CONVERGENCE PROPERTIES OF PSEUDO-MARGINAL MARKOV CHAIN MONTE CARLO ALGORITHMS
Author(s): Christophe Andrieu and Matti Vihola
Source: *The Annals of Applied Probability*, Vol. 25, No. 2 (April 2015), pp. 1030-1077
Published by: Institute of Mathematical Statistics
Stable URL: https://www.jstor.org/stable/24519941
Accessed: 18-09-2019 13:29 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve and extend access to *The Annals of Applied Probability*

JSTOR

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

The Annals of Applied Probability
2015, Vol. 25, No. 2, 1030–1077
DOI: 10.1214/14-AAP1022
© Institute of Mathematical Statistics, 2015

CONVERGENCE PROPERTIES OF PSEUDO-MARGINAL MARKOV CHAIN MONTE CARLO ALGORITHMS

BY CHRISTOPHE ANDRIEU[1] AND MATTI VIHOLA[2]

University of Bristol and University of Jyväskylä

We study convergence properties of pseudo-marginal Markov chain Monte Carlo algorithms (Andrieu and Roberts [*Ann. Statist.* **37** (2009) 697–725]). We find that the asymptotic variance of the pseudo-marginal algorithm is always at least as large as that of the marginal algorithm. We show that if the marginal chain admits a (right) spectral gap and the weights (normalised estimates of the target density) are uniformly bounded, then the pseudo-marginal chain has a spectral gap. In many cases, a similar result holds for the absolute spectral gap, which is equivalent to geometric ergodicity. We consider also unbounded weight distributions and recover polynomial convergence rates in more specific cases, when the marginal algorithm is uniformly ergodic or an independent Metropolis–Hastings or a random-walk Metropolis targeting a super-exponential density with regular contours. Our results on geometric and polynomial convergence rates imply central limit theorems. We also prove that under general conditions, the asymptotic variance of the pseudo-marginal algorithm converges to the asymptotic variance of the marginal algorithm if the accuracy of the estimators is increased.

1. Introduction. Assume that one is interested in sampling from a probability distribution π defined on some measurable space $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$. One practical recipe to achieve this in complex scenarios consists of using Markov chain Monte Carlo (MCMC) methods, of which the Metropolis–Hastings update is the main workhorse [15, 24]. We may write the Markov kernel related to a Metropolis–Hastings algorithm in the form

$$P(x, \mathrm{d}y) := \min\{1, r(x, y)\}q(x, \mathrm{d}y) + \delta_x(\mathrm{d}y)\rho(x), \tag{1}$$

where $r(x, y)$ is the Radon–Nikodym derivative as defined in [34]

$$r(x, y) := \frac{\pi(\mathrm{d}y)q(y, \mathrm{d}x)}{\pi(\mathrm{d}x)q(x, \mathrm{d}y)} \quad \text{and} \quad \rho(x) := 1 - \int \min\{1, r(x, y)\}q(x, \mathrm{d}y), \tag{2}$$

Received June 2012; revised March 2014.

[1]Supported in part by an EPSRC advanced research fellowship and a Winton Capital research award.

[2]Supported by the Academy of Finland (project 250575) and by the Finnish Academy of Science and Letters, Vilho, Yrjö and Kalle Väisälä Foundation.

MSC2010 subject classifications. Primary 65C40; secondary 60J05, 65C05.

Key words and phrases. Asymptotic variance, geometric ergodicity, Markov chain Monte Carlo, polynomial ergodicity, pseudo-marginal algorithm.

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

where q is the so-called proposal kernel (or proposal distribution). We follow the terminology of [4] and call this method the *marginal algorithm*.

In some situations, the marginal algorithm cannot be implemented due to the intractability of the distribution π. For example, assuming that π and q have densities (also denoted π and q) with respect to some σ-finite measure, it may be that π cannot be evaluated point-wise, and although $r(x, y)$ may be well defined theoretically, it cannot be evaluated either. However, in some situations unbiased nonnegative estimates $\hat{\pi}(x) = W_x \pi(x)$ may be available; that is, $W_x \sim Q_x(\cdot) \geq 0$ and $\mathbb{E}[W_x] = 1$ for any $x \in \mathsf{X}$ (we will refer to W_x as a "weight" throughout the paper). A naive idea may be to use such estimates in place of the true values in order to compute the acceptance probability. A remarkable property is that such an algorithm is in fact correct [4]. This can be seen by considering the following probability distribution:

$$\tilde{\pi}(\mathrm{d}x, \mathrm{d}w) := \pi(\mathrm{d}x)\pi_x(\mathrm{d}w) \qquad \text{with } \pi_x(\mathrm{d}w) := Q_x(\mathrm{d}w)w \tag{3}$$

on the product space $(\mathsf{X} \times \mathsf{W}, \mathcal{B}(\mathsf{X}) \times \mathcal{B}(\mathsf{W}))$ where W is a Borel subset of $\mathbb{R}_+$ and $\mathcal{B}(\mathsf{W})$ are the Borel sets on W. Here $\pi_x(\mathrm{d}w)$ is a probability measure for each $x \in \mathsf{X}$, and therefore π is a marginal distribution of $\tilde{\pi}$.

It is possible to implement a Metropolis–Hastings algorithm targeting $\tilde{\pi}(\mathrm{d}x, \mathrm{d}w)$ using a proposal kernel $\tilde{q}(x, w; \mathrm{d}y, \mathrm{d}u) := q(x, \mathrm{d}y)Q_y(\mathrm{d}u)$ by defining

$$\begin{aligned} &\tilde{P}(x, w; \mathrm{d}y, \mathrm{d}u) \\ &\quad := \min\left\{1, r(x, y)\frac{u}{w}\right\} q(x, \mathrm{d}y)Q_y(\mathrm{d}u) + \delta_{x,w}(\mathrm{d}y, \mathrm{d}u)\tilde{\rho}(x, w), \end{aligned} \tag{4}$$

where the probability of rejection is given as

$$\tilde{\rho}(x, w) := 1 - \iint \min\left\{1, r(x, y)\frac{u}{w}\right\} q(x, \mathrm{d}y)Q_y(\mathrm{d}u).$$

This is the *pseudo-marginal algorithm* [4], which targets π marginally since it is a marginal distribution of $\tilde{\pi}$, and may be implemented in situations where the marginal algorithm may not. As a particular instance of the Metropolis–Hastings algorithm, the pseudo-marginal algorithm converges to $\tilde{\pi}$ under mild assumptions (e.g., [28]), and although it may be seen as a "noisy" version of the marginal algorithm, it is exact since it allows us to target the distribution of interest π. The aim of this paper is to study some of the theoretical properties of such algorithms in terms of the properties of the weights and those of the marginal algorithm. More precisely we investigate the rate of convergence of the pseudo-marginal algorithm to equilibrium and characterise the approximation of the marginal algorithm by the pseudo-marginal algorithm in terms of the variability of their respective ergodic averages.

The apparently abstract structure of the pseudo-marginal algorithm is in fact shared by several practical algorithms which have recently been proposed in order to sample from intractable distributions. The distribution of w is most often

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

implicit, as we illustrate now with one of the simplest examples. Assume for simplicity that the space X is (a Borel subset of) $\mathbb{R}^d$, and $\mathcal{B}(\mathsf{X})$ consists of the Borel subsets of X and that both π and $q(x, \cdot)$ (for any $x \in \mathsf{X}$) have densities with respect to the Lebesgue measure. Consider a situation where the target density is of the form $\pi(x) = \int \pi(x, z)\,\mathrm{d}z$ where the integral cannot be computed analytically. One can suggest approximating this density with an importance sampling estimate of the integral,

$$(5) \qquad W_x \pi(x) = \hat{\pi}(x) = \frac{1}{N} \sum_{n=1}^{N} \frac{\pi(x, Z_k)}{h_x(Z_k)}, \qquad Z_k \sim h_x(\cdot) \text{ independently,}$$

where h_x is a probability density for each $x \in \mathsf{X}$. Note that it is in fact possible to consider unbiased estimators up to a normalising constant since such a constant cancels in the acceptance ratio of the pseudo-marginal algorithm, and without loss of generality, we will assume this constant to be equal to one throughout. This setting was considered by Beaumont in the seminal paper [9] and various extensions proposed in [4]. There are more involved applications of this idea. In the context of state-space models, it has been shown in [1] that W_x can be obtained with a particle filter—resulting in "particle MCMC" algorithms. In [10] it was shown how exact sampling methods can be used to carry out inference in discretely observed diffusion models for which the transition probability is intractable. See also the discussion [20] on the connection with pseudo-marginal MCMC and approximate Bayesian computation.

We now summarise our main findings, which are of two different types, although some of their underpinnings and consequences are related.

Rates of convergence. In previous work [4] it has been shown that a pseudo-marginal chain is uniformly ergodic whenever the marginal algorithm targeting $\pi(x)$ is uniformly ergodic, and the weights are bounded uniformly in x. It was also shown that geometric ergodicity is not possible as soon as the weights W_x are unbounded on a set of positive π-probability. We extend the analysis of the convergence rates of pseudo-marginal algorithms in several directions.

In Section 3, we show that if the marginal chain admits a nonzero (right) spectral gap, and the weights are bounded uniformly in x, then the pseudo-marginal chain has also a nonzero spectral gap. Our proof relies on an explicit lower bound on the spectral gap (Propositions 8 and 10). Our results imply that geometric ergodicity of a marginal algorithm is inherited by the pseudo-marginal chain as soon as the weights are uniformly bounded, either through a slight modification (Remark 15) or directly in many cases by observing that the pseudo-marginal Markov operator is positive (Proposition 16).

We also restate in a more explicit form a result of Andrieu and Roberts [4] which establishes the necessity of the existence of a function $\bar{w}: \mathsf{X} \to [0, \infty)$ such that $Q_x([0, \bar{w}(x)]) = 1$ for the geometric ergodicity of pseudo-marginal algorithms to

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

hold. Assuming that Q_x has positive mass in any neighbourhood of $\bar{w}(x)$, we show through specific examples that $\sup_{x \in \mathsf{X}} \bar{w}(x) < \infty$ may in some cases be a necessary condition for geometric ergodicity of a pseudo-marginal algorithm to hold (second part of Remark 34) while in other situations the existence of such a uniform upper bound is not a requirement (Remark 26 and the first part of Remark 34). Intuitively, the latter will correspond to situations where the marginal algorithm possesses some robustness properties which allow it to counter, up to a limit, the perturbations brought in by the pseudo-marginal approximation.

In Section 5 we consider the particular case where the pseudo-marginal algorithm is an independent Metropolis–Hastings (IMH) algorithm. The primary interest of this example is pedagogical, since the corresponding pseudo-marginal implementation is also an IMH, which lends itself to a straightforward, yet very instructive, analysis. For example it allows us to establish that the existence of (not necessarily uniformly bounded) moments for the weights leads to polynomial convergence rates, while the existence of exponential moments leads to subexponential rates.

In the light of this pedagogical example, we pursue our analysis by considering more general scenarios where the supports of the weight distributions may be unbounded, that is, such that on some set of positive π-probability $Q_x([0, \bar{w}]) < 1$ for any $\bar{w} < \infty$, implying that the corresponding pseudo-marginal algorithms cannot be geometric.

In Section 6, we only assume that the marginal algorithm is uniformly ergodic (together with a mild additional condition) and that the weight distributions are uniformly integrable. We establish the existence of a Lyapunov function satisfying a sub-geometric drift condition toward a small set (Proposition 30 and Lemma 32). In particular, if the weight distributions possess finite power moments, we establish polynomial ergodicity (Corollary 31).

In Section 7 we consider the popular random-walk Metropolis (RWM). Assuming standard tail conditions on π which ensure the geometric ergodicity of the RWM [16] and the existence of uniformly bounded moments we show that the corresponding pseudo-marginal algorithm is polynomially ergodic (Theorem 38). We extend this result to the situation where moments of the weights are assumed to exist but are not necessarily uniformly bounded in x (i.e., we allow them to grow in the tails of π) in Theorem 45. We note in Remark 34 that one of the intermediate results (Lemma 34) in fact implies the existence of a geometric drift when $Q_x([0, \bar{w}(x)]) = 1$ for some appropriate function $\bar{w}$, possibly divergent in the tails of π, which is a consequence of the fast vanishing assumptions on the tails of π.

Asymptotic variance. It is natural to compare the asymptotic performance of ergodic averages obtained from a marginal algorithm and its pseudo-marginal counterpart. One can in fact ask a more general question of practical relevance. In practice, it is often possible to choose the weight distributions Q_x from a family

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

$\{Q_x^N\}_{N\in\mathbb{N}}$ indexed by an accuracy parameter N, as for example in (5). In such situations $\pi_x^N(\mathrm{d}w) = Q_x^N(\mathrm{d}w)w$ converge weakly to $\delta_1(\mathrm{d}w)$ as $N \to \infty$, and one may wonder if the asymptotic variance of the corresponding ergodic averages converge to that of the marginal algorithm.

In Section 2 we first show that the pseudo-marginal and marginal algorithms are ordered both in terms of the mean acceptance probability (Corollary 4) and the asymptotic variance (Theorem 7). The latter result relies on a generalisation of the argument due to Peskun [29, 34], which may be of independent interest. This supports and generalises the empirical observation on examples that the pseudo-marginal algorithm cannot be more efficient than its marginal version.

When the weights are uniformly bounded in x, we start Section 4 with a simple upper bound on the asymptotic variance of the pseudo-marginal algorithm (Corollary 11) from which it is straightforward to deduce that it converges to that of the marginal when the weight upper bound goes to one. We generalise this result to the situation where the weights are unbounded, but $\pi_x^N(\mathrm{d}w)$ converges weakly to $\delta_1(\mathrm{d}w)$ as $N \to \infty$ (Theorem 21). We also show how the sub-geometric ergodicity results proved earlier are essential to establish the conditions of this theorem in practice (Proposition 25).

We conclude in Section 8 where we briefly discuss additional implications of our results such as the existence of central limit theorems, the possibility to compute quantitative expressions for the asymptotic variance, and the analysis of generalisations of pseudo-marginal algorithms.

2. Ordering of the marginal and pseudo-marginal algorithms.

We first introduce some standard notation related to probability measures and Markov transition probabilities. For Π a Markov kernel and μ a probability measure defined on some measurable space $(\mathsf{E}, \mathcal{B}(\mathsf{E}))$ and f a measurable real-valued function on E, we let for any $x \in \mathsf{E}$, $\Pi^0 f(x) := f(x)$,

$$\mu(f) := \int f(x)\mu(\mathrm{d}x) \quad \text{and} \quad \Pi^n f(x) := \int \Pi(x, \mathrm{d}y)\Pi^{n-1} f(y) \qquad \text{for } n \geq 1.$$

We will also denote the inner product between two real-valued functions f and g on E as $\langle f, g\rangle_\mu := \int f(x)g(x)\mu(\mathrm{d}x)$ and the associated norm $\|f\|_\mu := \langle f, f\rangle_\mu^{1/2}$.

We start by a simple lemma, which plays a key role in the ordering of the marginal and the pseudo-marginal algorithms.

LEMMA 1. *For any $x, y \in \mathsf{X}$, we have*

$$\iint Q_x(\mathrm{d}w)wQ_y(\mathrm{d}u)\min\left\{1, r(x,y)\frac{u}{w}\right\} \leq \min\{1, r(x,y)\}.$$

PROOF. Notice that $t \mapsto \min\{1, t\}$ is a concave function. Therefore, one can apply Jensen's inequality, with the probability measure $Q_x(\mathrm{d}w)wQ_y(\mathrm{d}u)$, to get the desired inequality. $\square$

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

In order to facilitate the comparison of P and $\tilde{P}$ we follow [4] and introduce an auxiliary transition probability $\bar{P}$ which is defined on the same space as the pseudo-marginal kernel $\tilde{P}$ and is reversible with respect to $\tilde{\pi}$,

$$\bar{P}(x, w; \mathrm{d}y, \mathrm{d}u) := q(x, \mathrm{d}y)\pi_y(\mathrm{d}u)\min\{1, r(x, y)\} + \delta_{x,w}(\mathrm{d}y, \mathrm{d}u)\rho(x). \tag{6}$$

Application of Lemma 1 leads to the generic result below, which in turn implies an order between the expected acceptance rates (Corollary 4) and the asymptotic variances (Theorem 7) of the marginal and pseudo-marginal algorithms.

PROPOSITION 2. *Let $g : \mathsf{X}^2 \to [0, \infty)$ be a symmetric measurable function, that is, such that $g(x, y) = g(y, x)$ for all $x, y \in \mathsf{X}$. Define*

$$\Delta_{\bar{P}}(g) := \int \tilde{\pi}(\mathrm{d}x, \mathrm{d}w) \int q(x, \mathrm{d}y)\pi_y(\mathrm{d}u)\min\{1, r(x, y)\}g(x, y),$$

$$\Delta_{\tilde{P}}(g) := \int \tilde{\pi}(\mathrm{d}x, \mathrm{d}w) \int q(x, \mathrm{d}y)Q_y(\mathrm{d}u)\min\left\{1, r(x, y)\frac{u}{w}\right\}g(x, y).$$

Then we have $\Delta_{\bar{P}}(g) \geq \Delta_{\tilde{P}}(g)$ and whenever these quantities are finite,

$$\Delta_{\bar{P}}(g) - \Delta_{\tilde{P}}(g) \leq \int \pi(\mathrm{d}x)Q_x(\mathrm{d}w)|w - 1| \int q(x, \mathrm{d}y)\min\{1, r(x, y)\}g(x, y).$$

PROOF. Denote $a(x, y, u, w) := \min\{1, r(x, y)\} - \min\{1, r(x, y)\frac{u}{w}\}$. Since $\int \pi_y(\mathrm{d}u) = 1 = \int Q_y(\mathrm{d}u)$, we may write for a bounded function g

$$\begin{aligned}\Delta_{\bar{P}}(g) - \Delta_{\tilde{P}}(g) &= \int \pi(\mathrm{d}x)q(x, \mathrm{d}y)g(x, y) \int Q_x(\mathrm{d}w)wQ_y(\mathrm{d}u)a(x, y, u, w)\\ &\geq 0,\end{aligned}$$

where the inequality is a consequence of Lemma 1. The general case follows by a truncation argument.

For the second bound, note that $\min\{1, r(x, y)\frac{u}{w}\} \geq \min\{1, r(x, y)\}\min\{1, \frac{u}{w}\}$ and $2\min\{u, w\} = u + w - |u - w|$, and observe that $\Delta_{\tilde{P}}(g)$ can be lower bounded by

$$\begin{aligned}&\int \pi(\mathrm{d}x)q(x, \mathrm{d}y)Q_x(\mathrm{d}w)Q_y(\mathrm{d}u)\min\{1, r(x, y)\}\min\{u, w\}g(x, y)\\ &\quad = \Delta_{\bar{P}}(g)\\ &\qquad - \frac{1}{2}\int \pi(\mathrm{d}x)q(x, \mathrm{d}y)Q_x(\mathrm{d}w)Q_y(\mathrm{d}u)\min\{1, r(x, y)\}|u - w|g(x, y)\\ &\quad \geq \Delta_{\bar{P}}(g) - \int \pi(\mathrm{d}x)Q_x(\mathrm{d}w)|1 - w| \int q(x, \mathrm{d}y)\min\{1, r(x, y)\}g(x, y),\end{aligned}$$

where the last inequality follows by the bound $|u - w| \leq |1 - u| + |1 - w|$, the symmetry of $g(x, y)$ and because

$$\pi(\mathrm{d}x)q(x, \mathrm{d}y)\min\{1, r(x, y)\} = \pi(\mathrm{d}y)q(y, \mathrm{d}x)\min\{1, r(y, x)\}. \qquad \square$$

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

REMARK 3. The upper bound $|u-w| \le |1-w|+|1-u|$ used in Proposition 2 adds at most a factor of two, because $\int Q_x(\mathrm{d}w)|u-w| \ge |1-w|$.

COROLLARY 4. *Let us denote the expected acceptance rates of the marginal and the pseudo-marginal algorithms as*

$$\alpha_P := \int \pi(\mathrm{d}x) \int q(x, \mathrm{d}y) \min\{1, r(x, y)\},$$

$$\alpha_{\tilde{P}} := \int \tilde{\pi}(\mathrm{d}x, \mathrm{d}w) \int q(x, \mathrm{d}y) Q_y(\mathrm{d}u) \min\left\{1, r(x, y)\frac{u}{w}\right\},$$

respectively. Then we have

$$0 \le \alpha_P - \alpha_{\tilde{P}} \le \int |w-1| \pi(\mathrm{d}x)\big(1-\rho(x)\big) Q_x(\mathrm{d}w) \le \int |w-1| \pi(\mathrm{d}x) Q_x(\mathrm{d}w).$$

PROOF. Observe first that

$$\alpha_{\bar{P}} := \int \tilde{\pi}(\mathrm{d}x, \mathrm{d}w) \int q(x, \mathrm{d}y) Q_y(\mathrm{d}u) \min\{1, r(x, y)\} = \alpha_P.$$

Applying then Proposition 2 with $g \equiv 1$ implies

$$0 \le \alpha_P - \alpha_{\tilde{P}} \le \int |w-1| \pi(\mathrm{d}x)\big(1-\rho(x)\big) Q_x(\mathrm{d}w).$$

The last inequality follows because $\rho(x) \in [0, 1]$ for all $x \in \mathsf{X}$. $\square$

REMARK 5. Corollary 4 implies also the following bounds:

$$\alpha_P - \alpha_{\tilde{P}} \le \begin{cases} \alpha_P \Big(\sup_{x \in \mathsf{X}} \int Q_x(\mathrm{d}w)|1-w| \Big), \\ \alpha_P^{1/p} \Big(\int \pi(\mathrm{d}x) Q_x(\mathrm{d}w) |1-w|^q \Big)^{1/q}, \end{cases}$$

where $p, q > 1$ with $1/p + 1/q = 1$.

