Bernoulli **19**(5A), 2013, 2033–2066
DOI: 10.3150/12-BEJ442

Nonasymptotic bounds on the estimation error of MCMC algorithms

KRZYSZTOF ŁATUSZYŃSKI[1], BŁAŻEJ MIASOJEDOW[2] and WOJCIECH NIEMIRO[3]

[1]*Department of Statistics, University of Warwick, CV4 7AL, Coventry, UK. E-mail: latuch@gmail.com; url: http://www.warwick.ac.uk/go/klatuszynski*
[2]*Institute of Applied Mathematics and Mechanics, University of Warsaw, Banacha 2, 02-097 Warszawa, Poland. E-mail: bmia@mimuw.edu.pl*
[3]*Faculty of Mathematics and Computer Science, Nicolaus Copernicus University, Chopina 12/18, 87-100 Toruń, Poland. E-mail: wniem@mat.uni.torun.pl*

We address the problem of upper bounding the mean square error of MCMC estimators. Our analysis is nonasymptotic. We first establish a general result valid for essentially all ergodic Markov chains encountered in Bayesian computation and a possibly unbounded target function f. The bound is sharp in the sense that the leading term is exactly $\sigma^2_{\mathrm{as}}(P,f)/n$, where $\sigma^2_{\mathrm{as}}(P,f)$ is the CLT asymptotic variance. Next, we proceed to specific additional assumptions and give explicit computable bounds for geometrically and polynomially ergodic Markov chains under quantitative drift conditions. As a corollary, we provide results on confidence estimation.

Keywords: asymptotic variance; computable bounds; confidence estimation; drift conditions; geometric ergodicity; mean square error; polynomial ergodicity; regeneration

1. Introduction

Let π be a probability distribution on a Polish space $\mathcal{X}$ and $f:\mathcal{X}\to\mathbb{R}$ be a Borel function. The objective is to compute (estimate) the quantity

$$\theta := \pi(f) = \int_{\mathcal{X}} \pi(\mathrm{d}x) f(x).$$

Typically $\mathcal{X}$ is a high dimensional space, f need not be bounded and the density of π is known up to a normalizing constant. Such problems arise in Bayesian inference and are often solved using Markov chain Monte Carlo (MCMC) methods. The idea is to simulate a Markov chain (X_n) with transition kernel P such that $\pi P = \pi$, that is π is stationary with respect to P. Then averages along the trajectory of the chain,

$$\hat{\theta}_n := \frac{1}{n}\sum_{i=0}^{n-1} f(X_i)$$

are used to estimate θ. It is essential to have explicit and reliable bounds which provide information about how long the algorithms must be run to achieve a prescribed level of accuracy (cf.

[28,29,53]). The aim of our paper is to derive nonasymptotic and explicit bounds on the mean square error,

$$\text{MSE} := \mathbb{E}(\hat{\theta}_n - \theta)^2. \tag{1.1}$$

To upper bound (1.1), we begin with a general inequality valid for all ergodic Markov chains that admit a one step small set condition. Our bound is sharp in the sense that the leading term is exactly $\sigma^2_{\text{as}}(P, f)/n$, where $\sigma^2_{\text{as}}(P, f)$ is the asymptotic variance in the central limit theorem. The proof relies on the regeneration technique, methods of renewal theory and statistical sequential analysis.

To obtain explicit bounds, we subsequently consider geometrically and polynomially ergodic Markov chains. We assume appropriate drift conditions that give quantitative information about the transition kernel P. The upper bounds on MSE are then stated in terms of the drift parameters.

We note that most MCMC algorithms implemented in Bayesian inference are geometrically or polynomially ergodic (however establishing the quantitative drift conditions we utilize may be prohibitively difficult for complicated models). Uniform ergodicity is stronger then geometrical ergodicity considered here and is often discussed in literature. However, few MCMC algorithms used in practice are uniformly ergodic. MSE and confidence estimation for uniformly ergodic chains are discussed in our accompanying paper [34].

The Subgeometric condition, considered in, for example, [10], is more general than polynomial ergodicity considered here. We note that with some additional effort, the results for polynomially ergodic chains (Section 5) can be reformulated for subgeometric Markov chains. Motivated by applications, we avoid these technical difficulties.

Upper bounding the mean square error (1.1) leads immediately to confidence estimation by applying the Chebyshev inequality. One can also apply the more sophisticated median trick of [25], further developed in [45]. The median trick leads to an exponential inequality for the MCMC estimate whenever the MSE can be upper bounded, in particular in the setting of geometrically and polynomially ergodic chains.

We illustrate our results with benchmark examples. The first, which is related to a simplified hierarchical Bayesian model and similar to [29], Example 2, allows to compare the bounds provided in our paper with actual MCMC errors. Next, we demonstrate how to apply our results in the Poisson–Gamma model of [18]. Finally, the contracting normals toy-example allows for a numerical comparison with our earlier work [35].

The paper is organised as follows: in Section 2 we give background on the regeneration technique and introduce notation. The general MSE upper bound is derived in Section 3. Geometrically and polynomially ergodic Markov chains are considered in Sections 4 and 5, respectively. The applicability of our results is discussed in Section 6, where also numerical examples are presented. Technical proofs are deferred to Sections 7 and 8.

1.1. Related nonasymptotic results

A vast literature on nonasymptotic analysis of Markov chains is available in various settings. To place our results in this context, we give a brief account.

In the case of *finite state space*, an approach based on the spectral decomposition was used in [2,19,36,45] to derive results of related type.

For *bounded* functionals of *uniformly* ergodic chains on a general state space, exponential inequalities with explicit constants such as those in [20,33] can be applied to derive confidence bounds. In the accompanying paper [34], we compare the simulation cost of confidence estimation based on our approach (MSE bounds with the median trick) to exponential inequalities and conclude that while exponential inequalities have sharper constants, our approach gives in this setting the optimal dependence on the regeneration rate β and therefore will turn out more efficient in many practical examples.

Related results come also from studying concentration of measure phenomenon for dependent random variables. For the large body of work in this area see, for example, [40,58] and [32] (and references therein), where transportation inequalities or martingale approach have been used. These results, motivated in a more general setting, are valid for Lipschitz functions with respect to the Hamming metric. They also include expressions $\sup_{x,y\in\mathcal{X}} \|P^i(x,\cdot) - P^i(y,\cdot)\|_{\mathrm{tv}}$ and when applied to our setting, they are well suited for *bounded* functionals of *uniformly* ergodic Markov chains, but cannot be applied to geometrically ergodic chains. For details, we refer to the original papers and the discussion in Section 3.5 of [1].

For lazy reversible Markov chains, nonasymptotic mean square error bounds have been obtained for *bounded* target functions in [57] in a setting where explicit bounds on conductance are available. These results have been applied to approximating integrals over balls in $\mathbb{R}^d$ under some regularity conditions for the stationary measure, see [57] for details. The Markov chains considered there are in fact uniformly ergodic, however in their setting the regeneration rate β, can be verified for P^h, $h > 1$ rather then for P and turns out to be exponentially small in dimension. Hence, conductance seems to be the natural approach to make the problem tractable in high dimensions.

Tail inequalities for *bounded* functionals of Markov chains that are not uniformly ergodic were considered in [1,7] and [10] using regeneration techniques. These results apply for example, to *geometrically* or *subgeometrically* ergodic Markov chains, however they also involve nonexplicit constants or require tractability of moment conditions of random tours between regenerations. Computing explicit bounds from these results may be possible with additional work, but we do not pursue it here.

Nonasymptotic analysis of *unbounded* functionals of Markov chains is scarce. In particular, tail inequalities for *unbounded* target function f that can be applied to *geometrically* ergodic Markov chains have been established by Bertail and Clémençon in [6] by regenerative approach and using truncation arguments. However, they involve nonexplicit constants and can not be directly applied to confidence estimation. Nonasymptotic and explicit MSE bounds for geometrically ergodic MCMC samplers have been obtained in [35] under a geometric drift condition by exploiting computable convergence rates. Our present paper improves these results in a fundamental way. Firstly, the generic Theorem 3.1 allows to extend the approach to different classes of Markov chains, for example, polynomially ergodic in Section 5. Secondly, rather then resting on computable convergence rates, the present approach relies on upper-bounding the CLT asymptotic variance which, somewhat surprisingly, appears to be more accurate and consequently the MSE bound is much sharper, as demonstrated by numerical examples in Section 6.

Recent work [31] address error estimates for MCMC algorithms under positive curvature condition. The positive curvature implies geometric ergodicity in the Wasserstein distance and bivariate drift conditions (cf. [49]). Their approach appears to be applicable in different settings to ours and also rests on different notions, for example, employs the coarse diffusion constant instead of the exact asymptotic variance. Moreover, the target function f is assumed to be Lipschitz which is problematic in Bayesian inference. Therefore, our results and [31] appear to be complementary.

Nonasymptotic rates of convergence of geometrically, polynomially and subgeometrically ergodic Markov chains to their stationary distributions have been investigated in many papers [4, 11–13,16,30,43,51,52,54,55] under assumptions similar to our Section 4 and 5, together with an aperiodicity condition that is not needed for our purposes. Such results, although of utmost theoretical importance, do not directly translate into bounds on accuracy of estimation, as they allow to control only the bias of estimates and the so-called burn-in time.

2. Regeneration construction and notation

Assume P has invariant distribution π on $\mathcal{X}$, is π-irreducible and Harris recurrent. The following one step small set Assumption 2.1 is verifiable for virtually all Markov chains targeting Bayesian posterior distributions. It allows for the regeneration/split construction of Nummelin [46] and Athreya and Ney [3].

