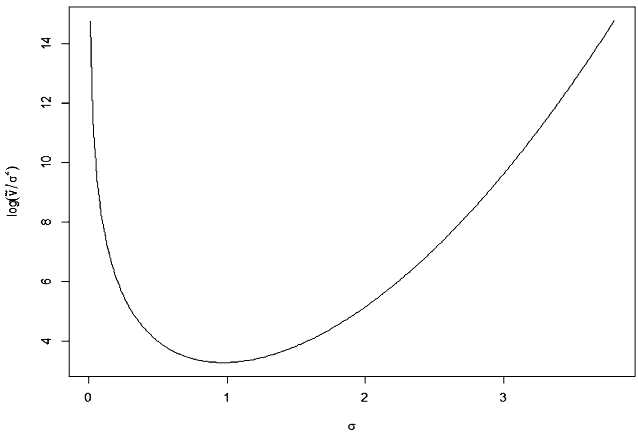

FIG. 1. *A plot of the function* $\sigma \mapsto \log(\tilde{v}(\sigma)/\sigma^2)$ *in the case* $C_\mathrm{P} = 1$.

LEMMA 59. *Fix a test function $f \in \mathrm{L}_0^2(\mu)$. In the setting of Theorem 21, for a reversible Markov kernel P, the asymptotic variance $v(f, P)$ is bounded by*

$$v(f, P) \le -\|f\|_2^2 + 4\Phi(f) \sum_{n=0}^{\infty} F^{-1}(n).$$

EXAMPLE 60. For our log-normal pseudo-marginal example, we can then ask for a given f, how to tune σ in order to minimise the resulting bound on the asymptotic variance. We can bound

$$\begin{aligned}\sum_{n=1}^{\infty} F_{\mathrm{PM}}^{-1}(n) &\le \sum_{n=1}^{\infty} \frac{2}{C_\mathrm{P}} \exp\left\{-\frac{1}{2\sigma^2}\mathsf{W}^2\left(\frac{C_\mathrm{P} n \sigma^2}{2\exp(\sigma^2/2)}\right)\right\} \\ &\le \frac{2}{C_\mathrm{P}} \int_0^{\infty} \exp\left(-\mathsf{a}\mathsf{W}^2(\mathsf{b}x)\right) \mathrm{d}x,\end{aligned}$$

where $\mathsf{a} := 1/(2\sigma^2)$ and $\mathsf{b} := C_\mathrm{P}\sigma^2/(2\exp(\sigma^2/2))$. Here, we used the fact that the Lambert function is increasing. Through routine calculations and making use of the substitution $\mathsf{b}x = u\exp(u) \Leftrightarrow u = \mathsf{W}(\mathsf{b}x)$, this integral can be simplified and written as

$$\tilde{v}(\sigma) := \frac{1}{\mathsf{b}}\left[\exp\left(1/(4\mathsf{a})\right)\left(1 + \frac{1}{2\mathsf{a}}\right)\mathsf{a}^{-1/2} \int_{-\mathsf{a}^{-1/2}/2}^{\infty} \exp\left(-w^2\right) \mathrm{d}w + \frac{1}{2\mathsf{a}}\right].$$

In this final expression, both a and b depend on σ, and the resulting function of $\sigma \mapsto \tilde{v}(\sigma)/\sigma^2$ can be optimised numerically, where we divide by σ^2 to take into account the additional computational cost; see Figure 1. Note that the optimal value σ_* of σ does not depend on C_P, and we find numerically that $\sigma_* \approx 0.973$. This is consistent with [16] who report optimal values in the range $\sigma_* \approx 1.0 - 1.7$ (dependent on the performance of the marginal algorithm) using another bound on the asymptotic variance, while [32] find $\sigma_* \approx 1.812$ using a scaling and diffusion approximation.

Acknowledgements. We would like to thank Gareth Roberts and Chris Sherlock, as well as the two anonymous referees and Associate Editor for useful comments which have improved the article.

Funding. Research of CA, AL and AQW supported by EPSRC grant "CoSInES (COmputational Statistical INference for Engineering and Security)" (EP/R034710/1), and research of CA and SP supported by EPSRC grant Bayes4Health, "New Approaches to Bayesian Data Science: Tackling Challenges from the Health Sciences" (EP/R018561/1).