We now define the notion of asymptotic variance for scaled ergodic averages of a Markov chain.

DEFINITION 6. Let Π be a reversible Markov kernel with invariant distribution μ defined on some measurable space $(\mathsf{E}, \mathcal{B}(\mathsf{E}))$, and denote by $(X_k)_{k \ge 0}$ the corresponding Markov chain at stationarity, that is, such that $X_0 \sim \mu$. Suppose $f : \mathsf{E} \to \mathbb{R}$ satisfies $\mu(f^2) < \infty$. The *asymptotic variance* of f under Π is defined as

$$\mathrm{var}(f, \Pi) := \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\left(\sum_{k=1}^{n} f(X_k) - \mu(f) \right)^2 \in [0, \infty]. \tag{7}$$

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

Whenever the *integrated autocorrelation time*

$$\tau(f,\Pi) := 1 + 2\sum_{k=1}^{\infty} \frac{\mathbb{E}[f(X_0)f(X_k)] - \pi(f)^2}{\mathrm{var}_\mu(f)}$$

$$\text{with } \mathrm{var}_\mu(f) := \mu\big(f - \mu(f)\big)^2,$$

exists and is finite, then $\mathrm{var}(f,\Pi) = \tau(f,\Pi)\,\mathrm{var}_\mu(f) \in [0,\infty)$.

Lemma 52 in Appendix A shows that the limit in (7) always exists (but may be infinite) and proves the relation between $\tau(f,\Pi)$ and $\mathrm{var}(f,\Pi)$. We now show that a pseudo-marginal algorithm is always dominated by its associated marginal algorithm in terms of asymptotic variance. The result can be regarded as an extension of Peskun's approach [29, 34]. We point out in the proof what makes the result not straightforward.

THEOREM 7. *Assume* $f:\mathsf{X}\to\mathbb{R}$ *satisfies* $\pi(f^2)<\infty$. *Denote* $\mathrm{var}(f,\tilde{P}) = \mathrm{var}(\tilde{f},\tilde{P})$ *where* $\tilde{f}(x,\cdot)\equiv f(x)$.

(i) *Then* $\mathrm{var}(f,\tilde{P}) \ge \mathrm{var}(f,P)$.
(ii) *More specifically,*

$$\mathrm{var}(f,\tilde{P}) \ge \mathrm{var}(f,P) + \liminf_{\lambda\to 1-}\big[\Delta_{\bar{P}}(g_\lambda) - \Delta_{\tilde{P}}(g_\lambda)\big],$$

where $\Delta_{\bar{P}}(g_\lambda)$ *and* $\Delta_{\tilde{P}}(g_\lambda)$ *are defined in Proposition* 2 *and* $g_\lambda(x,y) := [\phi_\lambda(x) - \phi_\lambda(y)]^2$ *with* $\phi_\lambda(x) := \sum_{k=0}^{\infty}\lambda^k[P^k f(x) - \pi(f)]$ *for* $\lambda\in[0,1)$.

PROOF. Our proof is inspired by the proof of Tierney [34], Theorem 4, but we cannot use his argument directly because Proposition 2 does not apply to functions depending also on u and w. Observe first from the definition of $\bar{P}$ that a Markov chain $(\bar{X}_n,\bar{W}_n)_{n\ge 0}$ with the kernel $\bar{P}$ and with $(\bar{X}_0,\bar{W}_0)\sim\tilde{\pi}$ coincides marginally with the marginal chain; that is, $(X_n)_{n\ge 0}$ following P with $X_0\sim\pi$ and $(\bar{X}_n)_{n\ge 0}$ have the same distribution. Therefore, $\mathrm{var}(f,\bar{P}) = \mathrm{var}(f,P)$. We denote

$$\bar{f}(x) := f(x) - \pi(f) \in L_0^2(\mathsf{X},\pi) := \{f:\mathsf{X}\to\mathbb{R} : \pi(f) = 0, \pi(f^2) < \infty\},$$

and with a slight abuse of notation define $\bar{f}(x,w) := \bar{f}(x)$ for all $(x,w)\in\mathsf{X}\times\mathsf{W}$. Notice that $\bar{f}\in L_0^2(\mathsf{X}\times\mathsf{W},\tilde{\pi})$. For $\lambda\in[0,1)$, we define the auxiliary quantities

$$\mathrm{var}_\lambda(\bar{f},H) = \langle\bar{f},(I-\lambda H)^{-1}(I+\lambda H)\bar{f}\rangle_{\tilde{\pi}},$$

for any Markov kernel H reversible with respect to $\tilde{\pi}$, where I stands for the identity operator. We note that from Lemma 51 in Appendix A, the quantity $\mathrm{var}_\lambda(\bar{f},H)$ is well defined and that from Lemma 52, it is sufficient to show that $\mathrm{var}_\lambda(\bar{f},\bar{P}) \le \mathrm{var}_\lambda(\bar{f},\tilde{P})$ holds for all $\lambda\in[0,1)$ in order to establish (i).

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

Using the notation of Lemma 51 with $P_1 = \bar{P}$ and $P_2 = \tilde{P}$, we can write

$$\begin{aligned}
\mathrm{var}_\lambda(\bar{f}, \tilde{P}) - \mathrm{var}_\lambda(\bar{f}, \bar{P}) &= \langle \bar{f}, A_\lambda(1)\bar{f}\rangle_{\tilde{\pi}} - \langle \bar{f}, A_\lambda(0)\bar{f}\rangle_{\tilde{\pi}} \\
&= \int_0^1 \langle \bar{f}, A'_\lambda(\beta)\bar{f}\rangle_{\tilde{\pi}}\, \mathrm{d}\beta \qquad (8) \\
&= \int_0^1 \int_0^\beta \langle \bar{f}, A''_\lambda(\gamma)\bar{f}\rangle_{\tilde{\pi}}\, \mathrm{d}\gamma\, \mathrm{d}\beta + \int_0^1 \langle \bar{f}, A'_\lambda(0)\bar{f}\rangle_{\tilde{\pi}}\, \mathrm{d}\beta.
\end{aligned}$$

Note that if $\tilde{P}$ and $\bar{P}$ would satisfy Peskun's order, then the second line is sufficient to conclude [34]. We show now that both terms on the right-hand side of the last line are nonnegative.

First observe that by Lemma 51,

$$\begin{aligned}
\langle \bar{f}, A'_\lambda(0)\bar{f}\rangle_{\tilde{\pi}} &= 2\lambda\langle \bar{f}, (I - \lambda\bar{P})^{-1}(\tilde{P} - \bar{P})(I - \lambda\bar{P})^{-1}\bar{f}\rangle_{\tilde{\pi}} \\
&= 2\lambda\langle \phi_\lambda, (\tilde{P} - \bar{P})\phi_\lambda\rangle_{\tilde{\pi}},
\end{aligned}$$

due to the reversibility of $\bar{P}$, where $\phi_\lambda := (I - \lambda\bar{P})^{-1}\bar{f} = \sum_{k=0}^\infty \lambda^k \bar{P}^k \bar{f}$ is well defined by Lemma 51. We notice that $\bar{P}^k\bar{f}(x, w) = P^k\bar{f}(x)$ implying $\phi_\lambda(x, w) = \phi_\lambda(x)$, and a straightforward calculation [cf. (9)] shows that

$$\begin{aligned}
&\langle \phi_\lambda, (\tilde{P} - \bar{P})\phi_\lambda\rangle_{\tilde{\pi}} \\
&\quad = \int \tilde{\pi}(\mathrm{d}x, \mathrm{d}w)\phi_\lambda(x)\phi_\lambda(y)\big(\tilde{P}(x, w; \mathrm{d}y, \mathrm{d}u) - \bar{P}(x, w; \mathrm{d}y, \mathrm{d}u)\big) \\
&\quad = \frac{1}{2}\int \big(\phi_\lambda(x) - \phi_\lambda(y)\big)^2 \tilde{\pi}(\mathrm{d}x, \mathrm{d}w)\big(\bar{P}(x, w; \mathrm{d}y, \mathrm{d}u) - \tilde{P}(x, w; \mathrm{d}y, \mathrm{d}u)\big) \\
&\quad = \frac{1}{2}[\Delta_{\bar{P}}(g_\lambda) - \Delta_{\tilde{P}}(g_\lambda)],
\end{aligned}$$

with $g_\lambda(x, y) = (\phi_\lambda(x) - \phi_\lambda(y))^2$, and Proposition 2 yields $\langle \bar{f}, A'_\lambda(0)\bar{f}\rangle_{\tilde{\pi}} \geq 0$. We therefore turn our attention to

$$\begin{aligned}
&\langle \bar{f}, A''_\lambda(\gamma)\bar{f}\rangle_{\tilde{\pi}} \\
&\quad = 4\lambda^2\langle \bar{f}, (I - \lambda H_\gamma)^{-1}(\tilde{P} - \bar{P})(I - \lambda H_\gamma)^{-1}(\tilde{P} - \bar{P})(I - \lambda H_\gamma)^{-1}\bar{f}\rangle_{\tilde{\pi}} \\
&\quad = 4\lambda^2\langle \varphi, (I - \lambda H_\gamma)^{-1}\varphi\rangle_{\tilde{\pi}},
\end{aligned}$$

where $\varphi := (\tilde{P} - \bar{P})(I - \lambda H_\gamma)^{-1}\bar{f}$, by the reversibility of $\bar{P}$ and $\tilde{P}$ and the interpolated kernel $H_\gamma = \bar{P} + \gamma(\tilde{P} - \bar{P})$. It is possible to check that $\varphi \in L_0^2(\mathsf{X} \times \mathsf{W}, \tilde{\pi})$, so we may conclude (i) by applying Lemma 52 implying $\langle \varphi, (I - \lambda H_\gamma)^{-1}\varphi\rangle_{\tilde{\pi}} \geq 0$.

The specific lower bound (ii) follows from (8) because the first term is always nonnegative. $\square$

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

3. Inheritance of the spectral gaps when the weights are uniformly bounded. We consider now an order between the spectral gaps of the pseudo-marginal kernel $\tilde{P}$ and the auxiliary kernel $\bar{P}$ defined in (6). In particular, we find that if w is always bounded from above by $\bar{w} \in [1, \infty)$, that is, $\mathsf{W} = (0, \bar{w}]$, and P has a nonzero (right) spectral gap (i.e., P is variance bounding; see [30], Theorem 14), then $\tilde{P}$ has a nonzero spectral gap as well. We will also examine the asymptotic variance constants using the spectral gap bound, and conclude the section by a discussion on how our results on the spectral gap can imply geometric ergodicity of $\tilde{P}$.

Suppose $f : \mathsf{X} \times \mathsf{W} \to \mathbb{R}$ is integrable with respect to $\tilde{\pi}$. We denote in this section the function centred with respect to w as

$$\bar{f}(x, w) := f(x, w) - f_0(x)$$
$$\text{with } f_0(x) := \pi_x\big(f(x, \cdot)\big) = \int_0^\infty f(x, w)\pi_x(\mathrm{d}w).$$

The Dirichlet form related to a Markov kernel Π with invariant distribution μ and a function g is given as

$$\mathcal{E}_\Pi(g) := \langle g, (I - \Pi)g\rangle_\mu = \frac{1}{2}\int \mu(\mathrm{d}x)\Pi(x, \mathrm{d}y)[g(x) - g(y)]^2, \tag{9}$$

where I is the identity operator. The spectral gap is defined through

$$\mathrm{Gap}(\Pi) := \inf_{g:\mathrm{var}_\mu(g)>0} \frac{\mathcal{E}_\Pi(g)}{\mathrm{var}_\mu(g)} = \inf_{g:\mu(g)=0, \|g\|_\mu=1} \mathcal{E}_\Pi(g), \tag{10}$$

where $\mathrm{var}_\mu(g)$ is given in Definition 6.

PROPOSITION 8. *The spectral gap of $\bar{P}$ defined in* (6) *satisfies*

$$\mathrm{Gap}(P) \wedge \Big(1 - \operatorname*{ess\,sup}_{x\in\mathsf{X}} \rho(x)\Big) \le \mathrm{Gap}(\bar{P}) \le \mathrm{Gap}(P),$$

where the essential supremum is with respect to π.

PROOF. Let $f : \mathsf{X} \times \mathsf{W} \to \mathbb{R}$ with $\tilde{\pi}(f) = 0$ and $\|f\|_{\tilde{\pi}} = 1$, and compute

$$\begin{aligned}
\mathcal{E}_{\bar{P}}(f) - \mathcal{E}_P(f_0) &= \frac{1}{2}\int \pi(\mathrm{d}x)\pi_x(\mathrm{d}w)q(x, \mathrm{d}y)\pi_y(\mathrm{d}u)\min\{1, r(x, y)\} \\
&\qquad \times \big([f(x, w) - f(y, u)]^2 - [f_0(x) - f_0(y)]^2\big) \\
&= \int \pi(\mathrm{d}x)\pi_x(\mathrm{d}w)q(x, \mathrm{d}y)\min\{1, r(x, y)\}[f^2(x, w) - f_0^2(x)] \\
&= \int \pi(\mathrm{d}x)\pi_x(\mathrm{d}w)[f(x, w) - f_0(x)]^2(1 - \rho(x)).
\end{aligned}$$

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

In other words,

$$\mathcal{E}_{\bar{P}}(f) = \mathcal{E}_P(f_0) + \int \pi(\mathrm{d}x)\pi_x(\mathrm{d}w)\big(1-\rho(x)\big)\bar{f}^2(x,w). \tag{11}$$

If $\mathrm{var}_\pi(f_0) > 0$, then we have by (11),

$$\begin{aligned}\mathcal{E}_{\bar{P}}(f) &\geq \mathrm{Gap}(P)\,\mathrm{var}_\pi(f_0) + \int \pi(\mathrm{d}x)\pi_x(\mathrm{d}w)\big(1-\rho(x)\big)\bar{f}^2(x,w)\\ &\geq \mathrm{Gap}(P)\big(1-\tilde{\pi}(\bar{f}^2)\big) + \Big(1 - \operatorname*{ess\,sup}_{x\in\mathsf{X}} \rho(x)\Big)\tilde{\pi}(\bar{f}^2),\end{aligned} \tag{12}$$

where we have used that $1 = \mathrm{var}_{\tilde{\pi}}(f) = \mathrm{var}_\pi(f_0) + \tilde{\pi}(\bar{f}^2)$ by the variance decomposition identity. We notice that (12) holds also when $\mathrm{var}_\pi(f_0) = 0$. We conclude with the bound $\mathcal{E}_{\bar{P}}(f) \geq \mathrm{Gap}(P) \wedge (1 - \operatorname{ess\,sup}_{x\in\mathsf{X}} \rho(x))$ which holds for all $\|f\|_{\tilde{\pi}} = 1$ with $\tilde{\pi}(f) = 0$, implying the first inequality.

For the second inequality, note that if $f(x,w) = f_0(x)$ for all $(x,w) \in \mathsf{X} \times \mathsf{W}$, then $\pi(f_0) = 0$ and $\pi(f_0^2) = 1$. Consequently, $\mathcal{E}_{\bar{P}}(f) = \mathcal{E}_P(f_0)$. Therefore, $\mathrm{Gap}(\bar{P}) \leq \mathrm{Gap}(P)$. □

REMARK 9. In the case where π is not concentrated on points, that is, $\pi(\{x\}) = 0$ for all $x \in \mathsf{X}$, the statement of Proposition 8 simplifies to $\mathrm{Gap}(\bar{P}) = \mathrm{Gap}(P)$, because then $1 - \operatorname{ess\,sup}_{x\in\mathsf{X}} \rho(x) \geq \mathrm{Gap}(P)$ by Lemma 54(ii) in Appendix B.

PROPOSITION 10. *Suppose that there exists a constant $\bar{w} \in [1,\infty)$ such that*

$$Q_x\big([0,\bar{w}]\big) = 1 \qquad \textit{for } \pi\textit{-almost every } x \in \mathsf{X}. \tag{13}$$

Then, the Dirichlet form of the pseudo-marginal algorithm satisfies

$$\mathcal{E}_{\tilde{P}}(f) \geq \bar{w}^{-1}\mathcal{E}_{\bar{P}}(f),$$

for any function with $\tilde{\pi}(f^2) < \infty$, implying $\mathrm{Gap}(\tilde{P}) \geq \bar{w}^{-1}\,\mathrm{Gap}(\bar{P})$.

PROOF. Because $\min\{1,ab\} \geq \min\{1,a\}\min\{1,b\}$ for all $a,b \geq 0$, we have, denoting $\Delta^2 f(x,w;y,u) := [f(x,w) - f(y,u)]^2$

$$\begin{aligned}2\mathcal{E}_{\tilde{P}}(f) &= \int \tilde{\pi}(\mathrm{d}x,\mathrm{d}w)q(x,\mathrm{d}y)Q_y(\mathrm{d}u)\min\Big\{1, r(x,y)\frac{u}{w}\Big\}\Delta^2 f(x,w;y,u)\\ &\geq \int_{u>0} \tilde{\pi}(\mathrm{d}x,\mathrm{d}w)q(x,\mathrm{d}y)\pi_y(\mathrm{d}u)\min\{1,r(x,y)\}\\ &\quad\times \min\Big\{\frac{1}{u},\frac{1}{w}\Big\}\Delta^2 f(x,w;y,u)\\ &\geq 2\bar{w}^{-1}\mathcal{E}_{\bar{P}}(f).\end{aligned}$$

□

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

COROLLARY 11. *Assume* $\mathrm{Gap}(P) > 0$, *and there exists some* $\bar{w} \in [1, \infty)$ *such that* (13) *holds. Let* $g : \mathsf{X} \to \mathbb{R}$ *satisfy* $\pi(g^2) < \infty$. *Then the asymptotic variances* (*Definition* 6) *satisfy*

$$\mathrm{var}(g, P) \le \mathrm{var}(g, \tilde{P}) \le \bar{w}\,\mathrm{var}(g, P) + (\bar{w} - 1)\,\mathrm{var}_\pi(g),$$

where $\mathrm{var}(g, \tilde{P}) := \mathrm{var}(\tilde{g}, \tilde{P})$ *with* $\tilde{g}(x, \cdot) \equiv g(x)$.

PROOF. Proposition 10 implies $\langle f, (I - \tilde{P})f\rangle_{\tilde{\pi}} \ge \langle f, \bar{w}^{-1}(I - \bar{P})f\rangle_{\tilde{\pi}}$ for all functions $\tilde{\pi}(f^2) < \infty$, and Lemma 53 in Appendix B implies

$$\langle \tilde{g}, (I - \tilde{P})^{-1}\tilde{g}\rangle_{\tilde{\pi}} \le \bar{w}\langle \tilde{g}, (I - \bar{P})^{-1}\tilde{g}\rangle_{\tilde{\pi}}.$$

Now note that $\mathrm{var}_{\tilde{\pi}}(\tilde{g}) = \mathrm{var}_\pi(g)$ and $\mathrm{var}(\tilde{g}, \bar{P}) = \mathrm{var}(g, P)$ hold because $\bar{P}$ and P coincide marginally; see the proof of Theorem 7. The above, together with Theorem 7, imply

$$\mathrm{var}_\pi(g) + \mathrm{var}(g, P) \le \mathrm{var}_{\tilde{\pi}}(\tilde{g}) + \mathrm{var}(\tilde{g}, \tilde{P}) \le \bar{w}\big(\mathrm{var}_{\tilde{\pi}}(\tilde{g}) + \mathrm{var}(\tilde{g}, \bar{P})\big),$$

and allows us to conclude. □

REMARK 12. From the proof of Proposition 10, one observes that in fact

$$\mathrm{Gap}(\tilde{P}) \ge \mathrm{Gap}(\check{P}) \ge \bar{w}^{-1}\,\mathrm{Gap}(\bar{P}),$$

where $\check{P}$ is the Markov kernel with the proposal $q(x, \mathrm{d}y)Q_y(\mathrm{d}u)$ and the acceptance probability $\min\{1, r(x, y)\}\min\{1, u/w\}$ reversible with respect to $\tilde{\pi}$. This implies, repeating the arguments in the proof of Corollary 11, that $\mathrm{var}(f, \tilde{P}) \le \mathrm{var}(f, \check{P})$ for all $\tilde{\pi}(f^2) < \infty$.

We also note that in our follow-up work [5], we upper bound the spectral gap of the pseudo-marginal algorithm by that of the marginal, $\mathrm{Gap}(\tilde{P}) \le \mathrm{Gap}(P)$.

Next we show that the boundedness of the support of the weight distributions Q_x for essentially all $x \in \mathsf{X}$ is a necessary condition for the spectral gap of the pseudo-marginal algorithm. The result is similar to Theorem 8 in [4], but its proof is different and the statement more explicit.

PROPOSITION 13. *If the pseudo-marginal kernel* $\tilde{P}$ *has a nonzero spectral gap, then there exists a function* $\bar{w} : \mathsf{X} \to [1, \infty)$ *such that* $Q_x([0, \bar{w}(x)]) = 1$ *for* π-*a.e.* $x \in \mathsf{X}$.