Assumption 2.1 (Small set). *There exist a Borel set $J \subseteq \mathcal{X}$ of positive π measure, a number $\beta > 0$ and a probability measure ν such that*

$$P(x, \cdot) \geq \beta \mathbb{I}(x \in J)\nu(\cdot).$$

Under Assumption 2.1, we can define a bivariate Markov chain (X_n, Γ_n) on the space $\mathcal{X} \times \{0, 1\}$ in the following way. Bell variable Γ_{n-1} depends only on X_{n-1} via

$$\mathbb{P}(\Gamma_{n-1} = 1 | X_{n-1} = x) = \beta \mathbb{I}(x \in J). \tag{2.1}$$

The rule of transition from (X_{n-1}, Γ_{n-1}) to X_n is given by

$$\mathbb{P}(X_n \in A | \Gamma_{n-1} = 1, X_{n-1} = x) = \nu(A),$$
$$\mathbb{P}(X_n \in A | \Gamma_{n-1} = 0, X_{n-1} = x) = Q(x, A),$$

where Q is the normalized "residual" kernel given by

$$Q(x, \cdot) := \frac{P(x, \cdot) - \beta \mathbb{I}(x \in J)\nu(\cdot)}{1 - \beta \mathbb{I}(x \in J)}.$$

Whenever $\Gamma_{n-1} = 1$, the chain regenerates at moment n. The regeneration epochs are

$$T := T_1 := \min\{n \geq 1 : \Gamma_{n-1} = 1\},$$
$$T_k := \min\{n \geq T_{k-1} : \Gamma_{n-1} = 1\}.$$

Write $\tau_k := T_k - T_{k-1}$ for $k = 2, 3, \dots$ and $\tau_1 := T$. Random blocks

$$\Xi := \Xi_1 := (X_0, \dots, X_{T-1}, T),$$
$$\Xi_k := (X_{T_{k-1}}, \dots, X_{T_k - 1}, \tau_k)$$

for $k = 1, 2, 3, \dots$ are independent.

We note that numbering of the bell variables Γ_n may differ between authors: in our notation $\Gamma_{n-1} = 1$ indicates regeneration at moment n, not $n - 1$. Let symbols $\mathbb{P}_\xi$ and $\mathbb{E}_\xi$ mean that $X_0 \sim \xi$. Note also that these symbols are unambiguous, because specifying the distribution of X_0 is equivalent to specifying the joint distribution of (X_0, Γ_0) via (2.1).

For $k = 2, 3, \dots$, every block Ξ_k under $\mathbb{P}_\xi$ has the same distribution as Ξ under $\mathbb{P}_\nu$. However, the distribution of Ξ under $\mathbb{P}_\xi$ is in general different. We will also use the following notations for the block sums:

$$\Xi(f) := \sum_{i=0}^{T-1} f(X_i), \qquad \Xi_k(f) := \sum_{i=T_{k-1}}^{T_k - 1} f(X_i).$$

3. A general inequality for the MSE

We assume that $X_0 \sim \xi$ and thus $X_n \sim \xi P^n$. Write $\bar{f} := f - \pi(f)$.

Theorem 3.1. *If Assumption 2.1 holds, then*

$$\sqrt{\mathbb{E}_\xi(\hat{\theta}_n - \theta)^2} \le \frac{\sigma_{\mathrm{as}}(P, f)}{\sqrt{n}}\left(1 + \frac{C_0(P)}{n}\right) + \frac{C_1(P, f)}{n} + \frac{C_2(P, f)}{n}, \tag{3.1}$$

where

$$\sigma_{\mathrm{as}}^2(P, f) := \frac{\mathbb{E}_\nu(\Xi(\bar{f}))^2}{\mathbb{E}_\nu T}, \tag{3.2}$$

$$C_0(P) := \mathbb{E}_\pi T - \frac{1}{2}, \tag{3.3}$$

$$C_1(P, f) := \sqrt{\mathbb{E}_\xi\big(\Xi(|\bar{f}|)\big)^2}, \tag{3.4}$$

$$C_2(P, f) = C_2(P, f, n) := \sqrt{\mathbb{E}_\xi\left(\mathbb{I}(T_1 < n) \sum_{i=n}^{T_{R(n)}-1} |\bar{f}|(X_i)\right)^2}, \tag{3.5}$$

$$R(n) := \min\{r \ge 1 \colon T_r > n\}. \tag{3.6}$$

Remark 3.2. The bound in Theorem 3.1 is meaningful only if $\sigma_{\mathrm{as}}^2(P, f) < \infty$, $C_0(P) < \infty$, $C_1(P, f) < \infty$ and $C_2(P, f) < \infty$. Under Assumption 2.1, we always have $\mathbb{E}_\nu T < \infty$ but not necessarily $\mathbb{E}_\nu T^2 < \infty$. On the other hand, finiteness of $\mathbb{E}_\nu(\Xi(\bar{f}))^2$ is a sufficient and necessary condition for the CLT to hold for Markov chain X_n and function f. This fact is proved in [5] in a more general setting. For our purposes, it is important to note that $\sigma^2_{\mathrm{as}}(P, f)$ in Theorem 3.1 is indeed the *asymptotic variance* which appears in the CLT, that is

$$\sqrt{n}(\hat{\theta}_n - \theta) \to_d \mathrm{N}\big(0, \sigma^2_{\mathrm{as}}(P, f)\big).$$

Moreover,

$$\lim_{n\to\infty} n\mathbb{E}_\xi(\hat{\theta}_n - \theta)^2 = \sigma^2_{\mathrm{as}}(P, f).$$

In this sense, the leading term $\sigma_{\mathrm{as}}(P, f)/\sqrt{n}$ in Theorem 3.1 is "asymptotically correct" and cannot be improved.

Remark 3.3. Under additional assumptions of geometric and polynomial ergodicity, in Sections 4 and 5 respectively, we will derive bounds for $\sigma^2_{\mathrm{as}}(P, f)$ and $C_0(P)$, $C_1(P, f)$, $C_2(P, f)$ in terms of some explicitly computable quantities.

Remark 3.4. In our related work [34], we discuss a special case of the setting considered here, namely when regeneration times T_k are identifiable. These leads to $X_0 \sim \nu$ and an regenerative estimator of the form

$$\hat{\theta}_{T_{R(n)}} := \frac{1}{T_{R(n)}} \sum_{i=1}^{R(n)} \Xi_i(f) = \frac{1}{T_{R(n)}} \sum_{i=0}^{T_{R(n)}-1} f(X_i). \tag{3.7}$$

The estimator $\hat{\theta}_{T_{R(n)}}$ is somewhat easier to analyze. We refer to [34] for details.

Proof of Theorem 3.1. Recall $R(n)$ defined in (3.6) and let

$$\Delta(n) := T_{R(n)} - n.$$

In words: $R(n)$ is *the first moment of regeneration past* n and $\Delta(n)$ is the *overshoot* or excess over n. Let us express the estimation error as follows.

$$\begin{aligned}\hat{\theta}_n - \theta &= \frac{1}{n}\sum_{i=0}^{n-1} \bar{f}(X_i) = \frac{1}{n}\left(\sum_{i=T_1}^{T_{R(n)}-1} \bar{f}(X_i) + \sum_{i=0}^{T_1-1} \bar{f}(X_i) - \sum_{i=n}^{T_{R(n)}-1} \bar{f}(X_i)\right) \\ &=: \frac{1}{n}(\mathcal{Z} + \mathcal{O}_1 - \mathcal{O}_2),\end{aligned}$$

with the convention that $\sum_l^u = 0$ whenever $l > u$. The triangle inequality entails

$$\sqrt{\mathbb{E}_\xi(\hat{\theta}_n - \theta)^2} \le \frac{1}{n}\left(\sqrt{\mathbb{E}_\xi \mathcal{Z}^2} + \sqrt{\mathbb{E}_\xi(\mathcal{O}_1 - \mathcal{O}_2)^2}\right). \tag{3.8}$$

Denote $C(P,f) := \sqrt{\mathbb{E}_\xi(\mathcal{O}_1 - \mathcal{O}_2)^2}$ and compute

$$\begin{aligned}
C(P,f) &= \Bigg(\mathbb{E}_\xi\Bigg\{\Bigg(\sum_{i=0}^{T-1}\bar f(X_i) - \sum_{i=n}^{T_{R(n)}-1}\bar f(X_i)\Bigg)\mathbb{I}(T \ge n) \\
&\qquad + \Bigg(\sum_{i=0}^{T-1}\bar f(X_i) - \sum_{i=n}^{T_{R(n)}-1}\bar f(X_i)\Bigg)\mathbb{I}(T < n)\Bigg\}^2\Bigg)^{1/2} \\
&\le \Bigg(\mathbb{E}_\xi\Bigg(\sum_{i=0}^{T-1}\big|\bar f(X_i)\big| + \sum_{i=n}^{T_{R(n)}-1}\big|\bar f(X_i)\big|\mathbb{I}(T<n)\Bigg)^2\Bigg)^{1/2} \qquad (3.9)\\
&\le \sqrt{\mathbb{E}_\xi\Bigg(\sum_{i=0}^{T-1}\big|\bar f(X_i)\big|\Bigg)^2} + \sqrt{\mathbb{E}_\xi\Bigg(\sum_{i=n}^{T_{R(n)}-1}\big|\bar f(X_i)\big|\mathbb{I}(T<n)\Bigg)^2} \\
&= C_1(P,f) + C_2(P,f).
\end{aligned}$$

It remains to bound the middle term, $\mathbb{E}_\xi \mathcal{Z}^2$, which clearly corresponds to the most significant portion of the estimation error. The crucial step in our proof is to show the following inequality:

$$\mathbb{E}_\nu\Bigg(\sum_{i=0}^{T_{R(n)}-1}\bar f(X_i)\Bigg)^2 \le \sigma^2_{\mathrm{as}}(P,f)\big(n + 2C_0(P)\big). \qquad (3.10)$$

Once this is proved, it is easy to see that

$$\begin{aligned}
\mathbb{E}_\xi \mathcal{Z}^2 &= \sum_{j=1}^{n}\mathbb{E}_\xi\big(\mathcal{Z}^2|T_1=j\big)\mathbb{P}_\xi(T_1=j) = \sum_{j=1}^{n}\mathbb{E}_\nu\Bigg(\sum_{i=0}^{T_{R(n-j)}-1}\bar f(X_i)\Bigg)^2\mathbb{P}_\xi(T_1=j) \\
&\le \sum_{j=1}^{n}\sigma^2_{\mathrm{as}}(P,f)\big(n-j+2C_0(P)\big)\mathbb{P}_\xi(T_1=j) \le \sigma^2_{\mathrm{as}}(P,f)\big(n+2C_0(P)\big),
\end{aligned}$$

consequently, $\sqrt{\mathbb{E}_\xi \mathcal{Z}^2} \le \sqrt{n}\sigma_{\mathrm{as}}(P,f)(1 + C_0(P)/n)$ and the conclusion will follow by recalling (3.8) and (3.9).