SUPPLEMENTARY MATERIAL

Supplement to: "Comparison of Markov chains via weak Poincaré inequalities with application to pseudo-marginal MCMC" (DOI: 10.1214/22-AOS2241SUPP; .pdf). This document contains a list of notation, deferred proofs and some auxiliary results.

REFERENCES

[1] ANDRIEU, C., DOUCET, A. and HOLENSTEIN, R. (2010). Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 269–342. MR2758115 https://doi.org/10.1111/j.1467-9868.2009.00736.x

[2] ANDRIEU, C., LEE, A., POWER, S. and WANG, A. Q. (2022). Supplement to "Comparison of Markov chains via weak Poincaré inequalities with application to pseudo-marginal MCMC." https://doi.org/10.1214/22-AOS2241SUPP

[3] ANDRIEU, C. and ROBERTS, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.* **37** 697–725. MR2502648 https://doi.org/10.1214/07-AOS574

[4] ANDRIEU, C. and VIHOLA, M. (2015). Convergence properties of pseudo-marginal Markov chain Monte Carlo algorithms. *Ann. Appl. Probab.* **25** 1030–1077. MR3313762 https://doi.org/10.1214/14-AAP1022

[5] BAKRY, D., BOLLEY, F. and GENTIL, I. (2010). Around Nash inequalities. In *Journées équations aux dérivées partielles* 1–16. Available at www.numdam.org/item/10.5802/jedp.59.pdf. https://doi.org/10.5802/jedp.59

[6] BAKRY, D., CATTIAUX, P. and GUILLIN, A. (2008). Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. *J. Funct. Anal.* **254** 727–759. MR2381160 https://doi.org/10.1016/j.jfa.2007.11.002

[7] BÉRARD, J., DEL MORAL, P. and DOUCET, A. (2014). A lognormal central limit theorem for particle approximations of normalizing constants. *Electron. J. Probab.* **19** no. 94, 28. MR3272327 https://doi.org/10.1214/EJP.v19-3428

[8] BOBKOV, S. G. and LEDOUX, M. (2009). Weighted Poincaré-type inequalities for Cauchy and other convex measures. *Ann. Probab.* **37** 403–427. MR2510011 https://doi.org/10.1214/08-AOP407

[9] BORNN, L., PILLAI, N., SMITH, A. and WOODARD, D. (2014). One pseudo-sample is enough in approximate Bayesian computation MCMC. *Biometrika* **99** 1–10.

[10] CARACCIOLO, S., PELISSETTO, A. and SOKAL, A. D. (1990). Nonlocal Monte Carlo algorithm for self-avoiding walks with fixed endpoints. *J. Stat. Phys.* **60** 1–53. MR1063212 https://doi.org/10.1007/BF01013668

[11] CATTIAUX, P., GOZLAN, N., GUILLIN, A. and ROBERTO, C. (2010). Functional inequalities for heavy tailed distributions and application to isoperimetry. *Electron. J. Probab.* **15** 346–385. MR2609591 https://doi.org/10.1214/EJP.v15-754

[12] DELIGIANNIDIS, G. and LEE, A. (2018). Which ergodic averages have finite asymptotic variance? *Ann. Appl. Probab.* **28** 2309–2334. MR3843830 https://doi.org/10.1214/17-AAP1358

[13] DIACONIS, P. and SALOFF-COSTE, L. (1996). Nash inequalities for finite Markov chains. *J. Theoret. Probab.* **9** 459–510. MR1385408 https://doi.org/10.1007/BF02214660

[14] DONOHUE, M. C., OVERHOLSER, R., XU, R. and VAIDA, F. (2011). Conditional Akaike information under generalized linear and proportional hazards mixed models. *Biometrika* **98** 685–700. MR2836414 https://doi.org/10.1093/biomet/asr023

[15] DOUC, R., MOULINES, E., PRIOURET, P. and SOULIER, P. (2018). *Markov Chains. Springer Series in Operations Research and Financial Engineering*. Springer, Cham. MR3889011 https://doi.org/10.1007/978-3-319-97704-1