PROOF. We prove the claim by contradiction. Assume that there exists a set $A \in \mathcal{B}(\mathsf{X})$ with $\pi(A) > 0$ such that $Q_x(([0, \tilde{w}]) < 1$ for all $x \in A$ and all $\tilde{w} \in [1, \infty)$. Fix $\varepsilon > 0$ and define a measurable function $\tilde{w}_\varepsilon(x) := \inf\{w \in \mathbb{N} : 1 - \tilde{\rho}(x, w) \le \varepsilon\}$, which is finite everywhere, because the term $\tilde{\rho}(x, w) \to 1$ as $w \to \infty$ (monotonically) for all $x \in \mathsf{X}$. Observe that $\tilde{\pi}(\tilde{A}_\varepsilon) > 0$ where $\tilde{A}_\varepsilon := \{(x, w) \in$

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

$A \times \mathsf{W} : w \geq \tilde{w}_\varepsilon(x)\}$. Because $\tilde{w}_\varepsilon$ increases to infinity as $\varepsilon \to 0$, we have $\tilde{\pi}(\tilde{A}_\varepsilon) \in (0, 1/2)$ for small enough $\varepsilon > 0$. For such $\varepsilon > 0$, we may apply Lemma 54(i) in Appendix B with the set $\tilde{A}_\varepsilon$, to conclude that $\mathrm{Gap}(\tilde{P}) \leq (1 - \tilde{\pi}(\tilde{A}_\varepsilon))^{-1}\varepsilon \leq 2\varepsilon$. □

REMARK 14. Proposition 13 implies the necessity of the existence of $\bar{w} : \mathsf{X} \to [1, \infty)$ for spectral gap and consequently geometric ergodicity to hold, but does not require the existence of a uniform upper bound $\bar{w}$ as in Proposition 10. Uniformity is indeed not necessary as illustrated in Remarks 26 and 34 with the independent MH and random walk MH algorithms, respectively; see also [21], Remark 1. However, the second part of Remark 34 implies that in some cases the existence of a uniform upper bound $\bar{w}$ is indeed necessary.

The above results are statements on the (right) spectral gap of $\tilde{P}$ only, which is equivalent to variance bounding property of $\tilde{P}$ [31]. In some applications, geometric ergodicity may be more desirable than variance boundedness. We first note that in general, geometricity can be enforced by a slight algorithmic modification.

REMARK 15. Suppose that $\tilde{P}$ is variance bounding. Then, for any $\varepsilon \in (0, 1)$, the lazy version of the pseudo-marginal algorithm $\tilde{P}_\varepsilon := \varepsilon I + (1 - \varepsilon)\tilde{P}$ is geometrically ergodic [31], Theorem 2.

In many cases, however, such a modification is unnecessary, because the pseudo-marginal algorithm can be shown to exhibit also a nonzero left spectral gap, defined using the notation in (10)

$$\operatorname*{Gap}_L(\Pi) := \inf_{g:\mu(g)=0, \|g\|_\mu=1} \big(2 + \mathcal{E}_\Pi(g)\big) = 1 + \inf_{g:\mu(g)=0, \|g\|_\mu=1} \langle g, \Pi g\rangle_\mu.$$

Nonzero left and right spectral gaps, or in other words the existence of an absolute spectral gap, is equivalent to geometric ergodicity of a reversible chain (e.g., [30], Theorem 2.1).

Of particular interest are positive Markov operators Π which satisfy $\langle g, \Pi g\rangle_\mu \geq 0$ for all functions g with $\|g\|_\mu < \infty$. For positive Π, clearly $\mathrm{Gap}_L(\Pi) \geq 1$ and establishing geometric ergodicity only requires focusing on the right spectral gap. We record the following easy proposition summarising two situations where the pseudo-marginal algorithm inherits the positivity of the marginal algorithm.

PROPOSITION 16. *The pseudo-marginal Markov operator is positive and therefore admits a left spectral gap in the following cases*:

(a) *if the marginal algorithm is an independent Metropolis–Hastings* (*IMH*);

(b) *if the marginal algorithm is a random-walk Metropolis* (*RWM*) *with a proposal distribution, which can be written in the form*

$$q(x, y) = \int \eta(z, x)\eta(z, y)\,\mathrm{d}z. \tag{14}$$

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

PROOF. Case (a) holds because the pseudo-marginal version of an IMH is also an IMH (see also Section 5), which is positive (e.g., [14]). Case (b) follows by using an argument of Baxendale [8], Lemma 3.1, by writing for $f : \mathsf{X} \times \mathsf{W} \to \mathbb{R}$ with $\|f\|_{\tilde{\pi}} < \infty$,

$$\langle f, \tilde{P} f\rangle_{\tilde{\pi}} \geq \int \tilde{\pi}(\mathrm{d}x, \mathrm{d}w) q(x, y) Q_y(\mathrm{d}u) \min\left\{1, \frac{\pi(y)}{\pi(x)} \frac{u}{w}\right\} f(x, w) f(y, u)$$

$$= \int \phi^2(t, z)\, \mathrm{d}t\, \mathrm{d}z \geq 0,$$

where $\phi(t, z) := \int f(x, w) \mathbb{I}\{t \leq \pi(x) w\} \eta(z, x) Q_x(\mathrm{d}w)\, \mathrm{d}x$. □

REMARK 17. Condition (14) holds, in particular, with $q(x, y) = \tilde{q}(y - x)$ where $\tilde{q}$ is "divisible;" that is, it is the density of the sum of two independent random variables sharing the same symmetric density q_0. Indeed, in such a scenario $\tilde{q}(y - x) = \int q_0(u) q_0(y - x - u)\, \mathrm{d}u = \int q_0(z - x) q_0(y - z)\, \mathrm{d}z$, and we may take $\eta(z, x) = q_0(z - x) = q_0(x - z)$. This covers the case where $\tilde{q}$ is a (possibly multivariate) Gaussian or Student.

We conjecture that geometric ergodicity is inherited in general as soon as the weights are uniformly bounded. More precisely, we believe that if the marginal algorithm is geometrically ergodic (admits a nonzero absolute spectral gap) and the weights are uniformly bounded, then the pseudo-marginal algorithm is also geometrically ergodic. We have not been able to prove this in general, but we have not found counter-examples either.

For completeness, we, however, provide the following counter-example which shows that the left spectral gap of the marginal algorithm may not be inherited by the pseudo-marginal algorithm without the uniform upper bound assumption on the weights.

EXAMPLE 18. Let $\mathsf{X} = \mathbb{N}$, $\pi(x) = 2^{-x-1}$ and $q(x, x+1) = q(x, x-1) = 1/2$ for all $x \in \mathsf{X}$. Direct calculation yields a geometric drift with function $V(x) = (3/2)^x$ toward an atom $\{0\}$, which shows that P is geometrically ergodic.

Let us then consider $\tilde{P}$ with the weight distributions $\{Q_x\}_{x \in \mathsf{X}}$ defined for $x = 10^k + n$ with $k \geq 1$ and $n \in [1, 10^k]$ by

$$Q_x(w) := (1 - \varepsilon_k) \delta_{a(k,n)}(w) + \varepsilon_k \delta_{b(k,n)}(w),$$

and $Q_x(w) := \delta_1(w)$ otherwise, where $\varepsilon_k := 10^{-k}$ and $a(k, n) := 2^{-10^k + n}$, and the constants $b(k, n) \in (1, \infty)$ are chosen so that $Q_x(w)$ have expectation one. Define the functions

$$f_k(x, w) := \begin{cases} +1, & \text{if } x = 10^k + n \text{ with } n \in [1, 10^k] \text{ odd and } w = a(k, n), \\ -1, & \text{if } x = 10^k + n \text{ with } n \in [1, 10^k] \text{ even and } w = a(k, n), \\ 0, & \text{otherwise.} \end{cases}$$

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

A straightforward calculation shows that $\lim_{k\to\infty}\langle f_k, \tilde{P} f_k\rangle_{\tilde{\pi}}/\|f_k\|_{\tilde{\pi}}^2 = -1$, which shows that there is no left spectral gap. See [6], Appendix E, for details.

4. Convergence of the asymptotic variance. In standard applications of the pseudo-marginal algorithm, one typically selects Q_x from a family of possible proposal distributions Q_x^N indexed by some precision parameter N which reflects the concentration of W on 1. In most relevant scenarios we are aware of, $N \in \mathbb{N}$ corresponds to the number of samples, particles or iterates of an algorithm used to compute an unbiased estimator of the density value, as exemplified in (5). It should be clear that this is not a restriction. Hereafter, we denote the pseudo-marginal kernels and the invariant measures associated with Q_x^N as $\tilde{P}_N$ and $\tilde{\pi}_N$, respectively.

It is easy to see that if for all $x \in \mathsf{X}$, $Q_x^N(\mathrm{d}w)w \to \delta_1(\mathrm{d}w)$ as $N \to \infty$ weakly, then $\tilde{\pi}_N(\mathrm{d}x, \mathrm{d}w) \to \pi(\mathrm{d}x)\delta_1(\mathrm{d}w)$ weakly, suggesting that a pseudo-marginal algorithm with invariant distribution $\tilde{\pi}_N$ may become similar to the marginal algorithm with invariant distribution π as $N \to \infty$. As pointed out earlier, whenever W_x is not bounded uniformly, a pseudo-marginal algorithm cannot be geometric, although its marginal algorithm may be. In fact it was shown in [4], Remark 1, that even in situations where the weights are uniformly bounded and the pseudo-marginal algorithm is uniformly ergodic, increasing N may not improve the rate of convergence of the algorithm, that is, there is not convergence in terms of rate of convergence.

In this section we, however, show that in many situations such a convergence takes place in terms of the asymptotic variance, or equivalently, the integrated autocorrelation time; see Definition 6. More precisely, we show here that under simple conditions $\mathrm{var}(g, \tilde{P}_N) \to \mathrm{var}(g, P)$ as $N \to \infty$. We start with a very simple result, which is a direct consequence of Corollary 11.

PROPOSITION 19. *Suppose that the marginal kernel P has a nonzero spectral gap and the weight distributions are bounded uniformly in $x \in \mathsf{X}$ by $\bar{w}^N \in (1,\infty)$, that is, $Q_x^N([0, \bar{w}^N]) = 1$ for all $x \in \mathsf{X}$ and $N \geq N_0$ for some $N_0 \in \mathbb{N}$, and $\lim_{N\to\infty}\bar{w}^N = 1$. Then, $\lim_{N\to\infty}\mathrm{var}(g, \tilde{P}_N) = \mathrm{var}(g, P)$ for any $g : \mathsf{X} \to \mathbb{R}$ with $\pi(g^2) < \infty$.*

PROOF. The result is direct consequence of Corollary 11. □

We now extend this result to situations where the distributions $\{Q_x^N\}_{N\in\mathbb{N}}$ may have an unbounded support, and therefore $\{\tilde{P}_N\}_{N\in\mathbb{N}}$ may not be geometrically ergodic. We formulate our result in terms of the following technical condition assuming uniform convergence of the integrated autocorrelation series. We will return to this assumption toward the end of this section and show that it can be checked in practice with for example Lyapunov type drift conditions; see Proposition 25.

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

CONDITION 20. For $g:\mathsf{X} \to \mathbb{R}$, suppose that the integrated autocorrelation time $\tau(g, P)$ (Definition 6) is well defined and finite. Denote by $(\tilde{X}_k^N)_{k\geq 0}$ the Markov chain with initial distribution $\tilde{\pi}_N$ and kernel $\tilde{P}_N$. Assume that there exists a constant $N_0 < \infty$ such that

$$\lim_{n\to\infty} \sup_{N\geq N_0} \left| \sum_{k=n}^{\infty} \mathbb{E}[\bar{g}(\tilde{X}_0^N)\bar{g}(\tilde{X}_k^N)] \right| = 0 \qquad \text{where } \bar{g} = g - \pi(g).$$

The main result of this section is the following:

THEOREM 21. *Assume that* $g:\mathsf{X} \to \mathbb{R}$ *satisfies* $\pi(|g|^{2+\delta}) < \infty$, *and Condition* 20 *holds for* g. *Suppose also that*

$$\lim_{N\to\infty} \int Q_x^N(\mathrm{d}w)|1-w| = 0 \qquad \textit{for all } x \in \mathsf{X}. \tag{15}$$

Then, $\lim_{N\to\infty} \mathrm{var}(g, \tilde{P}_N) = \mathrm{var}(g, P)$.

PROOF. If $\mathrm{var}_\pi(g) = 0$, the claim is trivial. If $\mathrm{var}_\pi(g) > 0$, our conditions imply that the autocorrelation times exist and are finite for both the marginal kernel P and the pseudo-marginal kernels $\tilde{P}_N$ for $N \geq N_0$; this follows from the finiteness of the terms in the autocorrelation series ensured by the Cauchy–Schwarz inequality and Condition 20. Therefore, without loss of generality, we prove the claim for autocorrelation times $\tau(g, \tilde{P}_N) \to \tau(g, P)$ for a function g with $\tilde{\pi}_N(g) = \pi(g) = 0$ and $\tilde{\pi}_N(g^2) = \pi(g^2) = 1$.

Consider the Markov kernels $\bar{P}_N$ defined as in (6) with Q_x^N and $\pi_x^N(\mathrm{d}w) := Q_x^N(\mathrm{d}w)w$. Denote by $(\bar{X}_k^N, \bar{W}_k^N)_{k\geq 0}$ the corresponding stationary Markov chain with $(\bar{X}_0^N, \bar{W}_0^N) \sim \bar{\pi}_N$. Denote similarly $(\tilde{X}_k^N, \tilde{W}_k^N)_{k\geq 0}$ the stationary Markov chain corresponding to the kernel $\tilde{P}_N$ with $(\tilde{X}_0^N, \tilde{W}_0^N) \sim \tilde{\pi}_N$. Notice that $\bar{P}_N$ and $\bar{\pi}_N$ coincide marginally with P and π, respectively; that is, $(\bar{X}_k^N)_{k\geq 0}$ has the same distribution as that of the stationary marginal chain $(X_k)_{k\geq 0}$ with kernel P and such that $X_0 \sim \pi$.

Choose $\varepsilon \in (0,1)$ and let $n_0 = n_0(\varepsilon) < \infty$ be such that for all $N \geq N_0$,

$$\left| \sum_{k=n_0}^{\infty} \mathbb{E}[g(\tilde{X}_0^N)g(\tilde{X}_k^N)] \right| \leq \varepsilon \quad \text{and} \quad \left| \sum_{k=n_0}^{\infty} \mathbb{E}[g(X_0)g(X_k)] \right| \leq \varepsilon, \tag{16}$$

where the existence of n_0 follows from Condition 20. We have for $N \geq N_0$,

$$\left|\tau(g, P) - \tau(g, \tilde{P}_N)\right| \leq 4\varepsilon + 2\left| \sum_{k=1}^{n_0 - 1} \mathbb{E}[g(\tilde{X}_0^N)g(\tilde{X}_k^N)] - \mathbb{E}[g(\bar{X}_0)g(\bar{X}_k)] \right|.$$

In order to control the last term, we consider a coupling argument. Denote $q := (2+\delta)/\delta \in (1, \infty)$. Lemma 22 applied with $\check{\varepsilon} = \varepsilon n_0^{-q-1}/2$ implies the existence

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

of $N_1 < \infty$ and a set $\bar{C} \in \mathcal{B}(\mathsf{X}) \times \mathcal{B}(\mathsf{W})$ such that for all $N \ge N_1$,

$$\tilde{\pi}_N(\bar{C}^{\complement}) \le \varepsilon n_0^{-q-1}/2,$$

$$\|\tilde{P}_N(x, w; \cdot) - \bar{P}_N(x, w; \cdot)\| \le \varepsilon n_0^{-q-1}/2 \qquad \text{for all } (x, w) \in \bar{C}.$$

Lemma 55 in Appendix C applied to $(\tilde{X}_k^N, \tilde{W}_k^N)_{0\le k\le n_0-1}$ and $(\bar{X}_k^N, \bar{W}_k^N)_{0\le k\le n_0-1}$ with the set $\bar{C}$ shows that the laws of these processes, $\tilde{\mu}$ and $\bar{\mu}$, respectively, satisfy the following total variation inequality for all $N \ge N_1$,

$$\|\tilde{\mu} - \bar{\mu}\| \le 2n_0\tilde{\pi}_N(\bar{C}^{\complement}) + n_0 \sup_{(x,w)\in\bar{C}} \|\tilde{P}^N(x, w; \cdot) - \bar{P}^N(x, w; \cdot)\| \le 2\varepsilon n_0^{-q}.$$

Therefore, for all $N \ge N_1$, there exists a probability space $(\bar{\Omega}_N, \bar{\mathbb{P}}_N, \bar{\mathcal{F}}_N)$ where both $(\tilde{X}_k^N, \tilde{W}_k^N)_{0\le k\le n_0-1}$ and $(\bar{X}_k^N, \bar{W}_k^N)_{0\le k\le n_0-1}$ are defined, and the set

$$\bar{A}_N := \{(\tilde{X}_k^N, \tilde{W}_k^N) \equiv (\bar{X}_k^N, \bar{W}_k^N), 0 \le k \le n_0 - 1\}$$

satisfies $\bar{\mathbb{P}}_N(\bar{A}_N^{\complement}) = \frac{1}{2}\|\tilde{\mu} - \bar{\mu}\| \le \varepsilon n_0^{-q}$ (e.g., [22], Theorem 5.2). Denote $p = 1 + \delta/2$, and note that $p^{-1} + q^{-1} = 1$. Now for $N \ge N_1$,

$$\begin{aligned}
&\left|\sum_{k=1}^{n_0-1} \mathbb{E}[g(\tilde{X}_0^N)g(\tilde{X}_k^N)] - \mathbb{E}[g(\bar{X}_0^N)g(\bar{X}_k^N)]\right| \\
&\quad = \left|\bar{\mathbb{E}}_N\left[\sum_{k=1}^{n_0-1} g(\tilde{X}_0^N)g(\tilde{X}_k^N) - g(\bar{X}_0^N)g(\bar{X}_k^N)\right]\right| \\
&\quad \le (\bar{\mathbb{P}}_N(\bar{A}_N^{\complement}))^{1/q}\left\{\left(\bar{\mathbb{E}}_N\left|\sum_{k=1}^{n_0-1} g(\tilde{X}_0^N)g(\tilde{X}_k^N) - g(\bar{X}_0^N)g(\bar{X}_k^N)\right|^p\right)^{1/p}\right\} \\
&\quad \le (\bar{\mathbb{P}}_N(\bar{A}_N^{\complement}))^{1/q}(n_0 - 1) \\
&\qquad \times \max_{1\le k\le n_0-1}[(\mathbb{E}|g(\tilde{X}_0^N)g(\tilde{X}_k^N)|^p)^{1/p} + (\mathbb{E}|g(X_0)g(X_k)|^p)^{1/p}] \\
&\quad \le 2\varepsilon^{1/q}(\pi(|g|^{2+\delta}))^{1/(2p)},
\end{aligned}$$

by the Hölder, Minkowski and Cauchy–Schwarz inequalities. $\square$

Let μ_1 and μ_2 be two probability distributions on the space $(\mathsf{E}, \mathcal{B}(\mathsf{E}))$. We define the total variation distance $\|\mu_1 - \mu_2\| := \sup_{|f|\le 1} |\mu_1(f) - \mu_2(f)| = 2\sup_{0\le f\le 1} |\mu_1(f) - \mu_2(f)| = 2\sup_{A\in\mathcal{B}(\mathsf{E})} |\mu_1(A) - \mu_2(A)|$.

LEMMA 22. *Assume that* (15) *is satisfied. Then, for any* $\check{\varepsilon} > 0$ *there exists a* $N_1 < \infty$ *and a set* $\check{C} \in \mathcal{B}(\mathsf{X}) \times \mathcal{B}(\mathsf{W})$ *such that for all* $N \ge N_1$,

$$\tilde{\pi}_N(\check{C}^{\complement}) \le \check{\varepsilon},$$

$$\|\tilde{P}_N(x, w; \cdot) - \bar{P}_N(x, w; \cdot)\| \le \check{\varepsilon} \qquad \textit{for all } (x, w) \in \check{C}.$$

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

PROOF. Choose $\check{\varepsilon} > 0$, and let $\bar{w} := 1 + \check{\varepsilon}/8$. It is not difficult to see that assumption (15) implies for all $x \in \mathsf{X}$,

$$\lim_{N\to\infty} \pi_x^N([\bar{w}^{-1}, \bar{w}]) = 1. \tag{17}$$

Because $\int Q_y^N(\mathrm{d}u)|1-u| \le 2$, the dominated convergence theorem together with (15) implies for all $x \in \mathsf{X}$,

$$\lim_{N\to\infty} \int q(x,\mathrm{d}y) Q_y^N(\mathrm{d}u)|1-u| = 0. \tag{18}$$

By Egorov's theorem, there exists a set $C \in \mathcal{B}(\mathsf{X})$ such that $\pi(C^{\complement}) \le \check{\varepsilon}/2$ and the convergence in both (17) and (18) is uniform in x.