We are therefore left with the task of proving (3.10). This is essentially a statement about sums of i.i.d. random variables. Indeed,

$$\sum_{i=0}^{T_{R(n)}-1}\bar f(X_i) = \sum_{k=1}^{R(n)}\Xi_k(\bar f) \qquad (3.11)$$

and all the blocks Ξ_k (including $\Xi = \Xi_1$) are i.i.d. under $\mathbb{P}_\nu$. By the general version of the Kac theorem ([42], Theorem 10.0.1, or [47], equation (3.3.7)), we have

$$\mathbb{E}_\nu \Xi(f) = \pi(f)\mathbb{E}_\nu T$$

(and $1/\mathbb{E}_\nu T = \beta\pi(J)$), so $\mathbb{E}_\nu \Xi(\bar{f}) = 0$ and $\mathrm{Var}_\nu\, \Xi(\bar{f}) = \sigma^2_{\mathrm{as}}(P, f)\mathbb{E}_\nu T$. Now we will exploit the fact that $R(n)$ is a stopping time with respect to $\mathcal{G}_k = \sigma((\Xi_1(\bar{f}), \tau_1), \dots, (\Xi_k(\bar{f}), \tau_k))$, a filtration generated by i.i.d. pairs. We are in a position to apply the two Wald's identities. The second identity yields

$$\mathbb{E}_\nu \left(\sum_{k=1}^{R(n)} \Xi_k(\bar{f}) \right)^2 = \mathrm{Var}_\nu\, \Xi(\bar{f})\mathbb{E}_\nu R(n) = \sigma^2_{\mathrm{as}}(P, f)\mathbb{E}_\nu T \mathbb{E}_\nu R(n).$$

But in this expression, we can replace $\mathbb{E}_\nu T \mathbb{E}_\nu R(n)$ by $\mathbb{E}_\nu T_{R(n)}$ because of the first Wald's identity:

$$\mathbb{E}_\nu T_{R(n)} = \mathbb{E}_\nu \sum_{k=1}^{R(n)} \tau_k = \mathbb{E}_\nu T \mathbb{E}_\nu R(n).$$

It follows that

$$\mathbb{E}_\nu \left(\sum_{k=1}^{R(n)} \Xi_k(\bar{f}) \right)^2 = \sigma^2_{\mathrm{as}}(P, f)\mathbb{E}_\nu T_{R(n)} = \sigma^2_{\mathrm{as}}(P, f)\bigl(n + \mathbb{E}_\nu \Delta(n)\bigr). \tag{3.12}$$

We now focus attention on bounding the "mean overshoot" $\mathbb{E}_\nu \Delta(n)$. Under $\mathbb{P}_\nu$, the cumulative sums $T = T_1 < T_2 < \cdots < T_k < \cdots$ form a (nondelayed) renewal process in discrete time. Let us invoke the following elegant theorem of Lorden ([37], Theorem 1):

$$\mathbb{E}_\nu \Delta(n) \le \frac{\mathbb{E}_\nu T^2}{\mathbb{E}_\nu T}. \tag{3.13}$$

By Lemma 7.1 with $g \equiv 1$ from Section 7, we obtain:

$$\mathbb{E}_\nu \Delta(n) \le 2\mathbb{E}_\pi T - 1. \tag{3.14}$$

Hence, substituting (3.14) into (3.12) and taking into account (3.11) we obtain (3.10) and complete the proof. □

4. Geometrically ergodic chains

In this section, we upper bound constants $\sigma^2_{\mathrm{as}}(P, f), C_0(P), C_1(P, f), C_2(P, f)$, appearing in Theorem 3.1, for geometrically ergodic Markov chains under a quantitative drift assumption. Proofs are deferred to Sections 7 and 8.

Using drift conditions is a standard approach for establishing geometric ergodicity. We refer to [50] or [42] for definitions and further details. The assumption below is the same as in [4]. Specifically, let J be the small set which appears in Assumption 2.1.

Assumption 4.1 (Geometric drift). *There exist a function $V : \mathcal{X} \to [1, \infty[$, constants $\lambda < 1$ and $K < \infty$ such that*

$$PV(x) := \int_{\mathcal{X}} P(x, \mathrm{d}y)V(y) \le \begin{cases} \lambda V(x), & \text{for } x \notin J, \\ K, & \text{for } x \in J. \end{cases}$$

In many papers conditions similar to Assumption 4.1 have been established for realistic MCMC algorithms in statistical models of practical relevance [14,17,21,27,30,56]. This opens the possibility of computing nonasymptotic upper bounds on MSE or nonasymptotic confidence intervals in these models.

In this section, we bound quantities appearing in Theorem 3.1 by expressions involving λ, β and K. The main result in this section is the following theorem.

Theorem 4.2. *If Assumptions 2.1 and 4.1 hold and f is such that*

$$\|\bar{f}\|_{V^{1/2}} := \sup_x |\bar{f}(x)| / V^{1/2}(x) < \infty,$$

then

(i) $$C_0(P) \le \frac{\lambda}{1-\lambda}\pi(V) + \frac{K - \lambda - \beta}{\beta(1-\lambda)} + \frac{1}{2},$$

(ii) $$\frac{\sigma^2_{\mathrm{as}}(P, f)}{\|\bar{f}\|^2_{V^{1/2}}} \le \frac{1 + \lambda^{1/2}}{1 - \lambda^{1/2}}\pi(V) + \frac{2(K^{1/2} - \lambda^{1/2} - \beta)}{\beta(1 - \lambda^{1/2})}\pi\bigl(V^{1/2}\bigr),$$

(iii) $$\frac{C_1(P, f)^2}{\|\bar{f}\|^2_{V^{1/2}}} \le \frac{1}{(1 - \lambda^{1/2})^2}\xi(V) + \frac{2(K^{1/2} - \lambda^{1/2} - \beta)}{\beta(1 - \lambda^{1/2})^2}\xi\bigl(V^{1/2}\bigr) + \frac{\beta(K - \lambda - \beta) + 2(K^{1/2} - \lambda^{1/2} - \beta)^2}{\beta^2(1 - \lambda^{1/2})^2},$$

(iv) $C_2(P, f)^2$ *satisfies an inequality analogous to* (iii) *with ξ replaced by ξP^n.*

Remark 4.3. Combining Theorem 4.2 with Theorem 3.1 yields the MSE bound of interest. Note that the leading term is of order $n^{-1}\beta^{-1}(1-\lambda)^{-1}$. A related result is Proposition 2 of [15] where the pth moment of $\hat{\theta}_n$ for $p \ge 2$ is controlled under similar assumptions. Specialised to $p = 2$, the leading term of the moment bound of [15] is of order $n^{-1}\beta^{-3}(1-\lambda)^{-4}$.

Remark 4.4. An alternative form of the first bound in Theorem 4.2 is

(i′) $$C_0(P) \le \frac{\lambda^{1/2}}{1 - \lambda^{1/2}}\pi\bigl(V^{1/2}\bigr) + \frac{K^{1/2} - \lambda^{1/2} - \beta}{\beta(1 - \lambda^{1/2})} + \frac{1}{2}.$$

Theorem 4.2 still involves some quantities which can be difficult to compute, such as $\pi(V^{1/2})$ and $\pi(V)$, not to mention $\xi P^n(V^{1/2})$ and $\xi P^n(V)$. The following proposition gives some simple complementary bounds.

Proposition 4.5. *Under Assumptions 2.1 and 4.1,*

(i) $\pi\left(V^{1/2}\right) \le \pi(J)\dfrac{K^{1/2}-\lambda^{1/2}}{1-\lambda^{1/2}} \le \dfrac{K^{1/2}-\lambda^{1/2}}{1-\lambda^{1/2}}$,

(ii) $\pi(V) \le \pi(J)\dfrac{K-\lambda}{1-\lambda} \le \dfrac{K-\lambda}{1-\lambda}$,

(iii) *if* $\xi\left(V^{1/2}\right) \le \dfrac{K^{1/2}}{1-\lambda^{1/2}}$ *then* $\xi P^n\left(V^{1/2}\right) \le \dfrac{K^{1/2}}{1-\lambda^{1/2}}$,

(iv) *if* $\xi(V) \le \dfrac{K}{1-\lambda}$ *then* $\xi P^n(V) \le \dfrac{K}{1-\lambda}$,

(v) $\|\bar{f}\|_{V^{1/2}}$ *can be related to* $\|f\|_{V^{1/2}}$ *by*

$$\|\bar{f}\|_{V^{1/2}} \le \|f\|_{V^{1/2}}\left[1+\frac{\pi(J)(K^{1/2}-\lambda^{1/2})}{(1-\lambda^{1/2})\inf_{x\in\mathcal{X}} V^{1/2}(x)}\right] \le \|f\|_{V^{1/2}}\left[1+\frac{K^{1/2}-\lambda^{1/2}}{1-\lambda^{1/2}}\right].$$

Remark 4.6. In MCMC practice, almost always the initial state is deterministically chosen, $\xi=\delta_x$ for some $x\in\mathcal{X}$. In this case in (ii) and (iii), we just have to choose x such that $V^{1/2}(x)\le K^{1/2}/(1-\lambda^{1/2})$ and $V(x)\le K/(1-\lambda)$, respectively (note that the latter inequality implies the former). It might be interesting to note that our bounds would not be improved if we added a burn-in time $t>0$ at the beginning of simulation. The standard practice in MCMC computations is to discard the initial part of trajectory and use the estimator

$$\hat{\theta}_{t,n} := \frac{1}{n}\sum_{i=t}^{n+t-1} f(X_i).$$

Heuristic justification is that the closer ξP^t is to the equilibrium distribution π, the better. However, for technical reasons, our upper bounds on error are the tightest if the initial point has the smallest value of V, and *not* if its distribution is close to π.

Remark 4.7. In many specific examples, one can obtain (with some additional effort) sharper inequalities than those in Proposition 4.5 or at least bound $\pi(J)$ away from 1. However, in general we assume that such bounds are not available.

5. Polynomially ergodic Markov chains

In this section, we upper bound constants $\sigma^2_{\mathrm{as}}(P,f), C_0(P), C_1(P,f), C_2(P,f)$, appearing in Theorem 3.1, for polynomially ergodic Markov chains under a quantitative drift assumption. Proofs are deferred to Sections 7 and 8.

The following drift condition is a counterpart of Drift in Assumption 4.1, and is used to establish polynomial ergodicity of Markov chains [9,10,22,42].