[16] DOUCET, A., PITT, M. K., DELIGIANNIDIS, G. and KOHN, R. (2015). Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator. *Biometrika* **102** 295–313. MR3371005 https://doi.org/10.1093/biomet/asu075

[17] FILL, J. A. (1991). Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process. *Ann. Appl. Probab.* **1** 62–87. MR1097464

[18] GÅSEMYR, J. (2006). The spectrum of the independent Metropolis–Hastings algorithm. *J. Theoret. Probab.* **19** 152–165. MR2256484 https://doi.org/10.1007/s10959-006-0009-2

[19] GREENBERG, E. R., BARON, J. A., STUKEL, T. A., STEVENS, M. M., MANDEL, J. S., SPENCER, S. K. et al. and SKIN CANCER PREVENTION STUDY GROUP (1990). A clinical trial of beta carotene to prevent basal-cell and squamous-cell cancers of the skin. *N. Engl. J. Med.* **323** 789–795.

[20] JARNER, S. F. and ROBERTS, G. O. (2002). Polynomial convergence rates of Markov chains. *Ann. Appl. Probab.* **12** 224–247. MR1890063 https://doi.org/10.1214/aoap/1015961162

[21] JARNER, S. F. and ROBERTS, G. O. (2007). Convergence of heavy-tailed Monte Carlo Markov chain algorithms. *Scand. J. Stat.* **34** 781–815. MR2396939 https://doi.org/10.1111/j.1467-9469.2007.00557.x

[22] KHARE, K. and HOBERT, J. P. (2011). A spectral analytic comparison of trace-class data augmentation algorithms and their sandwich variants. *Ann. Statist.* **39** 2585–2606. MR2906879 https://doi.org/10.1214/11-AOS916

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

[24] LEE, A. and ŁATUSZYŃSKI, K. (2014). Variance bounding and geometric ergodicity of Markov chain Monte Carlo kernels for approximate Bayesian computation. *Biometrika* **101** 655–671. MR3254907 https://doi.org/10.1093/biomet/asu027

[25] LEE, A., TIBERI, S. and ZANELLA, G. (2019). Unbiased approximations of products of expectations. *Biometrika* **106** 708–715. MR3992400 https://doi.org/10.1093/biomet/asz008

[26] MEYN, S. P. and TWEEDIE, R. L. (2012). *Markov Chains and Stochastic Stability*. Springer Science & Business Media. MR1287609 https://doi.org/10.1007/978-1-4471-3267-7

[27] PECCOUD, J. and YCART, B. (1995). Markovian modeling of gene-product synthesis. *Theor. Popul. Biol.* **48** 222–234.

[28] ROBERTS, G. O., GELMAN, A. and GILKS, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.* **7** 110–120. MR1428751 https://doi.org/10.1214/aoap/1034625254

[29] RÖCKNER, M. and WANG, F.-Y. (2001). Weak Poincaré inequalities and L^2-convergence rates of Markov semigroups. *J. Funct. Anal.* **185** 564–603. MR1856277 https://doi.org/10.1006/jfan.2001.3776

[30] ROSENTHAL, J. S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **90** 558–566. MR1340509

[31] SHERLOCK, C., THIERY, A. H. and LEE, A. (2017). Pseudo-marginal Metropolis–Hastings sampling using averages of unbiased estimators. *Biometrika* **104** 727–734. MR3694593 https://doi.org/10.1093/biomet/asx031

[32] SHERLOCK, C., THIERY, A. H., ROBERTS, G. O. and ROSENTHAL, J. S. (2015). On the efficiency of pseudo-marginal random walk Metropolis algorithms. *Ann. Statist.* **43** 238–275. MR3285606 https://doi.org/10.1214/14-AOS1278

[33] TIBERI, S., WALSH, M., CAVALLARO, M., HEBENSTREIT, D. and FINKENSTÄDT, B. (2018). Bayesian inference on stochastic gene transcription from flow cytometry data. *Bioinformatics* **34** i647–i655.

[34] TRAN, M. N., KOHN, R., QUIROZ, M. and VILLANI, M. (2016). The block pseudo-marginal sampler. arXiv preprint arXiv:1603.02485.