For any $x \in \mathsf{X}$, any $w > 0$ and any set $A \in \mathcal{B}(\mathsf{X}) \times \mathcal{B}(\mathsf{W})$,

$$\begin{aligned}
&|\tilde{P}_N(x,w;A) - \bar{P}_N(x,w;A)| \\
&\quad \le 2\int q(x,\mathrm{d}y) Q_y^N(\mathrm{d}u)\left|\min\{1, r(x,y)\}u - \min\left\{1, r(x,y)\frac{u}{w}\right\}\right| \\
&\quad \le 2\int q(x,\mathrm{d}y) Q_y^N(\mathrm{d}u)\left[|1-u| + \left|\min\{1, r(x,y)\} - \min\left\{1, r(x,y)\frac{u}{w}\right\}\right|\right] \\
&\quad \le 2\int q(x,\mathrm{d}y) Q_y^N(\mathrm{d}u)\left[|1-u| + \left|1 - \frac{u}{w}\right|\right] \\
&\quad \le 2\left|1 - \frac{1}{w}\right| + 4\int q(x,\mathrm{d}y) Q_y^N(\mathrm{d}u)|1-u|,
\end{aligned}$$

where the third inequality follows because

$$\big|\min\{1, ab\} - \min\{1, a\}\big| \le \min\{1, a\}|1-b| \qquad \text{for any } a, b \ge 0.$$

Therefore, letting $\check{C} := C \times [\bar{w}^{-1}, \bar{w}]$, we can bound the total variation by

$$\sup_{(x,w)\in\check{C}} \|\tilde{P}_N(x,w;\cdot) - \bar{P}_N(x,w;\cdot)\| \le \frac{\check{\varepsilon}}{2} + 8\sup_{x\in C}\int q(x,\mathrm{d}y) Q_y^N(\mathrm{d}u)|1-u|.$$

Because $\lim_{N\to\infty} \tilde{\pi}_N(\check{C}^{\complement}) = \pi(C^{\complement})$, we may conclude by choosing $N_1 < \infty$ such that $\sup_{x\in C} \int q(x,\mathrm{d}y) Q_y^N(\mathrm{d}u)|1-u| \le \check{\varepsilon}/16$ and $\tilde{\pi}_N(\check{C}^{\complement}) \le \check{\varepsilon}$ for all $N \ge N_1$. $\square$

REMARK 23. With additional assumptions in Condition 20 and (15) on the rates of convergence, one could obtain a rate of convergence in Theorem 21, that is find $\{r(n)\}_{n\in\mathbb{N}}$ such that

$$\big|\mathrm{var}(g, \tilde{P}_N) - \mathrm{var}(g, P)\big| \le r(N) \to 0 \qquad \text{as } N \to \infty,$$

by going through the proofs of Theorem 21 and Lemma 22.

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

We now provide sufficient conditions implying the conditions of Theorem 21. Condition 20 which essentially require quantitative bounds on the ergodic behaviour of the pseudo-marginal Markov chains. Our results rely on polynomial drift conditions which we establish for some standard algorithms in Sections 5 and 7. Weaker drift conditions can be shown to imply Condition 20 (e.g., [2, 3]), but we do not detail this here in order to keep presentation simple.

CONDITION 24. There exists a function $V : \mathsf{X} \times \mathsf{W} \to [1, \infty)$, a set $C \in \mathcal{B}(\mathsf{X}) \times \mathcal{B}(\mathsf{W})$ with $\sup_{(x,w)\in C} V(x, w) < \infty$, constants $\alpha \in (0, 1]$, $b_V \in [0, \infty)$, $\varepsilon_V \in (0, \infty)$ and $N_0 < \infty$, such that for all $N \geq N_0$

$$\tilde{P}_N V(x, w) \leq V(x, w) - \varepsilon_V V^{\alpha}(x, w) + b_V \mathbb{I}\{(x, w) \in C\} \qquad \forall x \in \mathsf{X}, w \in \mathsf{W},$$

and for any $v \in [1, \infty)$, there exists probability measures $\{\nu^N\}_{N \geq N_0}$ and a constant $\varepsilon_v \in (0, 1]$, such that for all $N \geq N_0$,

$$\tilde{P}_N(x, w; \cdot) \geq \varepsilon_v \nu^N(\cdot) \qquad \text{for all } (x, w) \in \mathsf{X} \times \mathsf{W} \text{ such that } V(x, w) \leq v.$$

PROPOSITION 25. *Assume Condition 24 holds for the pseudo-marginal kernels $\tilde{P}_N$, and that for some $\lambda \in [0, 1)$ and $\kappa \in [0, 1)$,*

$$\|g\|_{V^{\alpha_{\kappa,\lambda}}} = \sup_{(x,w)\in \mathsf{X}\times\mathsf{W}} \frac{|g(x)|}{V^{\alpha_{\kappa,\lambda}}(x, w)} < \infty,$$

$$\sup_{N \geq N_0} \tilde{\pi}_N\big((|g| + 1)V^{1-\lambda\alpha}\big) < \infty,$$

where $\alpha_{\kappa,\lambda} := \kappa\alpha(1 - \lambda)$. Then Condition 20 holds.

PROOF. From the assumptions, there exists a finite constant R such that for all $N \geq N_0$ and any $(x, w), (x', w') \in \mathsf{X} \times \mathsf{W}$,

$$\sum_{k\geq 0} r(k)\big|\tilde{P}_N^k g(x, w) - \tilde{P}_N^k g(x', w')\big|$$
$$\leq R\|g\|_{V^{\alpha_{\kappa,\lambda}}}\big(V^{1-\lambda\alpha}(x, w) + V^{1-\lambda\alpha}(x', w') - 1\big),$$

where $r(k) := (k + 1)^{\alpha(1-\lambda)(1-\kappa)/(1-\alpha)} \to \infty$ as $k \to \infty$ [3], Corollary 8; see also [2], Proposition 3.4. Note that we may write

$$\big|\mathbb{E}_{(x,w)}[\bar{g}(\tilde{X}_k^N)]\big| = \left|\tilde{P}_N^k g(x, w) - \int \tilde{\pi}_N(\mathrm{d}y, \mathrm{d}u)\tilde{P}_N^k g(y, u)\right|$$
$$\leq \int \tilde{\pi}_N(\mathrm{d}y, \mathrm{d}u)\big|\tilde{P}_N^k g(x, w) - \tilde{P}_N^k g(y, u)\big|.$$

Therefore, we have for $n \geq 0$,

$$\left|\sum_{k=n}^{\infty} \mathbb{E}[\bar{g}(\tilde{X}_0^N)\bar{g}(\tilde{X}_k^N)]\right| \leq \mathbb{E}\left[|\bar{g}(\tilde{X}_0^N)| \sum_{k=n}^{\infty} \big|\mathbb{E}_{(\tilde{X}_0^N, \tilde{W}_0^N)}[\bar{g}(\tilde{X}_k^N)]\big|\right]$$
$$\leq \frac{\|g\|_{V^{\alpha_{\kappa,\lambda}}}}{r(n)}\big[\tilde{\pi}_N(|g|V^{1-\lambda\alpha}) + \pi(|g|)\tilde{\pi}_N(V^{1-\lambda\alpha})\big]. \qquad \square$$

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

5. Sub-geometric ergodicity with an IMH as marginal algorithm. The independent Metropolis–Hastings (IMH) algorithm is a specific case of the Metropolis–Hastings in (1) corresponding to a proposal $q(x, \mathrm{d}y) = q(\mathrm{d}y)$ for all $x \in \mathsf{X}$, such that $\pi \ll q$. It is straightforward to check that a pseudo-marginal implementation of this algorithm is also an IMH. This fact allows for the easy exploration of conditions which ensure uniform and sub-geometric ergodicity of the pseudo-marginal IMH, and are illustrative of the general ideas we develop later in the paper. We note that these results may be relevant, for example, to the analysis of the Particle IMH-EM algorithm presented in [7].

REMARK 26. It is now well known that the IMH is uniformly (and geometrically) ergodic if and only if $\pi(\mathrm{d}x)/q(\mathrm{d}x)$ is bounded [23]. In the case of the pseudo-marginal IMH, this is equivalent to assuming that the ratio $\tilde{\pi}(\mathrm{d}x, \mathrm{d}w)/\tilde{q}(\mathrm{d}x, \mathrm{d}w) = w\pi(\mathrm{d}x)/q(\mathrm{d}x)$ is bounded; in other words, assuming that there exists a constant $c \in (0, \infty)$ such that $Q_x([0, \bar{w}(x)]) = 1$ for π-almost every $x \in \mathsf{X}$, where $\bar{w}(x) := cq(\mathrm{d}x)/\pi(\mathrm{d}x)$.

We then give two conditions which ensure sub-geometric ergodicity of the pseudo-marginal IMH. Our results rely on Lemma 56 in Appendix C, which is inspired by [17], which established polynomial ergodicity and [13], which explored more general sub-geometric rates for the IMH.

COROLLARY 27. *Suppose either of the following holds*:

(a) *for some* $\gamma > 0$, $\int \tilde{\pi}(\mathrm{d}x, \mathrm{d}w) \exp[(w\pi(\mathrm{d}x)/q(\mathrm{d}x))^{\gamma}] < \infty$,
(b) *for some* $\beta \geq 1$, $\int \tilde{\pi}(\mathrm{d}x, \mathrm{d}w)(w\pi(\mathrm{d}x)/q(\mathrm{d}x))^{\beta} < \infty$.

Then, there exist constants $M, c, c_V \in (0, \infty)$ *such that for* $w\pi(\mathrm{d}x)/q(\mathrm{d}x) \geq M$, *the following drift inequalities hold*:

$$\tilde{P} V_{(a)}(x, w) \leq V_{(a)}(x, w) - c\kappa\big(V_{(a)}(x, w)\big),$$
$$\tilde{P} V_{(b)}(x, w) \leq V_{(b)}(x, w) - c V_{(b)}^{1-1/\beta}(x, w),$$

respectively, where $V_{(a)}(x, w) = \exp((w\pi(\mathrm{d}x)/q(\mathrm{d}x))^{\gamma})$, $\kappa(t) = t(\log t)^{-1/\gamma}$ *and* $V_{(b)}(x, w) = (w\pi(\mathrm{d}x)/q(\mathrm{d}x))^{\beta} + 1$.

PROOF. Lemma 56 applied with (a) $\phi(t) = \exp(t^{\gamma})$ and (b) $\phi(t) = t^{\beta} + 1$. □

The type of drift in Corollary 27(a) implies faster than polynomial sub-geometric rates of convergence (cf. [12]), whereas Corollary 27(b) implies polynomial rates of convergence (cf. [17]). We notice that the result suggests that the pseudo-marginal algorithm may have a similar rate of convergence as that of the marginal algorithm.

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

6. Sub-geometric ergodicity with uniformly ergodic marginal algorithm. We consider next the situation where the marginal algorithm is uniformly ergodic. This often corresponds to scenarios where the state space $\mathsf{X} \subset \mathbb{R}^d$ is compact. It turns out that when the weight distributions $\{Q_x\}_{x\in\mathsf{X}}$ do not have bounded supports but are uniformly integrable, then the corresponding pseudo-marginal algorithm satisfies a sub-geometric drift condition toward a set $\mathsf{C} := \mathsf{X} \times (0, \bar{w}]$ for some $\bar{w} \in (1, \infty)$. Provided the marginal algorithm satisfies a practically mild additional condition in (19), the set C is guaranteed to be small for the pseudo-marginal chain.

We start by assuming uniform integrability in a form given by the de la Vallée–Poussin theorem (e.g., [25], page 19 T22). This allows us to quantify the strength of the sub-geometric drift in a convenient way, for example, indicating that moment conditions imply polynomial drifts and consequently polynomial ergodicity.

CONDITION 28. There exists a nondecreasing convex function $\phi : [0, \infty) \to [1, \infty)$ satisfying

$$\liminf_{t\to\infty} \frac{\phi(t)}{t} = \infty \quad \text{and} \quad M_W := \sup_{x\in\mathsf{X}} \int \phi(w) Q_x(\mathrm{d}w) < \infty.$$

We record a simple implication of Condition 28.

LEMMA 29. *Assume Condition* 28 *holds. Then, there exists a function* $a(w) : [0, \infty) \to [0, \infty)$ *depending only on* M_W *and* ϕ *such that*

$$\sup_{y\in\mathsf{X}} \int_{u\ge w} u Q_y(\mathrm{d}u) \le a(w) \quad \textit{and} \quad \lim_{w\to\infty} a(w) = 0.$$

PROOF. For any function $f : [0, \infty) \to [0, \infty)$ nondecreasing in $[w, \infty)$, we have

$$\int_{u\ge w} u Q_y(\mathrm{d}u) \le \int u \frac{f(u)}{f(w)} Q_y(\mathrm{d}u).$$

The function $f(w) := \phi(w)/w$ is nondecreasing for w sufficiently large, therefore

$$\sup_{y\in\mathsf{X}} \int_{u\ge w} u Q_y(\mathrm{d}u) \le M_W \frac{w}{\phi(w)} =: a(w) \xrightarrow{w\to\infty} 0. \qquad \square$$

The next result establishes a drift away from large values of w for the pseudo-marginal chain, given that the marginal algorithm has an acceptance probability uniformly bounded away from zero. All uniformly (and geometrically) ergodic Markov chains satisfy this property [32], Proposition 5.1.

PROPOSITION 30. *Suppose that the one-step expected acceptance probability of the marginal algorithm is bounded away from zero,*

$$\alpha_0 := \inf_{x\in\mathsf{X}} \int q(x, \mathrm{d}y) \min\{1, r(x, y)\} > 0,$$

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

and Condition 28 *holds.*

Then, there exist constants $\delta > 0$ *and* $\bar{w} \in (1, \infty)$ *such that*

$$\tilde{P}V(x, w) \leq V(w) - \delta \frac{V(w)}{w} \mathbb{I}\{w \in [\bar{w}, \infty)\} + M_W \mathbb{I}\{w \in (0, \bar{w})\},$$

where $V(x, w) := V(w) := \phi(w)$. *The constants* δ *and* $\bar{w}$ *can be chosen to depend on* α_0, ϕ *and* M_W *only.*

PROOF. We can estimate

$$\begin{aligned}
&\tilde{P}V(x, w) - V(w) \\
&\quad = \iint q(x, \mathrm{d}y) Q_y(\mathrm{d}u) \min\left\{1, r(x, y)\frac{u}{w}\right\}[\phi(u) - \phi(w)] \\
&\quad \leq M_W - \iint q(x, \mathrm{d}y) Q_y(\mathrm{d}u) \min\left\{1, r(x, y)\frac{u}{w}\right\}\mathbb{I}\{u < w\}[\phi(w) - \phi(u)] \\
&\quad \leq M_W - \phi(w) \int q(x, \mathrm{d}y) \min\{1, r(x, y)\} \int_{u<w/2} Q_y(\mathrm{d}u)\frac{u}{w}\left[1 - \frac{\phi(w/2)}{\phi(w)}\right],
\end{aligned}$$

because $\min\{1, ab\} \geq \min\{1, a\}\min\{1, b\}$ for all $a, b \geq 0$. The convexity of ϕ implies $2\phi(w/2) \leq 1 + \phi(w)$, and therefore $\limsup_{w\to\infty} \phi(w/2)/\phi(w) \leq 1/2$. Because $\int_{u<w/2} Q_y(\mathrm{d}u)u = 1 - \int_{u\geq w/2} Q_y(\mathrm{d}u)u$, we may apply Lemma 29. Now, for any $\delta_0 \in (0, \alpha_0/2)$, there exists $\bar{w}_0 \in (1, \infty)$ such that

$$\tilde{P}V(x, w) - V(w) \leq M_W - \delta_0 \frac{\phi(w)}{w} \qquad \text{for all } w \in [\bar{w}_0, \infty).$$

The claim follows by taking $\bar{w} \in [\bar{w}_0, \infty)$ sufficiently large such that $\phi(w)/w > M_W/\delta_0$ for all $w \in [\bar{w}, \infty)$. □

In practice, Condition 28 is often verified for moments, that is, $\phi(w) = w^\beta$. We record the following corollary to highlight the straightforward connection of β to the polynomial drift rate.

COROLLARY 31. *Suppose the conditions of Proposition* 30 *hold with* $\phi(w) = w^\beta + 1$ *for some* $\beta > 1$. *Then, the pseudo-marginal kernel satisfies the drift condition*

$$\tilde{P}V(x, w) \leq V(w) - \delta V^{(\beta-1)/\beta}(w) + b_V \mathbb{I}\{w \in (0, \bar{w})\},$$

where $V(w) := w^\beta + 1$ *and* $b_V := M_W + \delta V^{(\beta-1)/\beta}(\bar{w})$.

PROOF. Follows from Proposition 30 observing that $w \leq (1 + w^\beta)^{1/\beta} = V(w)^{1/\beta}$. □

Proposition 30 and Corollary 31 establish a drift toward the set $\mathsf{X} \times (0, \bar{w}]$. They imply sub-geometric convergence of the Markov chain, with the following lemma showing that the set $\mathsf{X} \times (0, \bar{w}]$ is small.

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

LEMMA 32. *Denote the (sub-probability) kernel* $P_{\mathrm{acc}}(x, A) := \int_A q(x, \mathrm{d}y) \times \min\{1, r(x, y)\}$. *Suppose there exists* $\varepsilon > 0$, *an integer* $n \in [1, \infty)$ *and a probability measure* ν *on* $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ *such that for any* $A \in \mathcal{B}(\mathsf{X})$,

$$P^n_{\mathrm{acc}}(x, A) \geq \varepsilon \nu(A) \qquad \text{for all } x \in \mathsf{X}. \tag{19}$$

Then, there exists $\bar{w}_0 \in (1, \infty)$, $\tilde{\varepsilon} > 0$ *and a probability measure* $\tilde{\nu}$ *on* $(\mathsf{X} \times \mathsf{W}, \mathcal{B}(\mathsf{X}) \times \mathcal{B}(\mathsf{W}))$ *such that for all* $\bar{w} \in [\bar{w}_0, \infty)$,

$$\tilde{P}^n(x, w; \cdot) \geq \frac{\tilde{\varepsilon}}{\bar{w}} \tilde{\nu}(\cdot) \qquad \text{for all } (x, w) \in \mathsf{X} \times (0, \bar{w}].$$

PROOF. Choose $\bar{w}_0 > 1$ sufficiently large so that $\varepsilon_W := \inf_{y \in \mathsf{X}} \int Q_y(\mathrm{d}u) \times \min\{\bar{w}_0, u\} > 0$; such $\bar{w}_0$ exists due to Lemma 29 because

$$\int Q_y(\mathrm{d}u) \min\{\bar{w}_0, u\} \geq \int_{u < \bar{w}_0} Q_y(\mathrm{d}u) u = 1 - \int_{u \geq \bar{w}_0} Q_y(\mathrm{d}u) u.$$

We may write for $A \times B \in \mathcal{B}(\mathsf{X}) \times \mathcal{B}(\mathsf{W})$ and for $w \in (0, \bar{w}]$,

$$\begin{aligned}
\tilde{P}(x, w; A, B) &\geq \int_A q(x, \mathrm{d}y) \int_B Q_y(\mathrm{d}u) \min\left\{1, r(x, y)\frac{u}{w}\right\} \\
&\geq \int_A q(x, \mathrm{d}y) \min\{1, r(x, y)\} \int_B Q_y(\mathrm{d}u) \min\left\{1, \frac{u}{\bar{w}}\right\} \\
&\geq \frac{1}{\bar{w}} \int P_{\mathrm{acc}}(x, \mathrm{d}y) \hat{P}_W(y, B),
\end{aligned}$$

where $\hat{P}_W(y, B) = \int_B Q_y(\mathrm{d}u) \min\{\bar{w}_0, u\}$. We deduce recursively that

$$\begin{aligned}
\tilde{P}^n(x, w; A, B) &\geq \frac{1}{\bar{w}^n} \Big[\inf_{y \in \mathsf{X}} \hat{P}_W\big(y, (0, \bar{w}]\big)\Big]^{n-1} \int P^n_{\mathrm{acc}}(x, \mathrm{d}y) \hat{P}_W(y, B) \\
&\geq \frac{\varepsilon_W^{n-1} \varepsilon}{\bar{w}^n} \int_A \nu(\mathrm{d}y) \hat{P}_W(y, B) =: \frac{\varepsilon_W^{n-1} \varepsilon}{\bar{w}^n} \tilde{\nu}_0(A \times B).
\end{aligned}$$

We may take $\tilde{\nu}(A \times B) = \tilde{\nu}_0(A \times B)/\tilde{\nu}_0(\mathsf{X} \times \mathsf{W})$ and $\tilde{\varepsilon} = \varepsilon \tilde{\nu}_0(\mathsf{X} \times \mathsf{W}) > 0$. $\square$

REMARK 33. The condition in (19) is more stringent than assuming P uniformly ergodic. However, it is the most common way to establish the n-step minorisation condition $P^n(x, \cdot) \geq \varepsilon \nu(\cdot)$ in practice, which holds if and only if P is uniformly ergodic. In the case of a continuous state-space X where $q(x, \{y\}) = 0$ and $\nu(\{y\}) = 0$ for all $x, y \in \mathsf{X}$ and $n = 1$, the condition in (19) is in fact equivalent to $P(x, \cdot) \geq \varepsilon \nu(\cdot)$.

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

7. Polynomial ergodicity with a RWM as marginal algorithm. We consider next conditions which allow us to establish a polynomial drift condition for the pseudo-marginal algorithm in the case where the marginal algorithm is a geometrically ergodic random-walk Metropolis (RWM) targeting a super-exponentially decaying target with regular contours [16]. The existence of such a drift, together with additional simple assumptions, imply polynomial rates of ergodicity, but also Condition 20 (essential for the convergence of the pseudo-marginal asymptotic variance to that of the marginal algorithm) and a central limit theorem for example.

Our results rely on moment conditions on the distributions $Q_x(\mathrm{d}w)$. In Section 7.1 we assume the moments to be (essentially) uniform in x, while in Section 7.2 we consider the case where the behaviour of $Q_x(\mathrm{d}w)$ can get worse as $|x| \to \infty$. Note that the conditions in Section 7.2 may appear more general, but that they do not include all the cases covered by those of Section 7.1. This can be seen, for example, by comparing Conditions 37 and 46 and the admissible values of η in Theorem 38 and Corollary 47.

It is possible to extend our results beyond the polynomial case. For example one may assume the existence of exponential moment conditions; see Remark 39. For the sake of clarity and brevity, we have opted to detail here the polynomial case only.