Assumption 5.1 (Polynomial drift). *There exist a function* $V : \mathcal{X} \to [1, \infty[$, *constants* $\lambda < 1$, $\alpha \le 1$ *and* $K < \infty$ *such that*

$$PV(x) \le \begin{cases} V(x) - (1-\lambda)V(x)^{\alpha}, & \text{for } x \notin J, \\ K, & \text{for } x \in J. \end{cases}$$

We note that Assumption 5.1 or closely related drift conditions have been established for MCMC samplers in specific models used in Bayesian inference, including independence samplers, random-walk Metropolis algorithms, Langevin algorithms and Gibbs samplers, see, for example, [14,23,24].

In this section, we bound quantities appearing in Theorem 3.1 by expressions involving λ, β, α and K. The main result in this section is the following theorem.

Theorem 5.2. *If Assumptions 2.1 and 5.1 hold with* $\alpha > \frac{2}{3}$ *and* f *is such that* $\|\bar{f}\|_{V^{(3/2)\alpha-1}} := \sup_x |\bar{f}(x)|/V^{(3/2)\alpha-1}(x) < \infty$, *then*

(i) $$C_0(P) \le \frac{1}{\alpha(1-\lambda)}\pi\bigl(V^{\alpha}\bigr) + \frac{K^{\alpha}-1-\beta}{\beta\alpha(1-\lambda)} + \frac{1}{\beta} - \frac{1}{2},$$

(ii) $$\frac{\sigma^2_{\mathrm{as}}(P,f)}{\|\bar{f}\|^2_{V^{(3/2)\alpha-1}}} \le \pi\bigl(V^{3\alpha-2}\bigr) + \frac{4\pi(V^{2\alpha-1})}{\alpha(1-\lambda)} + 2\biggl(\frac{2K^{\alpha/2}-2-2\beta}{\alpha\beta(1-\lambda)} + \frac{1}{\beta} - 1\biggr)\pi\bigl(V^{(3/2)\alpha-1}\bigr),$$

(iii) $$\begin{aligned} \frac{C_1(P,f)^2}{\|\bar{f}\|^2_{V^{(3/2)\alpha-1}}} \le{}& \frac{1}{(2\alpha-1)(1-\lambda)}\xi\bigl(V^{2\alpha-1}\bigr) + \frac{4}{\alpha^2(1-\lambda)^2}\xi\bigl(V^{\alpha}\bigr) \\ &+ \biggl(\frac{8K^{\alpha/2}-8-8\beta}{\alpha^2\beta(1-\lambda)^2} + \frac{4-4\beta}{\alpha\beta(1-\lambda)}\biggr)\xi\bigl(V^{\alpha/2}\bigr) \\ &+ \frac{\alpha(1-\lambda)+4}{\alpha\beta(1-\lambda)} + \frac{K^{2\alpha-1}-1-\beta}{(2\alpha-1)\beta(1-\lambda)} \\ &+ \frac{4(K^{\alpha}-1-\beta)}{\alpha^2\beta(1-\lambda)^2} + 2\biggl(\frac{2K^{\alpha/2}-2-2\beta}{\alpha\beta(1-\lambda)} + \frac{1}{\beta}\biggr)^2 \\ &- 2\biggl(\frac{2K^{\alpha/2}-2-2\beta}{\alpha\beta(1-\lambda)} + \frac{1}{\beta}\biggr), \end{aligned}$$

(iv) $$\begin{aligned} \frac{C_2(P,f)^2}{\|\bar{f}\|^2_{V^{(3/2)\alpha-1}}} \le{}& \frac{1}{(2\alpha-1)\beta^{(2\alpha-1)/\alpha}(1-\lambda)}\biggl(\frac{K-\lambda}{1-\lambda}\biggr)^{(4\alpha-2)/\alpha} + \frac{4(K-\lambda)^2}{\alpha^2\beta(1-\lambda)^4} \\ &+ \biggl(\frac{8K^{\alpha/2}-8-8\beta}{\alpha^2\beta(1-\lambda)^2} + \frac{4-4\beta}{\alpha\beta(1-\lambda)}\biggr)\frac{K-\lambda}{\sqrt{\beta}(1-\lambda)} \\ &+ \frac{\alpha(1-\lambda)+4}{\alpha\beta(1-\lambda)} + \frac{K^{2\alpha-1}-1-\beta}{(2\alpha-1)\beta(1-\lambda)} \end{aligned}$$

$$+\frac{4(K^{\alpha}-1-\beta)}{\alpha^2\beta(1-\lambda)^2}+2\left(\frac{2K^{\alpha/2}-2-2\beta}{\alpha\beta(1-\lambda)}+\frac{1}{\beta}\right)^2$$
$$-2\left(\frac{2K^{\alpha/2}-2-2\beta}{\alpha\beta(1-\lambda)}+\frac{1}{\beta}\right).$$

Remark 5.3. A counterpart of Theorem 5.2 parts (i)–(iii) for $\frac{1}{2}<\alpha\le\frac{2}{3}$ and functions s.t. $\|f\|_{V^{\alpha-1/2}}<\infty$ can be also established, using respectively modified but analogous calculations as in the proof of the above. For part (iv) however, an additional assumption $\pi(V)<\infty$ is necessary.

Theorem 5.2 still involves some quantities depending on π which can be difficult to compute, such as $\pi(V^{\eta})$ for $\eta\le\alpha$. The following proposition gives some simple complementary bounds.

Proposition 5.4. *Under Assumptions 2.1 and 5.1,*

(i) *For $\eta\le\alpha$ we have*

$$\pi\left(V^{\eta}\right)\le\left(\frac{K-\lambda}{1-\lambda}\right)^{\eta/\alpha}.$$

(ii) *If $\eta\le\alpha$, then $\|\bar{f}\|_{V^{\eta}}$ can be related to $\|f\|_{V^{\eta}}$ by*

$$\|\bar{f}\|_{V^{\eta}}\le\|f\|_{V^{\eta}}\left[1+\left(\frac{K-\lambda}{1-\lambda}\right)^{\eta/\alpha}\right].$$

6. Applicability in Bayesian inference and examples

To apply current results for computing MSE of estimates arising in Bayesian inference, one needs drift and small set conditions with explicit constants. The quality of these constants will affect the tightness of the overall MSE bound. In this section, we present three numerical examples. In Section 6.1, a simplified hierarchical model similar as [29], Example 2, is designed to compare the bounds with actual values and asses their quality. Next, in Section 6.2, we upperbound the MSE in the extensively discussed in literature Poisson–Gamma hierarchical model. Finally, in Section 6.3, we present the contracting normals toy-example to demonstrate numerical improvements over [35].

In realistic statistical models, the explicit drift conditions required for our analysis are very difficult to establish. Nevertheless, they have been recently obtained for a wide range of complex models of practical interest. Particular examples include: Gibbs sampling for hierarchical random effects models in [30]; van Dyk and Meng's algorithm for multivariate Student's t model [39]; Gibbs sampling for a family of Bayesian hierarchical general linear models in [26] (cf. also [27]); block Gibbs sampling for Bayesian random effects models with improper priors [59]; Data Augmentation algorithm for Bayesian multivariate regression models with Student's t regression errors [56]. Moreover, a large body of related work has been devoted to establishing a drift condition together with a small set to enable regenerative simulation for classes of statistical models. This kind of results, pursued in a number of papers mainly by James P. Hobert, Galin L. Jones and their coauthors, cannot be used directly for our purposes, but may provide substantial help in establishing quantitative drift and regeneration required here.

In settings where existence of drift conditions can be established, but explicit constants can not be computed (cf., e.g., [17,48]), our results do not apply and one must validate MCMC by asymptotic arguments. This is not surprising since qualitative existence results are not well suited for deriving quantitative finite sample conclusions.

6.1. A simplified hierarchical model

The simulation experiments described below are designed to compare the bounds proved in this paper with actual errors of MCMC estimation. We use a simple example similar as [29], Example 2. Assume that $y = (y_1, \dots, y_t)$ is an i.i.d. sample from the normal distribution $\mathrm{N}(\mu, \kappa^{-1})$, where κ denotes the reciprocal of the variance. Thus, we have

$$p(y|\mu,\kappa) = p(y_1,\dots,y_t|\mu,\kappa) \propto \kappa^{t/2} \exp\left[-\frac{\kappa}{2}\sum_{j=1}^{t}(y_j-\mu)^2\right].$$

The pair (μ, κ) plays the role of an unknown parameter. To make things simple, let us use the Jeffrey's noninformative (improper) prior $p(\mu,\kappa) = p(\mu)p(\kappa) \propto \kappa^{-1}$ (in [29] a different prior is considered). The posterior density is

$$p(\mu,\kappa|y) \propto p(y|\mu,\kappa)p(\mu,\kappa) \propto \kappa^{t/2-1} \exp\left[-\frac{\kappa t}{2}\bigl(s^2 + (\bar{y}-\mu)^2\bigr)\right],$$

where

$$\bar{y} = \frac{1}{t}\sum_{j=1}^{t} y_j, \qquad s^2 = \frac{1}{t}\sum_{j=1}^{t}(y_j - \bar{y})^2.$$

Note that $\bar{y}$ and s^2 only determine the location and scale of the posterior. We will be using a Gibbs sampler, whose performance does not depend on scale and location, therefore without loss of generality we can assume that $\bar{y} = 0$ and $s^2 = t$. Since $y = (y_1, \dots, y_t)$ is kept fixed, let us slightly abuse notation by using symbols $p(\kappa|\mu)$, $p(\mu|\kappa)$ and $p(\mu)$ for $p(\kappa|\mu, y)$, $p(\mu|\kappa, y)$ and $p(\mu|y)$, respectively. The Gibbs sampler alternates between drawing samples from both conditionals. Start with some (μ_0, κ_0). Then, for $i = 1, 2, \dots,$

- $\kappa_i \sim \mathrm{Gamma}(t/2, (t/2)(s^2 + \mu_{i-1}^2))$,
- $\mu_i \sim \mathrm{N}(0, 1/(\kappa_i t))$.

If we are chiefly interested in μ, then it is convenient to consider the two small steps $\mu_{i-1} \to \kappa_i \to \mu_i$ together. The transition density is

$$\begin{aligned} p(\mu_i|\mu_{i-1}) &= \int p(\mu_i|\kappa)p(\kappa|\mu_{i-1})\,\mathrm{d}\kappa \\ &\propto \int_0^\infty \kappa^{1/2} \exp\left[-\frac{\kappa t}{2}\mu_i^2\right]\bigl(s^2+\mu_{i-1}^2\bigr)^{t/2}\kappa^{t/2-1}\exp\left[-\frac{\kappa t}{2}\bigl(s^2+\mu_{i-1}^2\bigr)\right]\mathrm{d}\kappa \end{aligned}$$

$$= \left(s^2 + \mu_{i-1}^2\right)^{t/2} \int_0^\infty \kappa^{(t-1)/2} \exp\left[-\frac{\kappa t}{2}\left(s^2 + \mu_{i-1}^2 + \mu_i^2\right)\right] \mathrm{d}\kappa$$
$$\propto \left(s^2 + \mu_{i-1}^2\right)^{t/2}\left(s^2 + \mu_{i-1}^2 + \mu_i^2\right)^{-(t+1)/2}.$$

The proportionality constants concealed behind the $\propto$ sign depend only on t. Finally, we fix scale letting $s^2 = t$ and get

$$p(\mu_i|\mu_{i-1}) \propto \left(1 + \frac{\mu_{i-1}^2}{t}\right)^{t/2}\left(1 + \frac{\mu_{i-1}^2}{t} + \frac{\mu_i^2}{t}\right)^{-(t+1)/2}. \tag{6.1}$$

If we consider the RHS of (6.1) as a function of μ_i only, we can regard the first factor as constant and write

$$p(\mu_i|\mu_{i-1}) \propto \left(1 + \left(1 + \frac{\mu_{i-1}^2}{t}\right)^{-1} \frac{\mu_i^2}{t}\right)^{-(t+1)/2}.$$

It is clear that the conditional distribution of random variable

$$\mu_i \left(1 + \frac{\mu_{i-1}^2}{t}\right)^{-1/2} \tag{6.2}$$

is t-Student distribution with t degrees of freedom. Therefore, since the t-distribution has the second moment equal to $t/(t-2)$ for $t > 2$, we infer that

$$\mathbb{E}\left(\mu_i^2|\mu_{i-1}\right) = \frac{t + \mu_{i-1}^2}{t-2}.$$

Similar computation shows that the posterior marginal density of μ satisfies

$$p(\mu) \propto \left(1 + \frac{t-1}{t}\frac{\mu^2}{t-1}\right)^{-t/2}.$$

Thus, the stationary distribution of our Gibbs sampler is rescaled t-Student with $t-1$ degrees of freedom. Consequently, we have

$$\mathbb{E}_\pi \mu^2 = \frac{t}{t-3}.$$

Proposition 6.1 (Drift). *Assume that $t \geq 4$. Let $V(\mu) := \mu^2 + 1$ and $J = [-a, a]$. The transition kernel of the (2-step) Gibbs sampler satisfies*

$$PV(\mu) \leq \begin{cases} \lambda V(\mu), & \text{for } |\mu| > a; \\ K, & \text{for } |\mu| \leq a, \end{cases} \qquad \text{provided that } a > \sqrt{t/(t-3)}.$$

The quantities λ, K and $\pi(V)$ are given by

$$\lambda = \frac{1}{t-2}\left(\frac{2t-3}{1+a^2} + 1\right), \qquad K = 2 + \frac{a^2+2}{t-2} \quad \text{and} \quad \pi(V) = \frac{2t-3}{t-3}.$$

Proof. Since $a > \sqrt{t/t-3}$, we obtain that $\lambda = \frac{1}{t-2}(\frac{2t-3}{1+a^2}+1) < \frac{1}{t-2}(t-2) = 1$. Using the fact that

$$PV(\mu) = \mathbb{E}\big(\mu_i^2 + 1 | \mu_{i-1} = \mu\big) = \frac{t+\mu^2}{t-2} + 1$$

we obtain

$$\begin{aligned}
\lambda V(\mu) - PV(\mu) &= \frac{1}{t-2}\left(\frac{2t-3}{1+a^2}+1\right)\big(\mu^2+1\big) - \frac{t+\mu^2}{t-2} - 1 \\
&= \frac{1}{t-2}\left(\frac{2t-3}{1+a^2}\mu^2 + \frac{2t-3}{1+a^2} - 2t + 3\right) \\
&= \frac{2t-3}{(t-2)(1+a^2)}\big(\mu^2 + 1 - 1 - a^2\big) \\
&= \frac{2t-3}{(t-2)(1+a^2)}\big(\mu^2 - a^2\big).
\end{aligned}$$

Hence, $\lambda V(\mu) - PV(\mu) > 0$ for $|\mu| > a$. For μ such that $|\mu| \le a$, we get that

$$PV(\mu) = \frac{t+\mu^2}{t-2} + 1 \le \frac{t+a^2}{t-2} + 1 = 2 + \frac{t+a^2-t+2}{t-2} = 2 + \frac{a^2+2}{t-2}.$$

Finally,

$$\pi(V) = \mathbb{E}_\pi \mu^2 + 1 = \frac{t}{t-3} + 1 = \frac{2t-3}{t-3}. \qquad \square$$

Proposition 6.2 (Minorization). *Let $p_{\min}$ be a subprobability density given by*

$$p_{\min}(\mu) = \begin{cases} p(\mu|a), & \text{for } |\mu| \le h(a); \\ p(\mu|0), & \text{for } |\mu| > h(a), \end{cases}$$

where $p(\cdot|\cdot)$ is the transition density given by (6.1) and

$$h(a) = \left\{ a^2\left[\left(1+\frac{a^2}{t}\right)^{t/(t+1)} - 1\right]^{-1} - t\right\}^{1/2}.$$

Then $|\mu_{i-1}| \le a$ implies $p(\mu_i|\mu_{i-1}) \ge p_{\min}(\mu_i)$. Consequently, if we take for ν the probability measure with the normalized density $p_{\min}/\beta$ then the small set Assumption 2.1 holds for $J = [-a, a]$. Constant β is given by

$$\beta = 1 - \mathbb{P}\big(|\vartheta| \le h(a)\big) + \mathbb{P}\left(|\vartheta| \le \left(1+\frac{a^2}{t}\right)^{-1/2} h(a)\right),$$

where ϑ is a random variable with t-Student distribution with t degrees of freedom.

Proposition 6.2 is illustrated in Figure 1.

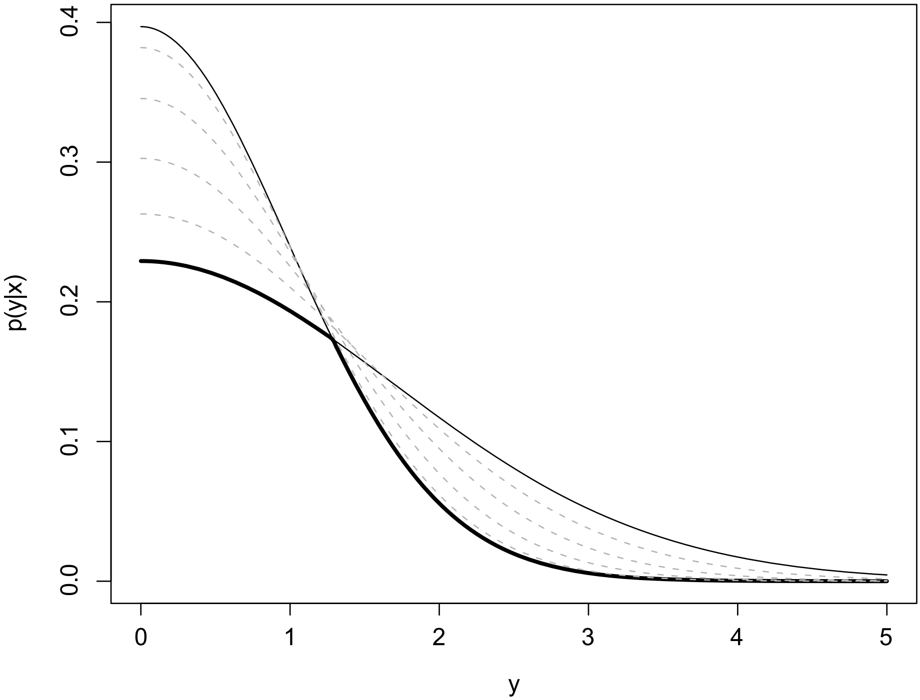

Figure 1. Illustration of Proposition 6.2, with $t = 50$ and $a = 10$. Solid lines are graphs of $p(\mu_i|0)$ and $p(\mu_i|a)$. Bold line is the graph of $p_{\min}(\mu_i)$. Gray dotted lines are graphs of $p(\mu_i|\mu_{i-1})$ for some selected positive $\mu_{i-1} \le a$.

Proof of Proposition 6.2. The formula for $p_{\min}$ results from minimisation of $p(\mu_i|\mu_{i-1})$ with respect to $\mu_{i-1} \in [-a, a]$. We use (6.1). First, compute $(\partial/\partial\mu_{i-1})p(\mu_i|\mu_{i-1})$ to check that for every μ_i the function $\mu_{i-1} \mapsto p(\mu_i|\mu_{i-1})$ has to attain minimum either at 0 or at a. Indeed,

$$\begin{aligned}\frac{\partial}{\partial\mu_{i-1}} p(\mu_i|\mu_{i-1}) &= const \cdot \biggl[\frac{t}{2}\bigl(s^2+\mu_{i-1}^2\bigr)^{t/2-1}\bigl(s^2+\mu_{i-1}^2+\mu_i^2\bigr)^{-(t+1)/2} \cdot 2\mu_{i-1} \\ &\qquad - \frac{t+1}{2}\bigl(s^2+\mu_{i-1}^2\bigr)^{t/2}\bigl(s^2+\mu_{i-1}^2+\mu_i^2\bigr)^{-(t+1)/2-1} \cdot 2\mu_{i-1}\biggr] \\ &= \mu_{i-1}\bigl(s^2+\mu_{i-1}^2\bigr)^{t/2-1}\bigl(s^2+\mu_{i-1}^2+\mu_i^2\bigr)^{-(t+1)/2-1} \\ &\quad \cdot \bigl[t\bigl(s^2+\mu_{i-1}^2+\mu_i^2\bigr) - (t+1)\bigl(s^2+\mu_{i-1}^2+\mu_i^2\bigr)\bigr].\end{aligned}$$