REMARK 34. While our main focus here is on unbounded weight distributions, we will see that Lemma 49 suggests that geometric ergodicity is still possible when $Q_x((0, \bar{w}(x)]) = 1$ for all $x \in \mathbb{R}^d$, where $\bar{w} : \mathbb{R}^d \to [1, \infty)$ tends to infinity as $|x| \to \infty$. This is, however, a consequence of the strong assumption properties on the tails of π which confer the algorithm with a robustness property with respect to perturbations. Indeed, consider now the RWM on a compact subset $\mathsf{X} \subset \mathbb{R}^d$ with π bounded away from zero and infinity on X. It is not difficult to establish that if there does not exist $\bar{w} < \infty$ such that $Q_x([0, \bar{w}]) = 1$ for π-almost every $x \in \mathsf{X}$, then the chain cannot be geometrically ergodic; see, for example, the proof of Proposition 13.

Throughout this section, we denote the regions of almost sure acceptance and possible rejection for the marginal and pseudo-marginal and algorithms as

$$A_x := \left\{z \in \mathsf{X} : \frac{\pi(x+z)}{\pi(x)} \geq 1\right\}, \qquad R_x := A_x^{\complement},$$

$$A_{x,w} := \left\{(z, u) \in \mathsf{X} \times \mathsf{W} : \frac{\pi(x+z)}{\pi(x)} \frac{u}{w} \geq 1\right\}, \qquad R_{x,w} := A_{x,w}^{\complement},$$

respectively, for all $x \in \mathsf{X}$ and $w \in \mathsf{W}$.

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

7.1. *Uniform moment bounds.* Consider the following moment condition on the distributions $\{Q_x\}_{x\in\mathsf{X}}$ where $\mathsf{X} = \mathbb{R}^d$.

CONDITION 35. Suppose there exist constants $\alpha' > 0$ and $\beta' > 1$ such that

$$M_W := \operatorname*{ess\,sup}_{x\in\mathsf{X}} \int (w^{-\alpha'} \vee w^{\beta'}) Q_x(\mathrm{d}w) < \infty, \tag{20}$$

where $a \vee b := \max\{a, b\}$ and the essential supremum is taken with respect to the Lebesgue measure.

We first establish the following simple lemma, used throughout this section, which guarantees that the moment condition above holds also for any intermediate exponents.

LEMMA 36. *Given* (20), *then for all* $\alpha \in [0, \alpha']$ *and* $\beta \in [0, \beta']$ *and any* $\gamma \in [-\alpha', \beta]$

$$\operatorname*{ess\,sup}_{x\in\mathsf{X}} \int (w^{-\alpha} \vee w^{\beta}) Q_x(\mathrm{d}w) \le M_W \quad \textit{and} \quad \operatorname*{ess\,sup}_{x\in\mathsf{X}} \int w^{\gamma} Q_x(\mathrm{d}w) \le M_W.$$

PROOF. The first inequality follows by observing that $w^{-\alpha} \vee w^{\beta} \le w^{-\alpha'} \vee w^{\beta'}$ for all $w > 0$. For the second one, suppose first that $\gamma \in [0, \beta']$. Then $w^{\gamma} \le w^{-\alpha'} \vee w^{\gamma}$, and the result follows from the first inequality. The case $\gamma \in [-\alpha', 0]$ is similar. □

The following condition for the target density π was introduced in [16].

CONDITION 37. The target distribution π has a density with respect to the Lebesgue measure (also denoted π) which is continuously differentiable and supported on $\mathbb{R}^d$. The tails of π are super-exponentially decaying and have regular contours, that is,

$$\lim_{|x|\to\infty} \frac{x}{|x|} \cdot \nabla \log \pi(x) = -\infty \quad \text{and} \quad \limsup_{|x|\to\infty} \frac{x}{|x|} \cdot \frac{\nabla\pi(x)}{|\nabla\pi(x)|} < 0,$$

respectively, where $|x|$ denotes the Euclidean norm of $x \in \mathbb{R}^d$. Moreover, the proposal distribution satisfies $q(x, A) = q(A - x) = \int_A q(y - x)\,\mathrm{d}y$ with a symmetric density q bounded away from zero in some neighbourhood of the origin.

The following theorem establishes a polynomial drift given the conditions above.

THEOREM 38. *Suppose* $\tilde{P}$ *is a pseudo-marginal kernel with distributions* $Q_x(\mathrm{d}w)$ *satisfying Condition* 35 *with some constants* $\alpha' > 0$ *and* $\beta' > 1$, *and that*

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

the corresponding marginal algorithm is a random walk Metropolis with invariant density π *and proposal density* q *satisfying Condition* 37.

Define $V : \mathsf{X} \times \mathsf{W} \to [1, \infty)$ *as follows*:

$$V(x, w) := c_\pi^\eta \pi^{-\eta}(x)\big(w^{-\alpha} \vee w^\beta\big) \qquad \textit{where } c_\pi := \sup_{z \in \mathsf{X}} \pi(z), \tag{21}$$

for some constants $\eta \in (0, \alpha' \wedge 1)$, $\alpha \in (\eta, \alpha']$ *and* $\beta \in (0, \beta' - \eta)$.

Then, there exists constants $\bar{w}, M, b \in [1, \infty)$, $\underline{w} \in (0, 1]$ *and* $\delta_V > 0$ *such that*

$$\tilde{P}V(x, w) \le \begin{cases} V(x, w) - \delta_V V^{(\beta-1)/\beta}(x, w), & \textit{for all } (x, w) \notin \mathsf{C}, \\ b, & \textit{for all } (x, w) \in \mathsf{C}, \end{cases} \tag{22}$$

where $\mathsf{C} := \{(x, w) \in \mathsf{X} \times \mathsf{W} : |x| \le M, w \in [\underline{w}, \bar{w}]\}$.

Moreover, b, δ_V *and* C *depend only on the marginal algorithm, the constants* α', β' *and* M_W *in Condition* 35 *and the chosen constants* α, β, η.

PROOF. Let $\bar{w} \in [1, \infty)$ and $\delta'_V > 0$ be as in Lemma 41, so that $\tilde{P}V(x, w) \le V(x, w) - \delta'_V V^{(\beta-1)/\beta}(x, w)$ for all $x \in \mathsf{X}$ and all $w \ge \bar{w}$. Then apply Lemma 42 with the fixed value of $\bar{w}$ to obtain a $M \in [1, \infty)$ and $\lambda \in [0, 1)$ such that

$$\tilde{P}V(x, w) \le \lambda V(x, w) = V(x, w) - (1 - \lambda)V(x, w), \tag{23}$$

for all $w \in (0, \bar{w}]$ and $|x| \ge M$. Lemma 43 implies that (23) holds with all $x \in \mathsf{X}$ and $w \in (0, \underline{w}]$, with some $\lambda' \in [0, 1)$. Because $V \ge 1$, we conclude the claim for $(x, w) \notin \mathsf{C}$ with $\delta_V := \min\{\delta'_V, 1 - \lambda, 1 - \lambda'\}$. Lemma 43 implies the case $(x, w) \in \mathsf{C}$.

The dependence on b, δ_V and C is clear from the proofs of Lemmas 42 and 43. □

REMARK 39. It is possible to generalise Theorem 38 for nonpolynomial moments. In particular, we may let $V(x, w) = c_\pi^\eta \pi^{-\eta}(x)\phi(w)$ where $\phi : (0, \infty) \to [1, \infty)$ is defined by

$$\phi(w) := \begin{cases} a(w), & w \in (0, 1], \\ b(w), & w \in (1, \infty), \end{cases}$$

with nonincreasing $a : (0, 1] \to [1, \infty)$ and nondecreasing $b : (1, \infty) \to [1, \infty)$ satisfying

$$\lim_{w \to 0+} w^{-\eta} a(w) = \infty \quad \text{and} \quad \lim_{w \to \infty} b(w)/w = \infty,$$

and for some $\gamma > \eta$

$$\operatorname*{ess\,sup}_{x \in \mathsf{X}} \int_0^1 a(w) Q_x(\mathrm{d}w) < \infty \quad \text{and} \quad \operatorname*{ess\,sup}_{x \in \mathsf{X}} \int_1^\infty b(w) w^\gamma Q_x(\mathrm{d}w) < \infty.$$

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

Note that $a(w)$ and $b(w)$ must grow at least polynomially as $w \to 0+$ and $w \to \infty$, respectively. For example, $b(w) = \exp(c_b w)$ allows one to establish the claim with the stronger drift condition

$$\tilde{P}V(x,w) \le V(x,w) - \hat{\delta}_V \frac{V(x,w)}{\log \circ V(x,w)} \qquad (x,w) \notin \mathsf{C},$$

instead of the polynomial drift in (21).

We conjecture that the negative moment condition and the presence of $w^{-\alpha}$ in the drift function are not necessary in order to establish polynomial ergodicity in general. It seems, however, difficult to establish a one-step drift condition without any control of the behaviour of the distributions Q_x near zero.

We first consider a simple result which is auxiliary to the other lemmas.

LEMMA 40. *We have the following bounds for all* $x, z \in \mathsf{X}$, $w > 0$, $\hat{\alpha} > 0$, *and* $\hat{\beta} > 1$:

$$\text{(i)} \quad \int \min\left\{1, \frac{u}{w}\right\} Q_x(\mathrm{d}u) \ge \frac{1}{w}\left(1 - \frac{1}{w^{\hat{\beta}-1}} \int u^{\hat{\beta}} Q_x(\mathrm{d}u)\right),$$

$$\text{(ii)} \quad \int_{\{u:(z,u)\in A_{x,w}\}} Q_{x+z}(\mathrm{d}u) \ge 1 - w^{\hat{\alpha}}\left(\frac{\pi(x)}{\pi(x+z)}\right)^{\hat{\alpha}} \int u^{-\hat{\alpha}} Q_{x+z}(\mathrm{d}z).$$

PROOF. The bound (i) follows by writing

$$\begin{aligned}\int \min\left\{1, \frac{u}{w}\right\} Q_x(\mathrm{d}u) &= \frac{1}{w}\left(1 - \int_{u \ge w} (u-w) Q_x(\mathrm{d}u)\right) \\ &\ge \frac{1}{w}\left(1 - \int_{u\ge w} u Q_x(\mathrm{d}u)\right),\end{aligned}$$

and using the estimate $\mathbb{I}\{u \ge w\} \le (u/w)^{\hat{\beta}-1}$. For (ii), similarly

$$\int_{\{u:(z,u)\in A_{x,w}\}} Q_{x+z}(\mathrm{d}u) = 1 - \int_{\{u < w(\pi(x)/(\pi(x+z)))\}} Q_{x+z}(\mathrm{d}u)$$

and use $\mathbb{I}\{u < w\frac{\pi(x)}{\pi(x+z)}\} \le u^{-\hat{\alpha}}(w\frac{\pi(x)}{\pi(x+z)})^{\hat{\alpha}}$. $\square$

We next consider the case where w is large, and establish a polynomial drift in this case.

LEMMA 41. *Suppose the conditions of Theorem* 38 *hold. Then, there exist constants* $\delta_V > 0$ *and* $\bar{w} \in [1, \infty)$ *such that*

$$\tilde{P}V(x,w) \le V(x,w) - \delta_V V^{(\beta-1)/\beta}(x,w) \qquad \textit{for all } x \in \mathsf{X} \textit{ and } w \in [\bar{w}, \infty).$$

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

PROOF. We may write for $w \geq \bar{w} \geq 1$

$$\frac{\tilde{P}V(x,w)}{V(x,w)} = \iint_{A_{x,w}} a_{x,w}(z,u) Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z) + \iint_{R_{x,w}} b_{x,w}(z,u) Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z),$$

where

$$a_{x,w}(z,u) := \left(\frac{\pi(x)}{\pi(x+z)}\right)^{\eta} \frac{u^{-\alpha} \vee u^{\beta}}{w^{\beta}}, \tag{24}$$

$$b_{x,w}(z,u) := \left(\frac{\pi(x+z)}{\pi(x)}\right)^{1-\eta} \frac{u^{1-\alpha} \vee u^{1+\beta}}{w^{1+\beta}} + \left(1 - \frac{\pi(x+z)}{\pi(x)} \frac{u}{w}\right). \tag{25}$$

We now estimate both integrals by partitioning their integration domains into their intersections with the acceptance and the rejection sets of the marginal algorithm. For notational simplicity we denote $A_{x,w} \cap R_x = A_{x,w} \cap (R_x \times \mathsf{W})$ etc.

The bound for the first integral is straightforward,

$$\iint_{A_{x,w} \cap A_x} a_{x,w}(z,u) Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z) \leq \frac{M_W}{w^{\beta}}.$$

For the second one, observe that $1 \leq (\frac{\pi(x+z)}{\pi(x)} \frac{u}{w})^{\eta}$ on $A_{x,w}$, implying

$$\begin{aligned}
&\iint_{A_{x,w} \cap R_x} a_{x,w}(z,u) Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z) \\
&\quad \leq \frac{1}{w^{\beta+\eta}} \iint_{A_{x,w} \cap R_x} u^{\eta-\alpha} \vee u^{\eta+\beta} Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z) \leq \frac{M_W}{w^{\beta+\eta}},
\end{aligned}$$

because $\beta + \eta \leq \beta'$. Similarly, because $(\frac{\pi(x+z)}{\pi(x)} \frac{u}{w})^{1-\eta} \leq 1$ on $R_{x,w}$ we have

$$\begin{aligned}
&\iint_{R_{x,w}} \left(\frac{\pi(x+z)}{\pi(x)}\right)^{1-\eta} \frac{u^{1-\alpha} \vee u^{1+\beta}}{w^{1+\beta}} Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z) \\
&\quad \leq \frac{1}{w^{\beta+\eta}} \iint_{R_{x,w}} u^{\eta-\alpha} \vee u^{\eta+\beta} Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z) \leq \frac{M_W}{w^{\beta+\eta}}.
\end{aligned}$$

We now turn to the crucial remainder, which approaches unity as w grows.

$$\begin{aligned}
&\iint_{R_{x,w}} \left(1 - \frac{\pi(x+z)}{\pi(x)} \frac{u}{w}\right) Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z) \\
&\quad = 1 - \iint \min\left\{1, \frac{\pi(x+z)}{\pi(x)} \frac{u}{w}\right\} Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z) \\
&\quad \leq 1 - \iint \min\left\{1, \frac{\pi(x+z)}{\pi(x)}\right\} \min\left\{1, \frac{u}{w}\right\} Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z) \\
&\quad \leq 1 - \frac{\nu}{w} \int_{\{z:(\pi(x+z)/\pi(x)) \geq \nu\}} \left(1 - \frac{M_W}{w^{\beta'-1}}\right) q(\mathrm{d}z),
\end{aligned}$$

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

by Lemma 40(i), where $\nu \in (0,1)$. Lemma 58(ii) in Appendix D implies the existence of a $\nu > 0$ such that $\inf_{x \in \mathsf{X}} q(\{z : \frac{\pi(x+z)}{\pi(x)}\} \geq \nu) > 0$. Therefore, there exists a $\nu_2 \in (0,\nu)$, such that whenever w is sufficiently large

$$\iint_{R_{x,w}} \left(1 - \frac{\pi(x+z)}{\pi(x)}\frac{u}{w}\right) Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z) \leq 1 - \frac{\nu_2}{w}.$$

Because $\beta > 1$, the terms of the order $w^{-\beta}$ or $w^{-\eta-\beta}$ vanish faster than w^{-1} when w increases. Consequently, we have for any $\nu_3 \in (0,\nu_2)$, by further assuming w sufficiently large, that

$$\begin{aligned}\tilde{P}V(x,w) &\leq \left(1 - \frac{\nu_3}{w}\right) V(x,w) \\ &= V(x,w) - \nu_3 V^{\kappa}(x,w)\bigl(c_\pi \pi^{-\eta}(x)\bigr)^{1-\kappa} \leq V(x,w) - \nu_3 V^{\kappa}(x,w),\end{aligned}$$

where $\kappa = \frac{\beta-1}{\beta} \in (0,1)$. □

Next we deduce that in the regime where w is bounded, we have a geometric drift.

LEMMA 42. *Assume the conditions of Theorem* 38 *hold, and let* $\bar{w} \in [1,\infty)$. *Then, there exist constants* $\lambda \in [0,1)$ *and* $M \in [1,\infty)$ *such that*

$$\tilde{P}V(x,w) \leq \lambda V(x,w) \qquad \textit{for all } w \in (0,\bar{w}],\ |x| \geq M.$$

PROOF. We may write

$$\begin{aligned}\frac{\tilde{P}V(x,w)}{V(x,w)} &= 1 + \iint_{A_{x,w}} \hat{a}_{x,w}(z,u) Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z) \\ &\quad + \iint_{R_{x,w}} \hat{b}_{x,w}(z,u) Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z),\end{aligned}$$

where

$$\hat{a}_{x,w}(z,u) := \left(\frac{\pi(x)}{\pi(x+z)}\right)^{\eta} \frac{u^{-\alpha} \vee u^{\beta}}{w^{-\alpha} \vee w^{\beta}} - 1, \tag{26}$$

$$\hat{b}_{x,w}(z,u) := \left(\frac{\pi(x+z)}{\pi(x)}\right)^{1-\eta} \frac{u}{w}\left[\frac{u^{-\alpha} \vee u^{\beta}}{w^{-\alpha} \vee w^{\beta}} - \left(\frac{\pi(x+z)}{\pi(x)}\right)^{\eta}\right]. \tag{27}$$

Fix a constant $c > 1$ and define the following subsets: $\bar{A}_x := \{z : \frac{\pi(x+z)}{\pi(x)} \geq c\}$ and $\bar{R}_x := \{z : \frac{\pi(x+z)}{\pi(x)} \leq \frac{1}{c}\}$, and the annulus between these two sets as $D_x := (\bar{A}_x \cup \bar{R}_x)^{\complement} = \{z : \frac{1}{c} < \frac{\pi(x+z)}{\pi(x)} < c\}$. Compute

$$\begin{aligned}&\int_{D_x}\int_{(z,u)\in A_{x,w}} \hat{a}_{x,w}(z,u) Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z) \\ &\quad \leq \frac{c^{\eta}}{w^{-\alpha} \vee w^{\beta}} \int_{D_x}\int u^{-\alpha} \vee u^{\beta} Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z) \leq M_W c^{\eta} q(D_x)\end{aligned} \tag{28}$$

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

and

$$\begin{aligned}
&\int_{D_x}\int_{(z,u)\in R_{x,w}} \hat{b}_{x,w}(z,u) Q_{x+z}(\mathrm{d}u)q(\mathrm{d}z) \\
&\le c^{1-\eta}\int_{D_x}\int_{u<cw}\left(\frac{u}{w}\right)\frac{u^{-\alpha}\vee u^{\beta}}{w^{-\alpha}\vee w^{\beta}} Q_{x+z}(\mathrm{d}u)q(\mathrm{d}z) \le M_W c^{2-\eta} q(D_x).
\end{aligned} \tag{29}$$

Let then $\gamma \in (\eta, \alpha \wedge 1)$ such that $\gamma + \beta \le \beta'$ and observe that $(\frac{\pi(x+z)}{\pi(x)}\frac{u}{w})^{1-\gamma} \le 1$ on $R_{x,w}$, and thereby

$$\begin{aligned}
&\int_{\bar{R}_x}\int_{(z,u)\in R_{x,w}} \hat{b}_{x,w}(z,u) Q_{x+z}(\mathrm{d}u)q(\mathrm{d}z) \\
&\le \int_{\bar{R}_x}\int_{(z,u)\in R_{x,w}}\left(\frac{\pi(x+z)}{\pi(x)}\right)^{\gamma-\eta}\frac{u^{\gamma-\alpha}\vee u^{\gamma+\beta}}{w^{\gamma-\alpha}\vee w^{\gamma+\beta}} Q_{x+z}(\mathrm{d}u)q(\mathrm{d}z) \\
&\le M_W c^{-(\gamma-\eta)}.
\end{aligned} \tag{30}$$

Similarly, observe that $(\frac{\pi(x)}{\pi(x+z)}\frac{w}{u})^{\gamma} \le 1$ on $A_{x,w}$ and so

$$\begin{aligned}
&\int_{\bar{R}_x}\int_{(z,u)\in A_{x,w}} \hat{a}_{x,w}(z,u) Q_{x+z}(\mathrm{d}u)q(\mathrm{d}z) \\
&\le \frac{c^{-(\gamma-\eta)}}{w^{\gamma-\alpha}\vee w^{\gamma+\beta}}\int_{\bar{R}_x}\int_{(z,u)\in A_{x,w}} u^{\gamma-\alpha}\vee u^{\gamma+\beta} Q_{x+z}(\mathrm{d}u)q(\mathrm{d}z) \\
&\le M_W c^{-(\gamma-\eta)}.
\end{aligned} \tag{31}$$

It holds that $1 \le (\frac{\pi(x)}{\pi(x+z)}\frac{w}{u})$ on $R_{x,w}$, so we have

$$\begin{aligned}
&\int_{\bar{A}_x}\int_{(z,u)\in R_{x,w}} \hat{b}_{x,w}(z,u) Q_{x+z}(\mathrm{d}u)q(\mathrm{d}z) \\
&\le \frac{1}{w^{-\alpha}\vee w^{\beta}}\int_{\bar{A}_x}\int_{(z,u)\in R_{x,w}}\left(\frac{\pi(x+z)}{\pi(x)}\right)^{-\eta} u^{-\alpha}\vee u^{\beta} Q_{x+z}(\mathrm{d}u)q(\mathrm{d}z) \\
&\le M_W c^{-\eta}.
\end{aligned} \tag{32}$$

We are left with the term that will yield the geometric drift when $|x|$ is large,

$$\begin{aligned}
&\int_{\bar{A}_x}\int_{(z,u)\in A_{x,w}} \hat{a}_{x,w}(z,u) Q_{x+z}(\mathrm{d}u)q(\mathrm{d}z) \\
&\qquad \le \frac{M_W c^{-\eta}}{w^{-\alpha}\vee w^{\beta}} - \int_{\bar{A}_x} q(\mathrm{d}z)\int_{\{u:(z,u)\in A_{x,w}\}} Q_{x+z}(\mathrm{d}u) \\
&\qquad \le M_W c^{-\eta} - q(\bar{A}_x)\left(1 - M_W\left(\frac{w}{c}\right)^{\alpha'}\right),
\end{aligned}$$

by Lemma 40(ii). Lemma 58(iii) implies that $\delta := \liminf_{|x|\to\infty} q(\bar{A}_x) > 0$.