Assuming that $\mu_{i-1} > 0$, the first factor at the right-hand side of the above equation is positive, so $(\partial/\partial\mu_{i-1})p(\mu_i|\mu_{i-1}) > 0$ iff $t(s^2 + \mu_{i-1}^2 + \mu_i^2) - (t+1)(s^2 + \mu_{i-1}^2 + \mu_i^2) > 0$, that is iff

$$\mu_{i-1}^2 < t\mu_i^2 - s^2.$$

Consequently, if $t\mu_i^2 - s^2 \le 0$ then the function $\mu_{i-1} \mapsto p(\mu_i|\mu_{i-1})$ is decreasing for $\mu_{i-1} > 0$ and $\min_{0\le\mu_{i-1}\le a} p(\mu_i|\mu_{i-1}) = p(\mu_i, a)$. If $t\mu_i^2 - s^2 > 0$, then this function first increases and then decreases. In either case we have $\min_{0\le\mu_{i-1}\le a} p(\mu_i|\mu_{i-1}) = \min[p(\mu_i|a), p(\mu_i|0)]$. Thus using symmetry, $p(\mu_i|\mu_{i-1}) = p(\mu_i|-\mu_{i-1})$, we obtain

$$p_{\min}(\mu_i) = \min_{|\mu_{i-1}|\le a} p(\mu_i|\mu_{i-1}) = \begin{cases} p(\mu_i|a), & \text{if } p(\mu_i|a) \le p(\mu_i|0); \\ p(\mu_i|0), & \text{if } p(\mu_i|a) > p(\mu_i|0). \end{cases}$$

Now it is enough to solve the inequality, say, $p(\mu|0) < p(\mu|a)$, with respect to μ. The following elementary computation shows that this inequality is fulfilled iff $|\mu| > h(a)$:

$$p(\mu|0) = \frac{(s^2)^{t/2}}{(s^2+\mu^2)^{(t+1)/2}} < \frac{(s^2+a^2)^{t/2}}{(s^2+a^2+\mu^2)^{(t+1)/2}} = p(\mu|a), \qquad \text{iff}$$

$$\left(\frac{s^2+a^2+\mu^2}{s^2+\mu^2}\right)^{(t+1)/2} < \left(\frac{s^2+a^2}{s^2}\right)^{t/2}, \qquad \text{iff}$$

$$\left(1+\frac{a^2}{s^2+\mu^2}\right)^{t+1} < \left(1+\frac{a^2}{s^2}\right)^{t}, \qquad \text{iff}$$

$$\frac{a^2}{s^2+\mu^2} < \left(1+\frac{a^2}{s^2}\right)^{t/(t+1)} - 1, \qquad \text{iff}$$

$$\mu^2 > a^2\left[\left(1+\frac{a^2}{s^2}\right)^{t/(t+1)} - 1\right]^{-1} - s^2.$$

It is enough to recall that $s^2 = t$ and thus the right-hand side above is just $h(a)^2$.

To obtain the formula for β, note that

$$\beta = \int p_{\min}(\mu)\,\mathrm{d}\mu = \int_{|\mu|\le h(a)} p(\mu|a)\,\mathrm{d}\mu + \int_{|\mu|>h(a)} p(\mu|0)\,\mathrm{d}\mu$$

and use (6.2). □

Remark 6.3. It is interesting to compare the asymptotic behaviour of the constants in Propositions 6.1 and 6.2 for $a \to \infty$. We can immediately see that $\lambda^2 \to 1/(t-2)$ and $K^2 \sim a^2/(t-2)$. Slightly more tedious computation reveals that $h(a) \sim const \cdot a^{1/(t+1)}$ and consequently $\beta \sim const \cdot a^{-t/(t+1)}$.

The parameter of interest is the posterior mean (Bayes estimator of μ). Thus, we let $f(\mu) = \mu$ and $\theta = \mathbb{E}_\pi \mu = 0$. Note that our chain $\mu_0, \ldots, \mu_i, \ldots$ is a sequence of martingale differences, so $\bar{f} = f$ and

$$\sigma_{\mathrm{as}}^2(P, f) = \mathbb{E}_\pi\big(f^2\big) = \frac{t}{t-3}.$$

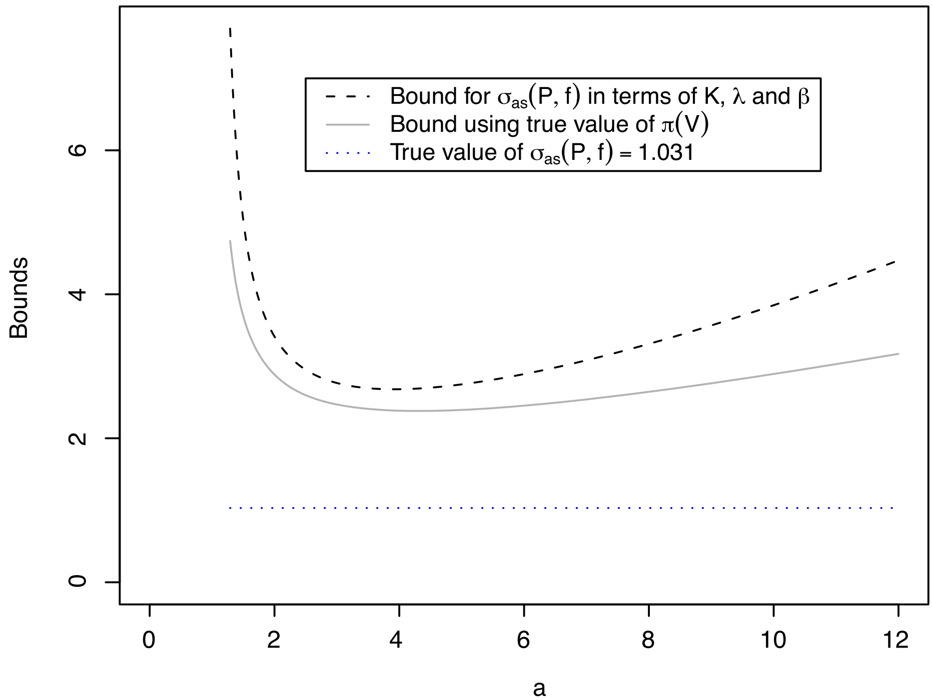

Figure 2. Bounds for the root asymptotic variance $\sigma_{\mathrm{as}}(P, f)$ as functions of a.

The MSE of the estimator $\hat{\theta}_n = \sum_{i=0}^{n-1} \mu_n$ can be also expressed analytically, namely

$$\mathrm{MSE} = \mathbb{E}_{\mu_0} \hat{\theta}_n^2 = \frac{t}{n(t-3)} - \frac{t(t-2)}{n^2(t-3)^2}\left[1 - \left(\frac{1}{t-2}\right)^n\right] + \frac{t-2}{n^2(t-3)}\left[1 - \left(\frac{1}{t-2}\right)^n\right]\mu_0^2.$$

Obviously, we have $\|f\|_{V^{1/2}} = 1$.

We now proceed to examine the bounds proved in Section 4 under the geometric drift condition, Assumption 4.1. Inequalities for the asymptotic variance play the crucial role in our approach. Let us fix $t = 50$. Figure 2 shows how our bounds on $\sigma_{\mathrm{as}}(P, f)$ depend on the choice of the small set $J = [-a, a]$.

The gray solid line gives the bound of Theorem 4.2(ii) which assumes the knowledge of πV (and uses the obvious inequality $\pi(V^{1/2}) \le (\pi V)^{1/2}$). The black dashed line corresponds to a bound which involves only λ, K and β. It is obtained if values of πV and $\pi V^{1/2}$ are replaced by their respective bounds given in Proposition 4.5(i) and (ii).

The best values of the bounds, equal to 2.68 and 2.38, correspond to $a = 3.91$ and $a = 4.30$, respectively. The actual value of the root asymptotic variance is $\sigma_{\mathrm{as}}(P, f) = 1.031$. In Table 1 below, we summarise the analogous bounds for three values of t.

The results obtained for different values of parameter t lead to qualitatively similar conclusions. From now on, we keep $t = 50$ fixed.

Table 1. Values of $\sigma_{\text{as}}(P, f)$ vs. bounds of Theorem 4.2(ii) combined with Proposition 4.5(i) and (ii) for different values of t

t	$\sigma_{\text{as}}(P, f)$	Bound with known πV	Bound involving only λ, K, β
5	1.581	6.40	11.89
50	1.031	2.38	2.68
500	1.003	2.00	2.08

Table 2 is analogous to Table 1 but focuses on other constants introduced in Theorem 3.1. Apart from $\sigma_{\text{as}}(P, f)$, we compare $C_0(P), C_1(P, f), C_2(P, f)$ with the bounds given in Theorem 4.2 and Proposition 4.5. The "actual values" of $C_0(P), C_1(P, f), C_2(P, f)$ are computed via a long Monte Carlo simulation (in which we identified regeneration epochs). The bound for $C_1(P, f)$ in Theorem 4.2(iii) depends on ξV, which is typically known, because usually simulation starts from a deterministic initial point, say x_0 (in our experiments, we put $x_0 = 0$). As for $C_2(P, f)$, its actual value varies with n. However, in our experiments the dependence on n was negligible and has been ignored (the differences were within the accuracy of the reported computations, provided that $n \geq 10$).

Finally, let us compare the actual values of the root mean square error, RMSE $:= \sqrt{\mathbb{E}_\xi(\hat{\theta}_n - \theta)^2}$, with the bounds given in Theorem 3.1. In column (a), we use the formula (3.1) with "true" values of $\sigma_{\text{as}}(P, f)$ and $C_0(P), C_1(P, f), C_2(P, f)$ given by (3.2) and (3.5). Column (b) is obtained by replacing those constants by their bounds given in Theorem 4.2 and using the true value of πV. Finally, the bounds involving only λ, K, β are in column (c).

Table 3 clearly shows that the inequalities in Theorem 3.1 are quite sharp. The bounds on RMSE in column (a) become almost exact for large n. However, the bounds on the constants in terms of minorization/drift parameters are far from being tight. While constants $C_0(P), C_1(P, f), C_2(P, f)$ have relatively small influence, the problem of bounding $\sigma_{\text{as}}(P, f)$ is of primary importance.

This clearly identifies the bottleneck of the approach: the bounds on $\sigma_{\text{as}}(P, f)$ under drift condition in Theorem 4.2 and Proposition 4.5 can vary widely in their sharpness in specific examples. We conjecture that this may be the case in general for any bounds derived under drift conditions. Known bounds on the rate of convergence (e.g., in total variation norm) obtained under drift conditions are typically very conservative, too (e.g., [4,30,52]). However, at present, drift conditions remain the main and most universal tool for proving computable bounds for

Table 2. Values of the constants appearing in Theorem 3.1 vs. bounds of Theorem 4.2 combined with Proposition 4.5

Constant	Actual value	Bound with known πV	Bound involving only λ, K, β
$C_0(P)$	0.568	1.761	2.025
$C_1(P, f)$	0.125	–	2.771
$C_2(P, f)$	1.083	–	3.752

Table 3. RMSE, its bound in Theorem 3.1 and further bounds based Theorem 4.2 combined with Proposition 4.5

n	$\sqrt{n}$ RMSE	Bound (3.1)		
		(a)	(b)	(c)
10	0.98	1.47	4.87	5.29
50	1.02	1.21	3.39	3.71
100	1.03	1.16	3.08	3.39
1000	1.03	1.07	2.60	2.89
5000	1.03	1.05	2.48	2.77
10,000	1.03	1.04	2.45	2.75
50,000	1.03	1.04	2.41	2.71

Markov chains on continuous spaces. An alternative might be working with conductance but to the best of our knowledge, so far this approach has been applied successfully only to examples with compact state spaces (see, e.g., [41,57] and references therein).

6.2. A Poisson–Gamma model

Consider a hierarchical Bayesian model applied to a well-known pump failure data set and analysed in several papers (e.g., [18,44,53,60]). Data are available for example, in [8], R package "SMPracticals" or in the cited Tierney's paper. They consist of $m = 10$ pairs (y_i, t_i) where y_i is the number of failures for ith pump, during t_i observed hours. The model assumes that:

$$\begin{aligned} y_i &\sim \operatorname{Poiss}(t_i\phi_i), \qquad \text{conditionally independent for } i = 1, \dots, m, \\ \phi_i &\sim \operatorname{Gamma}(\alpha, r), \qquad \text{conditionally i.i.d. for } i = 1, \dots, m, \\ r &\sim \operatorname{Gamma}(\sigma, \gamma). \end{aligned}$$

The posterior distribution of parameters $\phi = (\phi_1, \dots, \phi_m)$ and r is

$$p(\phi, r|y) \propto \left(\prod_{i=1}^{m} \phi_i^{y_i} \mathrm{e}^{-t_i\phi_i}\right) \cdot \left(\prod_{i=1}^{m} r^{\alpha}\phi_i^{\alpha-1} \cdot \mathrm{e}^{-r\phi_i}\right) \cdot r^{\sigma-1}\mathrm{e}^{-\gamma r},$$

where α, σ, γ are known hyperparameters. The Gibbs sampler updates cyclically r and ϕ using the following conditional distributions:

$$\begin{aligned} r|\phi, y &\sim \operatorname{Gamma}\Bigl(m\alpha + \sigma, \gamma + \sum \phi_i\Bigr), \\ \phi_i|\phi_{-i}, r, y &\sim \operatorname{Gamma}(y_i + \alpha, t_i + r). \end{aligned}$$

In what follows, the numeric results correspond to the same hyperparameter values as in the above cited papers: $\alpha = 1.802$, $\sigma = 0.01$ and $\gamma = 1$. For these values, Rosenthal in [53] constructed a small set $J = \{(\phi, r)\colon 4 \le \sum \phi_i \le 9\}$ which satisfies the one-step minorization condition (our Assumption 2.1) and established a geometric drift condition towards J (our Assumption 4.1) with $V(\phi, r) = 1 + (\sum \phi_i - 6.5)^2$. The minorization and drift constants were the following:

$$\beta = 0.14, \qquad \lambda = 0.46, \qquad K = 3.3.$$

Suppose we are to estimate the posterior expectation of a component ϕ_i. To get a bound on the (root-) MSE of the MCMC estimate, we combine Theorem 3.1 with Proposition 4.2 and Proposition 4.5. Suppose we start simulations at a point with $\sum \phi_i = 6.5$ that is, with initial value of V equal to 1. To get a better bound on $\|\bar{f}\|_{V^{1/2}}$ via Proposition 4.5(v), we first reduce $\|f\|_{V^{1/2}}$ by a vertical shift, namely we put $f(\phi, r) = \phi_i - b$ for $b = 3.327$ (expectation of ϕ_i can be immediately recovered from that of $\phi_i - b$). Elementary and easy calculations show that $\|f\|_{V^{1/2}} \leq 3.327$. We also use the bound taken from Proposition 4.5(ii) for $\pi(V)$ and the inequality $\pi(V^{1/2}) \leq \pi(V)^{1/2}$. Finally, we obtain the following values of the constants:

$$\sigma_{\mathrm{as}}(P, f) \leq 171.6 \quad \text{and} \quad C_0(P) \leq 27.5, \qquad C_1(P, f) \leq 547.7, \qquad C_2(P, f) \leq 676.1.$$

6.3. Contracting normals

As discussed in the Introduction, the results of the present paper improve over earlier MSE bounds of [35] for geometrically ergodic chains in that they are much more generally applicable and also tighter. To illustrate the improvement in tightness, we analyze the MSE and confidence estimation for the contracting normals toy-example considered in [35].

For the Markov chain transition kenel

$$P(x, \cdot) = \mathrm{N}\bigl(cx, 1 - c^2\bigr), \qquad \text{with} |c| < 1, \text{ on } \mathcal{X} = \mathbb{R},$$

with stationary distribution $\mathrm{N}(0, 1)$, consider estimating the mean, that is, put $f(x) = x$. Similarly as in [35] we take a drift function $V(x) = 1 + x^2$ resulting in $\|f\|_{V^{1/2}} = 1$. With the small set $J = [-d, d]$ with $d > 1$, the drift and regeneration parameters can be identified as

$$\lambda = c^2 + \frac{1(1 - c^2)}{1 + d^2} < 1, \qquad K = 2 + c^2\bigl(d^2 - 1\bigr),$$
$$\beta = 2\left[\Phi\left(\frac{(1 + |c|)d}{\sqrt{1 - c^2}}\right) - \Phi\left(\frac{|c|d}{\sqrt{1 - c^2}}\right)\right],$$

where Φ stands for the standard normal c.d.f. We refer to [4,35] for details on these elementary calculations.

To compare with the results of [35], we aim at confidence estimation of the mean. First, we combine Theorem 3.1 with Proposition 4.2 and Proposition 4.5 to upperbound the MSE of $\hat{\theta}_n$ and next we use the Chebyshev inequality. We derive the resulting minimal simulation length n guaranteeing

$$\mathbb{P}\bigl(|\hat{\theta}_n - \theta| < \varepsilon\bigr) > 1 - \alpha, \qquad \text{with } \varepsilon = \alpha = 0.1.$$

Table 4. Comparison of the total simulation effort n required for nonasymptotic confidence estimation $\mathbb{P}(|\hat{\theta}_n - \theta| < \varepsilon) > 1 - \alpha$ with $\varepsilon = \alpha = 0.1$ and the target function $f(x) = x$

Bound involving only λ, K, β	Bound with known πV	Bound from [35]	Reality
77,285	43,783	6,460,000,000	811

This is equivalent to finding minimal n s.t.

$$\mathrm{MSE}(\hat{\theta}_n) \le \varepsilon^2 \alpha.$$

Note that for small values of α a median trick can be applied resulting in an exponentially tight bounds, see [34,35,45] for details. The value of c is set to 0.5 and the small set half width d has been optimised numerically for each method yielding $d = 1.6226$ for the bounds from [35] and $d = 1.7875$ for the results based on our Section 4. The chain is initiated at 0, that is, $\xi = \delta_0$. Since in this setting the exact distribution of $\hat{\theta}_n$ can be computed analytically, both bounds are compared to reality, which is the exact true simulation effort required for the above confidence estimation.

As illustrated by Table 4, we obtain an improvement of 5 orders of magnitude compared to [35] and remain less then 2 orders of magnitude off the truth.

7. Preliminary lemmas

Before we proceed to the proofs for Sections 4 and 5, we need some auxiliary results that might be of independent interest.

We work under Assumptions 2.1 (small set) and 5.1 (the drift condition). Note that Assumption 4.1 is the special case of Assumption 5.1, with $\alpha = 1$. Assumption 4.1 implies

$$PV^{1/2}(x) \le \begin{cases} \lambda^{1/2} V^{1/2}(x), & \text{for } x \notin J, \\ K^{1/2}, & \text{for } x \in J, \end{cases} \tag{7.1}$$

because by Jensen's inequality $PV^{1/2}(x) \le \sqrt{PV(x)}$. Whereas for $\alpha < 1$, Lemma 3.5 of [22] for all $\eta \le 1$ yields

$$PV^{\eta}(x) \le \begin{cases} V^{\eta}(x) - \eta(1-\lambda)V(x)^{\eta+\alpha-1}, & \text{for } x \notin J, \\ K^{\eta}, & \text{for } x \in J. \end{cases} \tag{7.2}$$

The following lemma is a well-known fact which appears for example, in [47] (for bounded g). The proof for nonnegative function g is the same.

Lemma 7.1. *If* $g \ge 0$, *then*

$$\mathbb{E}_\nu \Xi(g)^2 = \mathbb{E}_\nu T \left(\mathbb{E}_\pi g(X_0)^2 + 2 \sum_{n=1}^{\infty} \mathbb{E}_\pi g(X_0) g(X_n) \mathbb{I}(T > n) \right).$$

We shall also use the generalised Kac lemma, in the following form that follows as an easy corollary from Theorem 10.0.1 of [42].

Lemma 7.2. *If $\pi(|f|) < \infty$, then*

$$\pi(f) = \int_J \mathbb{E}_x \sum_{i=1}^{\tau(J)} f(X_i)\pi(\mathrm{d}x), \qquad \textit{where}$$
$$\tau(J) := \min\{n > 0\colon X_n \in J\}. \tag{7.3}$$

The following lemma is related to other calculations in the drift conditions setting, for example, [4,10,11,13,38,55].