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

Let $\delta' \in (0, \delta)$ and fix $\varepsilon > 0$ sufficiently small so that $6\varepsilon - \delta(1-\varepsilon)^2 \le -\delta'$, and let $c > 1$ be sufficiently large so that $M_W c^{-\eta} \le \varepsilon$ and $M_W(\frac{\bar{w}}{c})^{\alpha'} \le \varepsilon$, and also that all (30), (31) and (32) are bounded by ε. Condition 37 implies that $\limsup_{|x|\to\infty} q(D_x) = 0$, and therefore there exists $M = M(c, \varepsilon) > 0$ such that (28) + (29) $\le \varepsilon$ for all $|x| \ge M$. By possibly increasing the bound M to ensure that $q(\bar{A}_x) \ge \delta(1-\varepsilon)$, we have that the claim holds for all $|x| \ge M$ with the constant $\lambda = 1 - \delta'$. □

We complete the results above by considering in particular very small values of w.

LEMMA 43. *Suppose the conditions of Theorem* 38 *hold, and let* $\bar{w}, M \in [1, \infty)$. *Then, there exist constants* $\underline{w} \in (0, 1)$, $\lambda \in (0, 1)$ *and* $b \in [1, \infty)$ *such that*

$$\tilde{P}V(x, w) \le b, \qquad \textit{for } |x| \le M \textit{ and } w \in [\underline{w}, \bar{w}], \tag{33}$$

$$\tilde{P}V(x, w) \le \lambda V(x, w), \qquad \textit{for } x \in \mathsf{X} \textit{ and } w \in (0, \underline{w}]. \tag{34}$$

PROOF. From the proof of Lemma 42, we have

$$\frac{\tilde{P}V(x, w)}{V(x, w)} \le 1 - \left(\iint_{A_{x,w}} Q_{x+z}(\mathrm{d}u)q(\mathrm{d}z)\right) + \tilde{a}_{x,w} + \tilde{b}_{x,w},$$

where

$$\tilde{a}_{x,w} := \iint_{A_{x,w}} \left(\frac{\pi(x)}{\pi(x+z)}\right)^{\eta} \frac{u^{-\alpha} \vee u^{\beta}}{w^{-\alpha} \vee w^{\beta}} Q_{x+z}(\mathrm{d}u)q(\mathrm{d}z),$$

$$\tilde{b}_{x,w} := \iint_{R_{x,w}} \left(\frac{\pi(x+z)}{\pi(x)}\right)^{1-\eta} \frac{u}{w} \frac{u^{-\alpha} \vee u^{\beta}}{w^{-\alpha} \vee w^{\beta}} Q_{x+z}(\mathrm{d}u)q(\mathrm{d}z).$$

Because $(\frac{\pi(x)}{\pi(x+z)}\frac{w}{u})^{\eta} \le 1$ on $A_{x,w}$ and $(\frac{\pi(x+z)}{\pi(x)}\frac{u}{w})^{1-\eta} \le 1$ on $R_{x,w}$,

$$\tilde{a}_{x,w} + \tilde{b}_{x,w} \le \iint \frac{u^{\eta-\alpha} \vee u^{\eta+\beta}}{w^{\eta-\alpha} \vee w^{\beta+\alpha}} Q_{x+z}(\mathrm{d}u)q(\mathrm{d}z) \le \frac{M_W}{w^{\eta-\alpha} \vee w^{\beta+\alpha}}.$$

This is enough to show that $\tilde{P}V(x, w) \le (1 + M_W)V(x, w)$ for all $(x, w) \in \mathsf{X} \times \mathsf{W}$. Because V is bounded on $\{|x| \le M, w \in [\underline{w}, \bar{w}]\}$, this implies the existence of $b = b(\bar{w}, \underline{w}, M) < \infty$ such that (33) holds.

Consider then (34). Let $\delta > 0$ be small enough so that $\inf_{x\in\mathsf{X}} q(A_x^{\delta}) \ge \varepsilon > 0$, where $A_x^{\delta} := \{z : \frac{\pi(x+z)}{\pi(x)} \ge \delta\}$. Then

$$\begin{aligned}\iint_{A_{x,w}} Q_{x+z}(\mathrm{d}u)q(\mathrm{d}z) &\ge \int_{A_x^{\delta}} q(\mathrm{d}z) \int_{\{u:(z,u)\in A_{x,w}\}} Q_{x+z}(\mathrm{d}u) \\ &\ge \int_{A_x^{\delta}} q(\mathrm{d}z)\left(1 - M_W\left(\frac{w}{\delta}\right)^{\alpha'}\right) \ge \frac{\varepsilon}{2}\end{aligned}$$

for $w \in (0, \underline{w}]$ if $\underline{w}$ is small enough. We may further decrease $\underline{w}$ to ensure that $\tilde{a}_{x,w} + \tilde{b}_{x,w} \le \varepsilon/4$ for all $w \in (0, \bar{w}]$ and conclude (34) with $\lambda = 1 - \varepsilon/4$. □

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

7.2. *Nonuniform moment bounds.* We replace the uniform moments in Condition 35 here with the following assumption, which allows the moments of the distributions $\{Q_x\}_{x \in \mathsf{X}}$ to grow in the tails of π.

CONDITION 44. Let $\hat{w} : \mathsf{X} \to [1, \infty)$ be a function bounded on compact sets and tending to infinity as $|x| \to \infty$. Let $\psi : (0, \infty) \to [1, \infty)$ be a nonincreasing function such that $\psi(t) \to \infty$ as $t \to 0$, and define $g(x) := \psi(\pi(x))$.

(i) There exist constants $\alpha' > 0$ and $\beta' > 1$ such that

$$\operatorname*{ess\,sup}_{x \in \mathsf{X}} g^{-1}(x) \int u^{-\alpha'} \vee u^{\beta'} Q_x(\mathrm{d}u) \leq 1,$$

where the essential supremum is taken with respect to the Lebesgue measure.

(ii) There exist constants $\xi_w \in (0, \beta' - 1)$ and $\xi_\pi \in (0, \beta' - 1 - \xi_w)$,

$$\sup_{x \in \mathsf{X}} \frac{g(x)}{\hat{w}^{\xi_\pi}(x)} \sup_{z \in R_x} \left[\left(\frac{\pi(x+z)}{\pi(x)} \right)^{\xi_\pi} \frac{g(x+z)}{g(x)} \right] < \infty, \tag{35}$$

where $R_x := \{z : \frac{\pi(x+z)}{\pi(x)} < 1\}$ is the set of possible rejection for the marginal random-walk Metropolis algorithm.

(iii) For any constant $b > 1$, one must have

$$\sup_{x \in \mathsf{X}} \frac{M_W(b(|x| \vee 1))}{\hat{w}^{\xi_w}(x)} < \infty, \tag{36}$$

where $M_W : (0, \infty) \to (0, \infty)$ is defined as follows:

$$M_W(r) := \operatorname*{ess\,sup}_{|x| \leq r} \int u^{-\alpha'} \vee u^{\beta'} Q_x(\mathrm{d}u) \leq \operatorname*{ess\,sup}_{|x| \leq r} g(x),$$

where the essential supremum is taken with respect to the Lebesgue measure.

The assumptions in Condition 44 may appear rather implicit and technical at first. However they, together with additional assumptions required in Theorem 45 below, are implied by the more meaningful assumptions in Condition 46 and Corollary 47, whose proof may help the reader gain some intuition.

THEOREM 45. *Suppose $\tilde{P}$ is a pseudo-marginal kernel corresponding to a random walk Metropolis with invariant density π and increment proposal density q satisfying Condition* 37. *Suppose Condition* 44 *holds with some $\alpha' > 0$ and $\beta' > 1$. Define $V : \mathsf{X} \times \mathsf{W} \to [1, \infty)$ as in* (21), *where the constant exponents satisfy*

$$\eta \in (0, \alpha' \wedge (\beta' - 1 - \xi_w) \wedge (1 - \xi_\pi)), \qquad \alpha \in (\eta, \alpha'], \beta \in (1 + \xi_w - \eta, \beta' - \eta)$$

and $\eta \leq (\beta' - \beta) \wedge 1 - \xi_\pi$.

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

Furthermore, suppose that there exists a function $c : \mathsf{X} \to [1, \infty)$ *bounded on compact sets such that* $\limsup_{|x|\to\infty} c(x)e^{-x} < \infty$ *and*

$$\limsup_{|x|\to\infty} \frac{\hat{w}^{\xi_\pi}(x)}{c^{\xi_c}(x)} = 0 \qquad \text{where } \xi_c \in \big(0, \big[(\beta' - \beta) \wedge \alpha \wedge 1\big] - \eta - \xi_\pi\big), \tag{37}$$

and that for any constant $b \in [1, \infty)$

$$\limsup_{|x|\to\infty} M_W(b|x|) \max\left\{q(D_x), \frac{1}{c^{\eta}(x)}, \left(\frac{\hat{w}(x)}{c(x)}\right)^{\alpha'}\right\} = 0, \tag{38}$$

where $D_x := \{z : \frac{1}{c(x)} \le \frac{\pi(x+z)}{\pi(x)} \le c(x)\}$.

Then, there exist constants $\bar{w}, M, b \in [1, \infty)$, $\underline{w} \in (0, 1]$ *and* $\delta_V > 0$ *such that the polynomial drift inequality* (22) *holds. Furthermore, the constants depend only on those of the marginal algorithm, the quantities* α', β', ξ_w, ξ_π, ψ, $\hat{w}$ *involved in Condition* 44, *including the upper bounds in* (35) *and* (36) (*as a function of* b), *the chosen* η, α, β, c *and* ξ_c, *and the upper bounds* (37) *and* (38).

PROOF. The proof follows by applying Lemma 48 below and then Lemma 49 with c_w from Lemma 48, similarly to the proof of Theorem 38 by setting $\bar{w} := \sup_{|x|\le M} \bar{w}(x)$, and observing that V is bounded on C. The dependence on the various quantities is clear from the proofs of Lemmas 48 and 49. □

Before proving Lemmas 48 and 49, we give sufficient conditions to establish the conditions of Theorem 45.

CONDITION 46. Suppose Condition 37 holds and additionally there exists a constant $\rho > 1$ such that

$$\lim_{|x|\to\infty} \frac{x}{|x|^{\rho}} \cdot \nabla \log \pi(x) = -\infty.$$

Moreover, the increment proposal density q satisfies $q(x) \le \bar{q}(|x|)$ for some bounded differentiable nonincreasing function $\bar{q} : [0, \infty) \to [0, \infty)$ such that $\int_{\mathsf{X}} \bar{q}(|x|)\,\mathrm{d}x < \infty$.

COROLLARY 47. *Suppose Condition* 46 *is satisfied, and that*

$$\int u^{-\alpha'} \vee u^{\beta'} Q_x(\mathrm{d}u) \le c\big(1 \vee |x|\big)^{\rho'} \tag{39}$$

with some constants $c < \infty$ *and* $\rho' \in [0, \rho - 1)$. *Then, for any*

$$\eta \in \big(0, \alpha' \wedge (\beta' - 1) \wedge 1\big), \qquad \alpha \in (\eta, \alpha'], \beta \in (1 - \eta, \beta' - \eta)$$

and V *defined in* (21), *the drift inequality* (22) *holds, with constants* $\bar{w}, M, b \in [1, \infty)$, $\underline{w} \in (0, 1]$, *and* $\delta_V > 0$ *only depending on the marginal algorithm and* α', β', c, ρ' *in* (39) *and the chosen* α, β, *and* η.

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

PROOF. Choose the constants ξ_w and ξ_π sufficiently small so that the conditions on η, α, and β in Theorem 45 are satisfied.

Fix a unit vector $u \in \mathbb{R}^d$, and define the function $\hat{\psi} : \mathbb{R}_+ \to [1, \infty)$ such that

$$\hat{\psi}(\pi(ru)) = \begin{cases} r, & r \geq R_0, \\ R_0, & r \in [0, R_0), \end{cases}$$

where $R_0 \in [1, \infty)$; this is always possible because the function $r \mapsto \pi(ru)$ is bounded away from zero on compact sets and monotone decreasing on the tail.

Define then $g(x) = c_g \hat{\psi}^{\rho'}(\pi(x))$, where the value of the constant $c_g \geq 1$ will be fixed later. In order to guarantee that Condition 44(i) is satisfied for sufficiently large c_g, it is sufficient to show that

$$\limsup_{|x|\to\infty} g^{-1}(x)|x|^{\rho'} < \infty. \tag{40}$$

Due to Lemma 57 in Appendix D, if $|x|$ is sufficiently large, then $g(x) = g(\zeta_x|x|u)$ for some $\zeta_x \in [b^{-1}, b]$, where $b \in [1, \infty)$ is a constant. Therefore, $g^{-1}(x) \leq (b^{-1}|x|)^{-\rho'}$, implying (40).

Define then $\hat{w}(x) := g^{\zeta_w}(x)$, where $\zeta_w = \xi_\pi^{-1} \vee \xi_w^{-1} \in (1, \infty)$. It is easy to check similarly to (40) that

$$\sup_{x\in\mathsf{X}} \frac{g(x)}{\hat{w}^{\xi_\pi}(x)} + \frac{M_W(b(|x| \vee 1))}{\hat{w}^{\xi_w}(x)} \leq 1 + \sup_{x\in\mathsf{X}} \frac{c'(b|x|)^{\rho'}}{\hat{w}^{\xi_w}(x)} < \infty.$$

It is also easy to check that

$$\sup_{z\in R_x}\left[\left(\frac{\pi(x+z)}{\pi(x)}\right)^{\xi_\pi} \frac{g(x+z)}{g(x)}\right] = \sup_{z\in R_x}\left[\left(\frac{\pi(x+z)}{\pi(x)}\right)^{\xi_\pi} \left(\frac{\hat{\psi}(\pi(x+z))}{\hat{\psi}(\pi(x))}\right)^{\rho'}\right]$$

is uniformly bounded in $x \in \mathsf{X}$. This is because it is sufficient to check the condition in the tails along a ray, that is, only for $z = r|x|$, $r \geq 1$. We conclude about the existence of a constant $c_g \in [1, \infty)$ such that Condition 44 holds.

Choose $\varepsilon_c \in (0, \rho - 1 - \rho')$, and let $c(x) = \exp(|x|^{\varepsilon_c})$. It is easy to check that there exists ξ_c such that (37) and (38) hold, using Lemma 59 in Appendix D to estimate $q(D_x)$. $\square$

We start by establishing a polynomial drift when w is large.

LEMMA 48. *Suppose the conditions of Theorem 45 hold. Then there exist constants $c_w \in [1, \infty)$ and $\delta_V > 0$ such that letting $\bar{w}(x) := c_w \hat{w}(x)$,*

$$\tilde{P}V(x, w) \leq V(x, w) - \delta_V V^{(\beta-1)/\beta}(x, w) \qquad \textit{for all } x \in \mathbb{R}^d \textit{ and } w \in [\bar{w}(x), \infty).$$

PROOF. We may write

$$\frac{\tilde{P}V(x, w)}{V(x, w)} = \iint_{A_{x,w}} a_{x,w}(z, u) Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z) + \iint_{R_{x,w}} b_{x,w}(z, u) Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z),$$

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

where $a_{x,w}$ and $b_{x,w}$ are defined in (24) and (25), respectively.

In what follows, for any $\nu > 0$, we will denote by $b_\nu \in (0, \infty)$ a constant chosen so that for all $x \in \mathsf{X}$, $\{x + z : \frac{\pi(x+z)}{\pi(x)} \geq \nu\} \subset B(0, b_\nu(|x| \vee 1))$; see Lemma 58(i) in Appendix D. We also denote by $c \in [1, \infty)$ a constant whose value may change upon each appearance.

For the first integral, note that on $A_{x,w}$, $1 \leq (\frac{\pi(x+z)}{\pi(x)} \frac{u}{w})^\eta$, so denoting $\delta := \eta + \beta - 1 - \xi_w > 0$, we have for $w \geq \hat{w}(x)$,

$$\begin{aligned}
&\iint_{A_{x,w} \cap A_x} a_{x,w}(z, u) Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z) \\
&\quad \leq \iint_{A_{x,w} \cap A_x} \frac{u^{\eta-\alpha} \vee u^{\eta+\beta}}{w^{\eta+\beta}} Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z) \\
&\quad \leq \frac{1}{w^{1+\delta}} \left(\frac{M_W(b_1(|x| \vee 1))}{\hat{w}^{\xi_w}(x)} \right) \leq \frac{c}{w^{1+\delta}},
\end{aligned}$$

by Condition 44(iii). For the second one, let $\gamma \in (\eta + \xi_\pi, \beta' - \beta]$, $\gamma < 1$, and observe that $1 \leq (\frac{\pi(x+z)}{\pi(x)} \frac{u}{w})^\gamma$ on $A_{x,w}$, implying that with $\delta' := \gamma + \beta - 1 - \xi_\pi > 0$

$$\begin{aligned}
&\iint_{A_{x,w} \cap R_x} a_{x,w}(z, u) Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z) \\
&\quad \leq \int_{R_x} \left(\frac{\pi(x+z)}{\pi(x)} \right)^{\gamma-\eta} \frac{u^{\gamma-\alpha} \vee u^{\gamma+\beta}}{w^{\gamma+\beta}} Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z) \\
&\quad \leq \frac{1}{w^{1+\delta'}} \int_{R_x} \left[\left(\frac{\pi(x+z)}{\pi(x)} \right)^{\xi_\pi} \frac{g(x+z)}{g(x)} \right] \frac{g(x)}{\hat{w}^{\xi_\pi}(x)} q(\mathrm{d}z) \leq \frac{c}{w^{1+\delta'}},
\end{aligned}$$

whenever $w \geq \hat{w}(x)$, by Condition 44(i) and (ii). Similarly, because $(\frac{\pi(x+z)}{\pi(x)} \frac{u}{w})^{1-\gamma} \leq 1$ on $R_{x,w}$ we have for $w \geq \hat{w}(x)$,

$$\begin{aligned}
&\iint_{R_{x,w} \cap R_x} \left(\frac{\pi(x+z)}{\pi(x)} \right)^{1-\eta} \frac{u^{1-\alpha} \vee u^{1+\beta}}{w^{1+\beta}} Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z) \\
&\quad \leq \frac{1}{w^{1+\delta'}} \int_{R_x} \left[\left(\frac{\pi(x+z)}{\pi(x)} \right)^{\xi_\pi} \frac{g(x+z)}{g(x)} \right] \frac{g(x)}{\hat{w}^{\xi_\pi}(x)} q(\mathrm{d}z) \\
&\quad \leq \frac{c}{w^{1+\delta'}},
\end{aligned}$$

and similarly, because $(\frac{\pi(x+z)}{\pi(x)} \frac{u}{w})^{1-\eta} \leq 1$,

$$\begin{aligned}
&\iint_{R_{x,w} \cap A_x} \left(\frac{\pi(x+z)}{\pi(x)} \right)^{1-\eta} \frac{u^{1-\alpha} \vee u^{1+\beta}}{w^{1+\beta}} Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z) \\
&\quad \leq \frac{1}{w^{1+\delta}} \left(\frac{M_W(b_1(|x| \vee 1))}{\hat{w}^{\xi_w}(x)} \right) \leq \frac{c}{w^{1+\delta}}.
\end{aligned}$$

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

As in the proof of Lemma 41, we may apply Lemma 40(i) to obtain

$$\begin{aligned}
&\iint_{R_{x,w}} \left(1 - \frac{\pi(x+z)}{\pi(x)} \frac{u}{w}\right) Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z) \\
&\quad \le 1 - \frac{\nu}{w} \int_{\{z:(\pi(x+z)/\pi(x)) \ge \nu\}} \left(1 - \frac{1}{w^{\beta'-1}} \int u^{\beta'} Q_{x+z}(\mathrm{d}u)\right) q(\mathrm{d}z) \\
&\quad \le 1 - \frac{\nu}{w} \int_{\{z:(\pi(x+z)/\pi(x)) \ge \nu\}} q(\mathrm{d}z) \left(1 - \frac{1}{w^{\beta'-1-\xi_w}} \left(\frac{M_W(b_\nu(|x| \vee 1))}{\hat{w}^{\xi_w}(x)}\right)\right) \\
&\quad \le 1 - \frac{\nu}{w} \int_{\{z:(\pi(x+z)/\pi(x)) \ge \nu\}} q(\mathrm{d}z) \left(1 - \frac{c}{w^{\beta'-1-\xi_w}}\right),
\end{aligned}$$

where we may choose $\nu \in (0, 1)$ such that $\inf_{x \in \mathsf{X}} q(z : \frac{\pi(x+z)}{\pi(x)} \ge \nu) > 0$; Lemma 58(ii) ensures the existence of such a ν.

The terms of the order $w^{-(1+\delta)}$ or $w^{-(1+\delta')}$ vanish faster than w^{-1} as w increases. Consequently, we can choose $c_w \in [1, \infty)$ sufficiently large so that there exists a $\nu' > 0$ such that for all $x \in \mathsf{X}$ and $w \ge \bar{w}(x)$,

$$\begin{aligned}
\tilde{P}V(x, w) &\le \left(1 - \frac{\nu'}{w}\right) V(x, w) \\
&= V(x, w) - \delta_V V^\kappa(x, w) \big(c_\pi^\eta \pi^{-\eta}(x)\big)^{1-\kappa} \le V(x, w) - \delta_V V^\kappa(x, w),
\end{aligned}$$

where $\kappa = \frac{\beta-1}{\beta} \in (0, 1)$. □

Our last lemma concentrates on the cases where either $|x|$ is large and w bounded, or w is small.

LEMMA 49. *Assume the conditions of Theorem* 45 *hold and let* $\bar{w}(x) := c_w \hat{w}(x)$ *for some constant* $c_w \in [1, \infty)$. *Then, there exist constants* $\lambda \in (0, 1)$, $\underline{w} \in (0, 1)$, $M \in [1, \infty)$, *and* $c_V \in [1, \infty)$ *such that*

$$\tilde{P}V(x, w) \le \lambda V(x, w) \qquad \textit{for } |x| \ge M, w \in \big(\underline{w}, \bar{w}(x)\big], \tag{41}$$

$$\tilde{P}V(x, w) \le \lambda V(x, w) \qquad \textit{for } x \in \mathsf{X}, w \in (0, \underline{w}], \tag{42}$$

$$\tilde{P}V(x, w) \le c_V V(x, w) \qquad \textit{for } (x, w) \in \mathsf{X} \times \mathsf{W}. \tag{43}$$

PROOF. We may write

$$\begin{aligned}
\frac{\tilde{P}V(x, w)}{V(x, w)} = 1 &+ \iint_{A_{x,w}} \hat{a}_{x,w}(z, u) Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z) \\
&+ \iint_{R_{x,w}} \hat{b}_{x,w}(z, u) Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z),
\end{aligned}$$

where $\hat{a}_{x,w}$ and $\hat{b}_{x,w}$ are given as in (26) and (27).

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

Define the subsets $\bar{A}_x := \{z : \frac{\pi(x+z)}{\pi(x)} \geq c(x)\}$, $\bar{R}_x := \{z : \frac{\pi(x+z)}{\pi(x)} \leq \frac{1}{c(x)}\}$ and $D_x := (\bar{A}_x \cup \bar{R}_x)^{\complement} = \{z : \frac{1}{c(x)} < \frac{\pi(x+z)}{\pi(x)} < c(x)\}$. Lemma 57 in Appendix D implies the existence of $b_1 \in [1, \infty)$ and $M_0 \in [1, \infty)$ such that $\bar{A}_x \cup D_x + x \subset B(0, b_1(|x| \vee 1))$ for all $x \in \mathsf{X}$. We decompose the two sums above into sub-sums on $\bar{A}_x$ and $\bar{R}_x$, with again an obvious abuse of notation.

Observe that $1 \leq (\frac{\pi(x+z)}{\pi(x)} \frac{u}{w})^{\eta}$ on $A_{x,w}$ and $(\frac{\pi(x+z)}{\pi(x)} \frac{u}{w})^{1-\eta} \leq 1$ on $R_{x,w}$, implying

$$
\begin{aligned}
&\iint_{D_x \cap A_{x,w}} \hat{a}_{x,w}(z,u) Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z) + \iint_{D_x \cap R_{x,w}} \hat{b}_{x,w}(z,u) Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z) \\
&\leq \int_{D_x} \int \frac{u^{\eta-\alpha} \vee u^{\eta+\beta}}{w^{\eta-\alpha} \vee w^{\eta+\beta}} Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z) \qquad (44) \\
&\leq \frac{M_W(b_1(|x| \vee 1)) q(D_x)}{w^{\eta-\alpha} \vee w^{\eta+\beta}},
\end{aligned}
$$

because $\eta \leq (\beta' - \beta) \wedge \alpha$.

Let then $\gamma := \eta + \xi_\pi + \xi_c < (\beta' - \beta) \wedge \alpha \wedge 1$ and notice again that $(\frac{\pi(x+z)}{\pi(x)} \frac{u}{w})^{1-\gamma} \leq 1$ on $R_{x,w}$ and $(\frac{\pi(x)}{\pi(x+z)} \frac{w}{u})^{\gamma} \leq 1$ on $A_{x,w}$. Therefore,

$$
\begin{aligned}
&\iint_{\bar{R}_x \cap A_{x,w}} \hat{a}_{x,w}(z,u) Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z) + \iint_{\bar{R}_x \cap R_{x,w}} \hat{b}_{x,w}(z,u) Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z) \\
&\leq \int_{\bar{R}_x} \left(\frac{\pi(x+z)}{\pi(x)}\right)^{\gamma-\eta} \int \frac{u^{\gamma-\alpha} \vee u^{\gamma+\beta}}{w^{\gamma-\alpha} \vee w^{\gamma+\beta}} Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z) \\
&\leq \frac{1}{w^{\gamma-\alpha} \vee w^{\gamma+\beta}} \left(\frac{\hat{w}^{\xi_\pi}(x)}{c^{\xi_c}(x)}\right) \int_{\bar{R}_x} \left[\left(\frac{\pi(x+z)}{\pi(x)}\right)^{\xi_\pi} \frac{g(x+z)}{g(x)}\right] \frac{g(x)}{\hat{w}^{\xi_\pi}(x)} q(\mathrm{d}z),
\end{aligned}
$$

because $\frac{\pi(x+z)}{\pi(x)} \leq c^{-1}(x)$ on $\bar{R}_x$.

It holds that $1 \leq (\frac{\pi(x)}{\pi(x+z)} \frac{w}{u})$ on $R_{x,w}$, so we have

$$
\begin{aligned}
&\int_{\bar{A}_x} \int_{(z,u) \in R_{x,w}} \hat{b}_{x,w}(z,u) Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z) \\
&\leq \int_{\bar{A}_x} \left(\frac{\pi(x)}{\pi(x+z)}\right)^{\eta} \int_{(z,u) \in R_{x,w}} \frac{u^{-\alpha} \vee u^{\beta}}{w^{-\alpha} \vee w^{\beta}} Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z) \\
&\leq \frac{M_W(b_1(|x| \vee 1)) c^{-\eta}(x)}{w^{-\alpha} \vee w^{\beta}}.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&\int_{\bar{A}_x} \int_{(z,u) \in A_{x,w}} \hat{a}_{x,w}(z,u) Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z) \\
&\leq \frac{M_W(b_1(|x| \vee 1)) c^{-\eta}(x)}{w^{-\alpha} \vee w^{\beta}} - \int_{\bar{A}_x \cap A_{x,w}} Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z).
\end{aligned}
$$

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

Now, by Lemma 40(ii),

$$\begin{aligned}
&\int_{\bar{A}_x \cap A_{x,w}} Q_{x+z}(\mathrm{d}u) q(\mathrm{d}z) \\
&\quad \geq \int \left(1 - \left(\frac{w}{c(x)}\right)^{\alpha'} \int u^{-\alpha'} Q_{x+z}(\mathrm{d}u)\right) q(\mathrm{d}z) \\
&\quad \geq q(\bar{A}_x)\left[1 - M_W\big(b_1(|x| \vee 1)\big) c_w^{\alpha'} \left(\frac{\hat{w}(x)}{c(x)}\right)^{\alpha'}\right],
\end{aligned}$$

for all $w \in (0, c_w \hat{w}(x)]$.

Lemma 58(iii) in Appendix D implies that $\delta := \liminf_{|x| \to \infty} q(\bar{A}_x) > 0$. Condition 44 together with (37) and (38) imply

$$\limsup_{|x| \to \infty} \frac{\tilde{P} V(x, w)}{V(x, w)} \leq 1 - \delta, \tag{45}$$

and we may conclude (41), by choosing any $\lambda \in (1 - \delta, 1)$ and finding a sufficiently large $M \in [1, \infty)$ such that the claim holds.

Consider then (42) and assume $|x| \leq M$. It is easy to verify that (45) holds with some $\delta' > 0$ when taking $\limsup_{w \to 0+}$ in the terms of the earlier decomposition. Finally, it is easy to check that (43) holds for $|x| \leq M$ similarly as (44), and the general case follows from (41) and Lemma 48. $\square$

8. Concluding remarks. Our convergence rate results in Sections 3 and 5–7 allow one to establish central limit theorems. In the case where the pseudo-marginal kernel is variance bounding, that is, $\tilde{P}$ admits a spectral gap as discussed in Section 3, the central limit theorem (CLT) holds for all functions $f : \mathsf{X} \times \mathsf{W} \to \mathbb{R}$ such that $\tilde{\pi}(f^2) < \infty$ [31], Theorem 7. Specifically, we have for all $g : \mathsf{X} \to \mathbb{R}$ with $\pi(g^2) < \infty$,

$$\frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} [g(\tilde{X}_k) - \pi(g)] \xrightarrow{n \to \infty} \mathcal{N}(0, \mathrm{var}(g, \tilde{P})) \qquad \text{in distribution,} \tag{46}$$

where $\mathrm{var}(g, \tilde{P}) \in [0, \infty)$ is given in Definition 6. It is possible to deduce upper bounds for the asymptotic variance $\mathrm{var}(g, \tilde{P})$. Namely, Corollary 11 relates $\mathrm{var}(g, \tilde{P})$ to $\mathrm{var}(g, P)$, and from Lemma 52, (49),

$$\mathrm{var}(g, P) \leq \frac{1 + (1 - \mathrm{Gap}(P))}{1 - (1 - \mathrm{Gap}(P))} \int e_{g - \pi(g), P}(\mathrm{d}x) = \frac{2 - \mathrm{Gap}(P)}{\mathrm{Gap}(P)} \mathrm{var}_\pi(g),$$

where $e_{g - \pi(g), P}$ is a positive measure on $[-1, 1]$; see Lemma 52 in Appendix A. If the spectral gap of the marginal algorithm is not directly accessible, it can be bounded by the drift constants; see [8] and references therein, and also [19], Theorem 4.2(ii).

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

When $\tilde{P}$ is polynomially ergodic, the class of functions g for which the CLT (46) holds is related to the exponent in the polynomial drift. For the convenience of the reader, we reformulate here a result due to Jarner and Roberts [17].

THEOREM 50. *Suppose P is irreducible and aperiodic. Assume there exists $V : \mathsf{X} \times \mathsf{W} \to [1, \infty)$, $\alpha \in [0, 1)$, $b \in [0, \infty)$, $c \in (0, \infty)$, a petite set (e.g.,* [16, 26]) *$C \in \mathcal{B}(\mathsf{X}) \times \mathcal{B}(\mathsf{W})$ such that*

$$\tilde{P}V(x, w) \le V(x, w) - cV^{\alpha}(x, w) + b\mathbb{I}\{(x, w) \in C\}, \tag{47}$$

and that there exists $\eta \in [1 - \alpha, 1]$ with $\tilde{\pi}(V^{2\eta}) < \infty$ and

$$\sup_{(x,w)\in\mathsf{X}\times\mathsf{W}} \frac{|g(x)|}{V^{\alpha+\eta-1}(x, w)} < \infty,$$

then $\mathrm{var}(g, \tilde{P}) \in [0, \infty)$ *and the CLT* (46) *holds.*

Theorem 50 is a restatement of [17], Theorem 4.2, because the pseudo-marginal kernel $\tilde{P}$ is also irreducible and aperiodic if the marginal kernel P is. The asymptotic variance can also be upper bounded in the polynomial case; see [3] and [19], Theorem 5.2(ii) and Remark 5.3. It is also possible to deduce nonasymptotic mean square error bounds [19].

Finally some of our results apply directly to extensions of pseudo-marginal algorithms which directly make use of noisy estimates of the marginal's acceptance ratio [18, 27]. However, despite some similitudes and simplifications, the corresponding processes differ fundamentally in that $(X_k)_{k\ge 0}$ is a Markov chain in this case (as opposed to the pseudo-marginal scenario), and we are currently investigating these differences.

APPENDIX A: LEMMAS FOR SECTION 2

In this section, $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ is a generic measurable space, and μ is a probability measure on X. We consider the Hilbert space

$$L_0^2(\mathsf{X}, \mu) := \{f : \mathsf{X} \to \mathbb{R} : \mu(f) = 0, \mu(f^2) < \infty\},$$

equipped with the inner product $\langle f, g\rangle_\mu := \int_{\mathsf{X}} f(x)g(x)\mu(\mathrm{d}x)$. We denote the corresponding norm by $\|f\|_\mu := \langle f, f\rangle_\mu^{1/2}$ and the operator norm for $A : L_0^2(\mathsf{X}, \mu) \to L_0^2(\mathsf{X}, \mu)$ as $\|A\| := \sup\{\|Af\|_\mu : \|f\|_\mu = 1\}$.

LEMMA 51. *Let P_1 and P_2 be two Markov kernels on space X reversible with respect to μ, and define the family of interpolated kernels $H_\beta := P_1 + \beta(P_2 - P_1)$ for $\beta \in [0, 1]$ also reversible with respect to μ. Then*

$$A_\lambda(\beta) := (I - \lambda H_\beta)^{-1}(I + \lambda H_\beta) = I + 2\sum_{k=1}^{\infty} \lambda^k H_\beta^k$$

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

is a well-defined operator on $L_0^2(\mathsf{X}, \mu)$ for all $\lambda \in [0, 1)$ and $\beta \in [0, 1]$ as well as the right-hand derivatives, with limits taken with respect to the operator norm

$$\begin{aligned} A'_\lambda(\beta) &:= \lim_{h\to 0+} h^{-1}\big(A_\lambda(\beta+h) - A_\lambda(\beta)\big) \\ &= 2\lambda(I - \lambda H_\beta)^{-1}(P_2 - P_1)(I - \lambda H_\beta)^{-1}, \\ A''_\lambda(\beta) &:= \lim_{h\to 0+} h^{-1}\big(A'_\lambda(\beta+h) - A'_\lambda(\beta)\big) \\ &= 2\lambda(I - \lambda H_\beta)^{-1}(P_2 - P_1)A'_\lambda(\beta), \end{aligned}$$

for all $\lambda \in [0, 1)$ and $\beta \in [0, 1)$.

PROOF. The expression for $A_\lambda(\beta)$ follows by the Neumann series representation $(I - \lambda H_\beta)^{-1} = \sum_{k=0}^\infty (\lambda H_\beta)^k$ which is well defined because $\|(\lambda H_\beta)^k\| \le \lambda^k$. Let us check that $\beta \mapsto A_\lambda(\beta)$ is right differentiable on $[0, 1)$. Write for any $h \in (0, 1-\beta)$

$$\begin{aligned} A_\lambda(\beta+h) - A_\lambda(\beta) = \lambda h(I - \lambda H_\beta)^{-1}(P_2 - P_1) + \Delta_{\lambda,\beta,h}(I + \lambda H_\beta) \\ + \lambda h \Delta_{\lambda,\beta,h}(P_2 - P_1), \end{aligned}$$

where $\Delta_{\lambda,\beta,h} = (I - \lambda H_{\beta+h})^{-1} - (I - \lambda H_\beta)^{-1}$. The differentiability follows as soon as we show $\lim_{h\to 0+} h^{-1}(\Delta_{\lambda,\beta,h})$ exists. By the Neumann series representation, it is sufficient to show that $\lim_{h\to 0+} h^{-1}(H_{\beta+h}^k - H_\beta^k)$ exists for all $k \ge 0$. The claim is trivial with $k = 0$, and the cases $k \ge 1$ follow inductively by writing

$$\begin{aligned} H_{\beta+h}^k - H_\beta^k = hH_\beta^{k-1}(P_2 - P_1) + \big(H_{\beta+h}^{k-1} - H_\beta^{k-1}\big)H_\beta \\ + h\big(H_{\beta+h}^{k-1} - H_\beta^{k-1}\big)(P_2 - P_1). \end{aligned}$$

Because $(I - \lambda H_\beta)A_\lambda(\beta) = I + \lambda H_\beta$, we may write

$$\lambda h(P_2 - P_1) = (I - \lambda H_{\beta+h})\big(A_\lambda(\beta+h) - A_\lambda(\beta)\big) - \lambda h(P_2 - P_1)A_\lambda(\beta),$$

from which, multiplying with h^{-1} and taking limit as $h \to 0+$, we obtain

$$\lambda(P_2 - P_1) = (I - \lambda H_\beta)A'_\lambda(\beta) - \lambda(P_2 - P_1)A_\lambda(\beta). \tag{48}$$

The desired expression for $A'_\lambda(\beta)$ follows by observing that $I + A_\lambda(\beta) = 2(I - \lambda H_\beta)^{-1}$. Consider then $A''_\lambda(\beta)$. From (48), we obtain

$$\begin{aligned} &(I - \lambda H_\beta)h^{-1}\big(A'_\lambda(\beta+h) - A'_\lambda(\beta)\big) \\ &\quad = \lambda(P_2 - P_1)A'_\lambda(\beta+h) + \lambda(P_2 - P_1)h^{-1}\big(A_\lambda(\beta+h) - A_\lambda(\beta)\big). \end{aligned}$$

We conclude by taking limits as $h \to 0+$. $\square$

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

LEMMA 52. *Suppose* Π *is a Markov kernel reversible with respect to* μ, *and* $(X_n)_{n\geq 0}$ *is a Markov chain corresponding to the transition* Π *with* $X_0 \sim \mu$. *Then, for a function* $f \in L_0^2(\mathsf{X}, \mu)$

$$\operatorname{var}(f, \Pi) = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left(\sum_{i=1}^{n} f(X_i)\right)^2 = \int \frac{1+x}{1-x} e_{f,\Pi}(\mathrm{d}x) \in [0, \infty], \tag{49}$$

where $e_{f,\Pi}$ *is a positive measure on* $S \subset [-1, 1]$ *satisfying* $e_{f,\Pi}(S) = \|f\|_\mu^2$.

For any $f \in L_0^2(\mathsf{X}, \mu)$, *whenever the series below is convergent, then the following equality holds*:

$$\operatorname{var}_\mu(f) + 2\sum_{k=1}^{\infty} \mathbb{E}[f(X_0)f(X_k)] = \operatorname{var}(f, \Pi) < \infty. \tag{50}$$

Moreover,

$$\operatorname{var}_\lambda(f, \Pi) := \langle f, (I - \lambda\Pi)^{-1}(I + \lambda\Pi)f\rangle_\mu \in [0, \infty)$$

is well defined for all $\lambda \in [0, 1)$ *and satisfies* $\lim_{\lambda\to 1-} \operatorname{var}_\lambda(f, \Pi) = \operatorname{var}(f, \Pi)$ *and* $\langle f, (I - \lambda\Pi)^{-1} f\rangle \geq 0$.

The results in Lemma 52 are well known; a full proof is given in [6].

APPENDIX B: LEMMAS FOR SECTION 3

We include the statement of [11], Theorem A.2, for the sake of self-containedness.

LEMMA 53. *Let* A *and* B *be self-adjoint operators on a Hilbert space* $\mathcal{H}$ *satisfying* $0 \leq \langle f, Af\rangle \leq \langle f, Bf\rangle$ *for all* $f \in \mathcal{H}$, *and the inverses* A^{-1} *and* B^{-1} *exist. Then* $0 \leq \langle f, B^{-1}f\rangle \leq \langle f, A^{-1}f\rangle$ *for all* $f \in \mathcal{H}$.

LEMMA 54. *Suppose* P *is a Metropolis–Hastings kernel given in* (1), *and* $\rho(x)$ *is given in* (2). *Then the spectral gap of* P *defined in* (10) *satisfies*:

(i) *for any set* $A \in \mathcal{B}(\mathsf{X})$ *with* $\pi(A) \in (0, 1)$,

$$\operatorname{Gap}(P) \leq (1 - \pi(A))^{-1}\Bigl(1 - \inf_{x\in A} \rho(x)\Bigr);$$

(ii) *if* π *does not have point masses, that is,* $\pi(\{x\}) = 0$ *for all* $x \in \mathsf{X}$, *then*

$$\operatorname{Gap}(P) \leq 1 - \rho(x) \qquad \text{for } \pi\text{-almost every } x \in \mathsf{X}.$$

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

PROOF. We first check (i). Denote $p = \mathbb{P}(A) \in (0,1)$ and define $f(x) = a\mathbb{I}\{x \in A\} - b\mathbb{I}\{x \notin A\}$ where the constants $a, b \in (0, \infty)$ are chosen so that $\pi(f) = ap - b(1-p) = 0$ and $\pi(f^2) = a^2 p + b^2(1-p) = 1$. We may compute

$$\begin{aligned}\mathcal{E}_P(f) &= \frac{1}{2}\int \pi(\mathrm{d}x) q(x, \mathrm{d}y)\min\{1, r(x,y)\}\big[f(x) - f(y)\big]^2 \\ &= (a+b)^2 \int_A \pi(\mathrm{d}x)\int_{A^\complement} q(x,\mathrm{d}y)\min\{1, r(x,y)\} \\ &\le (a+b)^2\int_A \pi(\mathrm{d}x)\big(1-\rho(x)\big) \le (a+b)^2 p\Big(1 - \inf_{x\in A}\rho(x)\Big).\end{aligned}$$

Now, according to our choice of a and b,

$$(a+b)^2 p = \big(1 - b^2(1-p)\big) + 2b^2(1-p) + b^2 p = 1 + b^2 = (1-p)^{-1}.$$

Consider then (ii). The case $\mathrm{Gap}(P) = 0$ is trivial, so assume $\mathrm{Gap}(P) > 0$ and assume the claim does not hold. Then there exists an $\varepsilon > 0$ and a set $A \in \mathcal{B}(\mathsf{X})$ with $p := \mathbb{P}(A) \in (0,1)$ such that $1 - \rho(x) \le \mathrm{Gap}(P) - \varepsilon$ for all $x \in A$. From (i), $\mathrm{Gap}(P) \le (1-p)^{-1}(\mathrm{Gap}(P) - \varepsilon)$. Because π is not concentrated on points, we may choose p as small as we want, which leads to a contradiction. □

APPENDIX C: LEMMAS FOR SECTIONS 4 AND 5

LEMMA 55. *Suppose $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_n)$ are Markov chains on a common state space $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ with kernels P and Q, and initial distributions π and ϖ, respectively, which are invariant such that $\pi P = \pi$ and $\varpi Q = \varpi$. Then, the distributions of X and Y denoted as μ_X and μ_Y satisfy the following inequality for any $C \in \mathcal{B}(\mathsf{X})$:*

$$\|\mu_X - \mu_Y\| \le \|\pi - \varpi\| + 2(n-1)\pi\big(C^\complement\big) + (n-1)\sup_{x\in C}\big\|P(x,\cdot) - Q(x,\cdot)\big\|,$$

where $\|\mu_X - \mu_Y\| := \sup_{|f|\le 1}|\mu_X(f) - \mu_Y(f)|$ denotes the total variation.

PROOF. Let $A \in \mathcal{B}(\mathsf{X})$. We shall use the shorthand notation $x = x_{1:n} = (x_1, \ldots, x_n)$ and denote $g_P^{(1:n)}(x) := \mathbb{I}\{x \in A\}$,

$$g_P^{(1:k)}(x_{1:k}) := \int P(x_k, \mathrm{d}x_{k+1})\cdots\int P(x_{n-1}, \mathrm{d}x_n)\mathbb{I}\{x \in A\}, \qquad 2 \le k \le n-1,$$

and $g_P^{(1:1)} := g_P^{(1)}$, and define $g_Q^{(\cdot)}$ similarly using the kernel Q.

Note that $g_P^{(\cdot)}$ and $g_Q^{(\cdot)}$ take values between zero and one and the total variation satisfies $\|\pi - \varpi\| = 2\sup_{0\le f\le 1}|\pi(f) - \varpi(f)| = 2\sup_{A\in\mathcal{B}(\mathsf{X})}|\pi(A) - \varpi(A)|$.

$$\begin{aligned}\big|\mu_X(A) - \mu_Y(A)\big| &= \big|\pi\big(g_P^{(1)}\big) - \varpi\big(g_Q^{(1)}\big)\big| \\ &\le \big|\pi\big(g_Q^{(1)}\big) - \varpi\big(g_Q^{(1)}\big)\big| + \big|\pi\big(g_P^{(1)} - g_Q^{(1)}\big)\big| \\ &\le \frac{1}{2}\|\pi - \varpi\| + \big|\pi\big(g_P^{(1)} - g_Q^{(1)}\big)\big|,\end{aligned}$$

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

showing the claim for $n = 1$. Assume then $n \geq 2$ and observe that we can write $|\pi(g_P^{(1)} - g_Q^{(1)})| = |\mathbb{E}[g_P^{(1)}(X_1) - g_Q^{(1)}(X_1)]|$. We may continue inductively

$$
\begin{aligned}
&|\mathbb{E}[(g_P^{(1:n-1)} - g_Q^{(1:n-1)})(X_{1:n-1})]| \\
&\quad \leq |\mathbb{E}[(g_P^{(1:n)} - g_Q^{(1:n)})(X_{1:n})]| + \left|\mathbb{E}\left[\int \Delta(X_{n-1}, \mathrm{d}x_n) g_Q^{(1:n)}(X_{1:n-1}, x_n)\right]\right|,
\end{aligned}
$$

where $\Delta(x, \mathrm{d}y) := P(x, \mathrm{d}y) - Q(x, \mathrm{d}y)$, and observe that

$$
\begin{aligned}
&\left|\mathbb{E}\left[\int \Delta(X_{n-1}, \mathrm{d}x_n) g_Q^{(1:n)}(X_{1:n-1}, x_n)\right]\right| \\
&\quad \leq \mathbb{P}(X_{n-1} \notin C) + \sup_{x_{1:n-2} \in \mathcal{X}^{n-2}} \sup_{x_{n-1} \in C} \left|\int \Delta(x_{n-1}, \mathrm{d}x_n) g_Q^{(1:n)}(x_{1:n})\right| \\
&\quad \leq \pi(C^{\complement}) + \frac{1}{2} \sup_{x \in C} \|P(x, \cdot) - Q(x, \cdot)\|,
\end{aligned}
$$

because $|\int \Delta(X_{n-1}, \mathrm{d}x_n) g_Q^{(1:n)}(X_{1:n-1}, x_n)| \leq 1$ and $0 \leq g_Q^{(1:n)} \leq 1$. $\square$

LEMMA 56. *Assume $q \gg \pi$ and denote $\mu(x) := \pi(\mathrm{d}x)/q(\mathrm{d}x)$. Suppose that there exists a strictly increasing $\phi : (0, \infty) \to [1, \infty)$ with $\liminf_{t\to\infty} \phi(t)/t > 0$, such that*

$$
\int \pi(\mathrm{d}x)\phi(\mu(x)) < \infty. \tag{51}
$$

Then, there exist constants $M, c, \varepsilon \in (0, \infty)$ and a probability measure ν on $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ such that for the independent Metropolis–Hastings P,

$$
PV(x) \leq V(x) - cV(x)/\phi^{-1}(V(x)) \qquad \text{if } \mu(x) > M, \tag{52}
$$

$$
P(x; \cdot) \geq \varepsilon\nu(\cdot) \qquad \text{if } \mu(x) \leq M, \tag{53}
$$

and $\nu(V) < \infty$, where $V(x) := \phi(\mu(x))$.

PROOF. Denote $A_x := \{y \in \mathsf{X} : \frac{\mu(y)}{\mu(x)} \geq 1\}$ and $R_x := A_x^{\complement}$ and write

$$
\begin{aligned}
PV(x) &= \int_{A_x} \frac{V(y)}{\mu(y)} \pi(\mathrm{d}y) + \int_{R_x} \frac{V(y)}{\mu(x)} \pi(\mathrm{d}y) + V(x, w) \int_{R_x} \left(1 - \frac{\mu(y)}{\mu(x)}\right) q(\mathrm{d}y) \\
&\leq \frac{1}{\mu(x)} \int \pi(\mathrm{d}y) V(y) + V(x)\left(1 - \frac{\pi(R_x)}{\phi^{-1}(V(x))}\right),
\end{aligned}
$$

because $\mu(y) \geq \mu(x)$ on $A_{x,w}$. The first term on the right vanishes and $\pi(R_x) \to 1$ as $\mu(x) \to \infty$, and $\liminf_{u\to\infty} u/\phi^{-1}(u) > 0$, implying (52). For (53), observe that for $\mu(x) \leq M$,

$$
P(x, B) \geq \int_B \min\left\{\frac{1}{M}, \frac{1}{\mu(y)}\right\} \pi(\mathrm{d}y) =: \tilde{\nu}(B),
$$

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

and we can take $\varepsilon = \tilde{\nu}(\mathsf{X})$ and $\nu = \varepsilon^{-1}\tilde{\nu}$, for which (51) implies $\nu(V) < \infty$. $\square$

APPENDIX D: LEMMAS FOR SECTION 7

We denote by $n(x) := x/|x|$ the unit vector pointing in the direction of $x \neq 0$ and by $B(x,r) := \{y \in \mathbb{R}^d : |x - y| \leq r\}$ the (closed) Euclidean ball.

LEMMA 57. *Assume π satisfies Condition* 37, *and that $c : \mathsf{X} \to [1,\infty)$ satisfies $\limsup_{|x|\to\infty} c(x)e^{-|x|} < \infty$. Then, there exist constants $M, b \in [1,\infty)$ such that for all $|x| \geq M$,*

$$D_x := \left\{ y \in \mathbb{R}^d : \frac{1}{c(x)} \leq \frac{\pi(y)}{\pi(x)} \leq c(x) \right\} \subset B(0, b|x|) \setminus B(0, b^{-1}|x|).$$

PROOF. Let $c' > \limsup_{|x|\to\infty} c(x)e^{-|x|}$. Choose any $C \in (4c', \infty)$ and let $M_0 \in [1 \vee \log c', \infty)$ be sufficiently large so that there exists a $\beta_\pi \in (0,1]$ such that for all $|x| \geq M_0$,

$$c(x) \leq c' e^{|x|}, \qquad n(x) \cdot \nabla \log \pi(x) \leq -C \quad \text{and} \quad n(x) \cdot n(\nabla \pi(x)) < -\beta_\pi.$$

Let $\delta \in (0,1)$. Then for any $|x| \geq M_0(1-\delta)^{-1}$ and all $z = tn(x)$ with $|t| \leq \delta$, we have

$$\left| \log \frac{\pi(x+z)}{\pi(x)} \right| = |t| \int_0^1 |n(x + \lambda z) \cdot \nabla \log \pi(x + \lambda z)| \, \mathrm{d}\lambda \geq C|t|. \tag{54}$$

Now, if $|x| > aM_0$ where $a := \exp(2\pi \tan(\arccos(\beta_\pi)))$, then [33], Lemma 22, implies

$$\{y \in \mathbb{R}^d : \pi(y) = \pi(x)\} \subset B(0, a|x|) \setminus B(0, a^{-1}|x|). \tag{55}$$

Take any $M > 4aM_0$, and choose $|x| \geq M$. Then, condition (54) implies that any $z = \lambda x \in D_x$, where $\lambda > 0$ satisfies

$$|(\lambda - 1)|x|| \leq C^{-1} \log c(x) \leq C^{-1}(\log(c') + |x|) \leq 2C^{-1}|x|.$$

We deduce that $|\lambda - 1| < 1/2$. Again, using (55), we deduce that the claim holds with $b = 2a$. $\square$

LEMMA 58. *Assume π satisfies Condition* 37.

(i) *Then, for any constant $\nu \in (0,\infty)$, there exists a constant $b_\nu \in [1,\infty)$ such that for all $x \in \mathsf{X}$, $\{x + z : \frac{\pi(x+z)}{\pi(x)} \geq \nu\} \subset B(0, b_\nu(|x| \vee 1))$. Assume also that q satisfies Condition* 37.

(ii) *There exists a constant $\nu \in (0,\infty)$ such that* $\inf_{x\in\mathsf{X}} q(\{z : \frac{\pi(x+z)}{\pi(x)} \geq \nu\}) > 0$.

(iii) *For any constant $\nu \in (0,\infty)$, there exists a constant $M = M(\nu) \in [1,\infty)$ such that* $\inf_{|x|\geq M} q(\{z : \frac{\pi(x+z)}{\pi(x)} \geq \nu\}) > 0$.

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

PROOF. Consider first (i). The existence of such a finite constant follows for x in compact sets by the continuity of π and in the tails by Lemma 57.

The claim (ii) follows on compact sets by the continuity of $\log \pi$, and in the tails as in [16], proof of Theorem 4.3; the last claim (iii) follows similarly. $\square$

When the target and the proposal distributions satisfy also Condition 46, we have a decay rate for $q(D_x)$.

LEMMA 59. *Assume Condition* 46, *and assume* $\limsup_{|x|\to\infty} c(x)e^{-|x|} < \infty$. *Then, for any* $\varepsilon' > 0$ *there exists a constant* $M_0 \in [M, \infty)$ *such that for all* $|x| \geq M_0$,

$$q(D_x) \leq \varepsilon' \frac{\log(c(x))}{|x|^{\rho-1}} \qquad \textit{where } D_x := \left\{z \in \mathbb{R}^d : \frac{1}{c(x)} \leq \frac{\pi(x+z)}{\pi(x)} \leq c(x)\right\}.$$

PROOF. Lemma 57 implies $b \in [1, \infty)$ and $M' \in [1, \infty)$ such that for all $|x| \geq M'$ the annulus $D_x \subset B(0, b|x|) \setminus B(0, b^{-1}|x|)$. This implies that for any constant $c_\ell \in [1, \infty)$ one can choose $M_\ell \in [M', \infty)$ such that

$$n(x+z) \cdot \nabla \log \pi(x+z) \leq -c_\ell |x+z|^{\rho-1} \qquad \text{for all } |x| \geq M_\ell, z \in D_x.$$

Denoting $\ell(x) := \log \pi(x)$, we write

$$D_x = \{z \in \mathbb{R}^d : |\ell(x+z) - \ell(x)| \leq \log c(x)\}.$$

Define the contour surface set $S_{\pi(x)} := \{y \in \mathbb{R}^d : \pi(y) = \pi(x)\}$ and

$$C_{\pi(x)}(\delta) := \{y + tn(y) : y \in S_{\pi(x)}, |t| \leq \delta\}.$$

We will now check that with our conditions, for $|x| \geq M_\ell b$,

$$D_x + x \subset C_{\pi(x)}(\delta_x) \qquad \text{where } \delta_x := \frac{b^{\rho-1}}{c_\ell} \cdot \frac{\log c(x)}{|x|^{\rho-1}}. \tag{56}$$

Because $D_x + x = D_y + y$ whenever $\pi(x) = \pi(y)$, it is sufficient to consider $z \in D_x$ such that $z = tn(x)$ As in the proof of Lemma 57,

$$\begin{aligned}|\ell(x+z) - \ell(x)| &= |t| \int_0^1 |n(x+\lambda z) \cdot \nabla \ell(x+\lambda z)| \,\mathrm{d}\lambda \\ &\geq |t| c_\ell |x|^{\rho-1} \int_0^1 \left|1 + \frac{t}{|x|}\right|^{\rho-1} \mathrm{d}t \geq c_\ell b^{-(\rho-1)} |x|^{\rho-1} |t|.\end{aligned}$$

Now $|\ell(x+z) - \ell(x)| \leq \log c(x)$ implies (56).

Write then, by Fubini's theorem,

$$\begin{aligned}q(D_x) &\leq \int_{C_{\pi(x)}(\delta_x)-x} \bar{q}(z)\,\mathrm{d}z \\ &= \int_0^{\bar{q}(0)} \mathcal{L}^d\big(z \in \mathbb{R}^d : \bar{q}(|z|) \geq t, z \in C_{\pi(x)}(\delta_x) - x\big)\,\mathrm{d}t \\ &= \int_0^\infty \mathcal{L}^d\big(z \in \mathbb{R}^d : |z| \leq u, z \in C_{\pi(x)}(\delta_x) - x\big)|\bar{q}'(u)|\,\mathrm{d}u.\end{aligned}$$

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

Now, [16], proof of Theorem 4.1, shows that for $u \le |x|/2$,

$$\mathcal{L}^d\big(C_{\pi(x)}(\delta_x) \cap B(x,u)\big) \le \delta_x \left(\frac{|x|+u}{|x|-u}\right)^{d-1} \frac{\mathcal{L}^d(B(x,3u))}{u} \le 3^{2d-1} c_d \delta_x u^{d-1},$$

where $c_d = \mathcal{L}^d(B(0,1))$. By polar integration,

$$\begin{aligned}\mathcal{L}^d\big(C_{\pi(x)}(\delta_x)\big) &\le c_d \sup_{y \in S_{\pi(x)}} \int_{|y|-\delta_x}^{|y|+\delta_x} r^{d-1}\,\mathrm{d}r \\ &\le 2c_d b^{d-1}\delta_x |x|^{d-1} \le 4c_d b^{d-1}\delta_x u^{d-1},\end{aligned}$$

where the latter inequality holds for $u \ge |x|/2$. We obtain

$$q(D_x) \le c'\delta_x\left(1 + \int_0^\infty u^d |\hat{q}'(u)|\,\mathrm{d}u\right),$$

and because $\hat{q}$ is monotone decreasing, integration by substitution yields

$$\int_0^M u^d|\hat{q}'(u)|\,\mathrm{d}u = d\int_0^M u^{d-1}\hat{q}(u)\,\mathrm{d}u - M^d\hat{q}(M) \le dc_d^{-1}\int \hat{q}(x)\,\mathrm{d}x < \infty.$$

We deduce $q(D_x) \le c''\delta_x$ and conclude by choosing c_ℓ sufficiently large. □

Acknowledgement. We thank Anthony Lee for fruitful discussions.

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

[2] ANDRIEU, C. and FORT, G. (2005). Explicit control of subgeometric ergodicity. Research Report No. 05:17, Univ. Bristol.

[3] ANDRIEU, C., FORT, G. and VIHOLA, M. (2014). Quantitative convergence rates for subgeometric Markov chains. Available at arXiv:1309.0622v2.

[4] ANDRIEU, C. and ROBERTS, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.* **37** 697–725. MR2502648

[5] ANDRIEU, C. and VIHOLA, M. (2014). Establishing some order amongst exact approximation MCMCs. Available at arXiv:1404.6909v1.

[6] ANDRIEU, C. and VIHOLA, M. (2014). Convergence properties of pseudo-marginal Markov chain Monte Carlo algorithms. Available at arXiv:1210.1484v2.

[7] ANDRIEU, C. and VIHOLA, M. (2014). Markovian stochastic approximation with expanding projections. *Bernoulli* **20** 545–585. MR3178509

[8] BAXENDALE, P. H. (2005). Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Probab.* **15** 700–738. MR2114987

[9] BEAUMONT, M. A. (2003). Estimation of population growth or decline in genetically monitored populations. *Genetics* **164** 1139–1160.

[10] BESKOS, A., PAPASPILIOPOULOS, O., ROBERTS, G. O. and FEARNHEAD, P. (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** 333–382. MR2278331

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

[11] 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

[12] DOUC, R., FORT, G., MOULINES, E. and SOULIER, P. (2004). Practical drift conditions for subgeometric rates of convergence. *Ann. Appl. Probab.* **14** 1353–1377. MR2071426

[13] DOUC, R., MOULINES, E. and SOULIER, P. (2007). Computable convergence rates for subgeometric ergodic Markov chains. *Bernoulli* **13** 831–848. MR2348753

[14] GÅSEMYR, J. (2006). The spectrum of the independent Metropolis–Hastings algorithm. *J. Theoret. Probab.* **19** 152–165. MR2256484

[15] HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109.

[16] JARNER, S. F. and HANSEN, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic Process. Appl.* **85** 341–361. MR1731030

[17] JARNER, S. F. and ROBERTS, G. O. (2002). Polynomial convergence rates of Markov chains. *Ann. Appl. Probab.* **12** 224–247. MR1890063

[18] KARAGIANNIS, G. and ANDRIEU, C. (2013). Annealed importance sampling reversible jump MCMC algorithms. *J. Comput. Graph. Statist.* **22** 623–648. MR3173734

[19] ŁATUSZYŃSKI, K., MIASOJEDOW, B. and NIEMIRO, W. (2013). Nonasymptotic bounds on the estimation error of MCMC algorithms. *Bernoulli* **19** 2033–2066. MR3129043

[20] LEE, A., ANDRIEU, C. and DOUCET, A. (2012). Discussion of "Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation" by Fearnhead and Prangle. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **74** 449–450.

[21] LEE, A. and ŁATUSZYŃSKI, K. (2012). Variance bounding and geometric ergodicity of Markov chain Monte Carlo kernels for approximate Bayesian computation. Available at arXiv:1210.6703v1.

[22] LINDVALL, T. (2002). *Lectures on the Coupling Method*. Dover, Mineola, NY. MR1924231

[23] MENGERSEN, K. L. and TWEEDIE, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* **24** 101–121. MR1389882

[24] METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. and TELLER, E. (1953). Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21** 1087–1092.

[25] MEYER, P. A. (1966). *Probability and Potentials*. Blaisdell, Waltham, MA.

[26] MEYN, S. and TWEEDIE, R. L. (2009). *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2509253

[27] NICHOLLS, G. K., FOX, C. and WATT, A. M. (2012). Coupled MCMC with a randomized acceptance probability. Available at arXiv:1205.6857v1.

[28] NUMMELIN, E. (2002). MC's for MCMC'ists. *Int. Statist. Rev.* **70** 215–240.

[29] PESKUN, P. H. (1973). Optimum Monte-Carlo sampling using Markov chains. *Biometrika* **60** 607–612. MR0362823

[30] ROBERTS, G. O. and ROSENTHAL, J. S. (1997). Geometric ergodicity and hybrid Markov chains. *Electron. Commun. Probab.* **2** 13–25 (electronic). MR1448322

[31] ROBERTS, G. O. and ROSENTHAL, J. S. (2008). Variance bounding Markov chains. *Ann. Appl. Probab.* **18** 1201–1214. MR2418242

[32] ROBERTS, G. O. and TWEEDIE, R. L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83** 95–110. MR1399158

[33] SAKSMAN, E. and VIHOLA, M. (2010). On the ergodicity of the adaptive Metropolis algorithm on unbounded domains. *Ann. Appl. Probab.* **20** 2178–2203. MR2759732

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms

[34] TIERNEY, L. (1998). A note on Metropolis–Hastings kernels for general state spaces. *Ann. Appl. Probab.* **8** 1–9. MR1620401

SCHOOL OF MATHEMATICS
UNIVERSITY OF BRISTOL
UNIVERSITY WALK
BRISTOL, BS8 1TW
UNITED KINGDOM
E-MAIL: C.Andrieu@bristol.ac.uk

DEPARTMENT OF STATISTICS
UNIVERSITY OF OXFORD
1 SOUTH PARKS ROAD
OXFORD
OX1 3TG
UNITED KINGDOM
E-MAIL: matti.vihola@iki.fi

This content downloaded from 148.88.67.84 on Wed, 18 Sep 2019 13:29:37 UTC
All use subject to https://about.jstor.org/terms