Lemma 7.3. *If Assumptions 2.1 and 5.1 hold, then for all $\eta \le 1$*

$$\begin{aligned}\mathbb{E}_x \sum_{n=1}^{T-1} V^{\alpha+\eta-1}(X_n) &\le \frac{V^\eta(x) - 1 + \eta(1-\lambda) - \eta(1-\lambda)V^{\alpha+\eta-1}(x)}{\eta(1-\lambda)}\mathbb{I}(x \notin J) \\ &\quad + \frac{K^\eta - 1}{\beta\eta(1-\lambda)} + \frac{1}{\beta} - 1 \\ &\le \frac{V^\eta(x)}{\eta(1-\lambda)} + \frac{K^\eta - 1 - \beta}{\beta\eta(1-\lambda)} + \frac{1}{\beta} - 1 \qquad (\textit{if additionally } \alpha+\eta \ge 1).\end{aligned}$$

Corollary 7.4. *For $\mathbb{E}_x \sum_{n=0}^{T-1} V^{\alpha+\eta-1}(X_n)$, we need to add the term $V^{\alpha+\eta-1}(x)$. Hence,*

$$\begin{aligned}\mathbb{E}_x \sum_{n=0}^{T-1} V^{\alpha+\eta-1}(X_n) &\le \frac{V^\eta(x) - 1 + \eta(1-\lambda) - \eta(1-\lambda)V^{\alpha+\eta-1}(x)}{\eta(1-\lambda)} \\ &\quad + \frac{K^\eta - 1}{\beta\eta(1-\lambda)} + \frac{1}{\beta} - 1 + V^{\alpha+\eta-1}(x) \\ &= \frac{V^\eta(x)}{\eta(1-\lambda)} + \frac{K^\eta - 1 - \beta}{\beta\eta(1-\lambda)} + \frac{1}{\beta}.\end{aligned}$$

In the case of geometric drift, the second inequality in Lemma 7.3 can be replaced by a slightly better bound. For $\alpha = \eta = 1$, the first inequality in Lemma 7.3 entails the following.

Corollary 7.5. *If Assumptions 2.1 and 4.1 hold, then*

$$\mathbb{E}_x \sum_{n=1}^{T-1} V(X_n) \le \frac{\lambda V(x)}{1-\lambda} + \frac{K - \lambda - \beta}{\beta(1-\lambda)}.$$

Proof of Lemma 7.3. The proof is given for $\eta = 1$, because for $\eta < 1$ it is identical and the constants can be obtained from (7.2).

Let $S := S_0 := \min\{n \geq 0\colon\ X_n \in J\}$ and $S_j := \min\{n > S_{j-1}\colon\ X_n \in J\}$ for $j = 1, 2, \dots$. Moreover, set

$$H(x) := \mathbb{E}_x \sum_{n=0}^{S} V^\alpha(X_n),$$

$$\tilde{H} := \sup_{x \in J} \mathbb{E}_x \left(\sum_{n=1}^{S_1} V^\alpha(X_n) \middle| \Gamma_0 = 0 \right) = \sup_{x \in J} \int Q(x, \mathrm{d}y) H(y).$$

Note that $H(x) = V^\alpha(x)$ for $x \in J$ and recall that Q denotes the normalized "residual kernel" defined in Section 2.

We will first show that

$$H(x) \leq \frac{V(x) - \lambda}{1 - \lambda} \qquad \text{for } x \in \mathcal{X}. \tag{7.4}$$

Let $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$ and remembering that $\eta = 1$, rewrite (7.2) as

$$V(X_n)^\alpha \mathbb{I}(X_n \notin J) \leq \frac{1}{1 - \lambda} \big[V(X_n) - \mathbb{E}\big(V(X_{n+1}) | \mathcal{F}_n\big)\big] \mathbb{I}(X_n \notin J). \tag{7.5}$$

Fix $x \notin J$. Since $\{X_n \notin J\} \supseteq \{S > n\} \in \mathcal{F}_n$, we can apply (7.5) and write

$$\begin{aligned}
\mathbb{E}_x \sum_{n=0}^{(S-1)\wedge m} V^\alpha(X_n) &= \mathbb{E}_x \sum_{n=0}^{m} V^\alpha(X_n) \mathbb{I}(S > n) \\
&\leq \frac{1}{1 - \lambda} \sum_{n=0}^{m} \mathbb{E}_x \big[V(X_n) - \mathbb{E}\big(V(X_{n+1}) | \mathcal{F}_n\big)\big] \mathbb{I}(S > n) \\
&= \frac{1}{1 - \lambda} \sum_{n=0}^{m} \big[\mathbb{E}_x V(X_n) \mathbb{I}(S > n) - \mathbb{E}_x \mathbb{E}\big(V(X_{n+1}) \mathbb{I}(S > n) | \mathcal{F}_n\big)\big] \\
&= \frac{1}{1 - \lambda} \sum_{n=0}^{m} \big[\mathbb{E}_x V(X_n) \mathbb{I}(S > n) - \mathbb{E}_x V(X_{n+1}) \mathbb{I}(S > n + 1) \\
&\qquad - \mathbb{E}_x V(X_{n+1}) \mathbb{I}(S = n + 1)\big] \\
&\leq \frac{1}{1 - \lambda} \Bigg[V(x) - \mathbb{E}_x V(X_{m+1}) \mathbb{I}(S > m + 1) \\
&\qquad - \sum_{n=0}^{m} \mathbb{E}_x V(X_{n+1}) \mathbb{I}(S = n + 1) \Bigg] \\
&= \frac{V(x) - \mathbb{E}_x V(X_{S \wedge (m+1)})}{1 - \lambda},
\end{aligned}$$

so

$$\begin{aligned}\mathbb{E}_x \sum_{n=0}^{S\wedge(m+1)} V^\alpha(X_n) &= \mathbb{E}_x \sum_{n=0}^{(S-1)\wedge m} V^\alpha(X_n) + \mathbb{E}_x V^\alpha(X_{S\wedge(m+1)}) \\ &\le \frac{V(x) - \mathbb{E}_x V(X_{S\wedge(m+1)})}{1-\lambda} + \mathbb{E}_x V(X_{S\wedge(m+1)}) \\ &= \frac{V(x) - \lambda \mathbb{E}_x V(X_{S\wedge(m+1)})}{1-\lambda} \le \frac{V(x)-\lambda}{1-\lambda}.\end{aligned}$$

Letting $m \to \infty$ yields equation (7.4) for $x \notin J$. For $x \in J$, (7.4) is obvious.

Next, from Assumption 5.1 we obtain $PV(x) = (1-\beta)QV(x) + \beta \nu V \le K$ for $x \in J$, so $QV(x) \le (K-\beta)/(1-\beta)$ and, taking into account (7.4),

$$\tilde{H} \le \frac{(K-\beta)/(1-\beta) - \lambda}{1-\lambda} = \frac{K - \lambda - \beta(1-\lambda)}{(1-\lambda)(1-\beta)}. \tag{7.6}$$

Recall that $T := \min\{n \ge 1\colon\ \Gamma_{n-1} = 1\}$. For $x \in J$, we thus have

$$\begin{aligned}\mathbb{E}_x \sum_{n=1}^{T-1} V^\alpha(X_n) &= \mathbb{E}_x \sum_{j=1}^{\infty} \sum_{n=S_{j-1}+1}^{S_j} V^\alpha(X_n)\mathbb{I}(\Gamma_{S_0} = \cdots = \Gamma_{S_{j-1}} = 0) \\ &= \sum_{j=1}^{\infty} \mathbb{E}_x\left(\sum_{n=S_{j-1}+1}^{S_j} V^\alpha(X_n) \middle| \Gamma_{S_0} = \cdots = \Gamma_{S_{j-1}} = 0\right)(1-\beta)^j \\ &\le \sum_{j=1}^{\infty} \tilde{H}(1-\beta)^j \le \frac{K-\lambda}{\beta(1-\lambda)} - 1,\end{aligned}$$

by (7.6). For $x \notin J$, we have to add one more term and note that the above calculation also applies.

$$\mathbb{E}_x \sum_{n=1}^{T-1} V^\alpha(X_n) = \mathbb{E}_x \sum_{n=1}^{S_0} V^\alpha(X_n) + \mathbb{E}_x \sum_{j=1}^{\infty} \sum_{n=S_{j-1}+1}^{S_j} V^\alpha(X_n)\mathbb{I}(\Gamma_{S_0} = \cdots = \Gamma_{S_{j-1}} = 0).$$

The extra term is equal to $H(x) - V^\alpha(x)$ and we use (7.4) to bound it. Finally, we obtain

$$\mathbb{E}_x \sum_{n=1}^{T-1} V^\alpha(X_n) \le \frac{V(x) - \lambda - (1-\lambda)V^\alpha(x)}{1-\lambda}\mathbb{I}(x \notin J) + \frac{K-\lambda}{\beta(1-\lambda)} - 1. \tag{7.7}$$

□

Lemma 7.6. *If Assumptions 2.1 and 5.1 hold, then*

(i) *for all* $\eta \le \alpha$

$$\pi\left(V^\eta\right) \le \left(\frac{K-\lambda}{1-\lambda}\right)^{\eta/\alpha},$$

(ii)

$$\pi(J) \ge \frac{1-\lambda}{K-\lambda},$$

(iii) *for all* $n \ge 0$ *and* $\eta \le \alpha$

$$\mathbb{E}_\nu V^\eta(X_n) \le \frac{1}{\beta^{\eta/\alpha}}\left(\frac{K-\lambda}{1-\lambda}\right)^{2\eta/\alpha}.$$

Proof. It is enough to prove (i) and (iii) for $\eta = \alpha$ and apply the Jensen inequality for $\eta < \alpha$. We shall need an upper bound on $E_x \sum_{n=1}^{\tau(J)} V^\alpha(X_n)$ for $x \in J$, where $\tau(J)$ is defined in (7.3). From the proof of Lemma 7.3,

$$\mathbb{E}_x \sum_{n=1}^{\tau(J)} V^\alpha(X_n) = PH(x) \le \frac{K-\lambda}{1-\lambda}, \qquad x \in J.$$

And by Lemma 7.2, we obtain

$$1 \le \pi V^\alpha = \int_J \mathbb{E}_x \sum_{n=1}^{\tau(J)} V^\alpha(X_n)\pi(\mathrm{d}x) \le \pi(J)\frac{K-\lambda}{1-\lambda},$$

which implies (i) and (ii).

By integrating the small set Assumption 2.1 with respect to π and from (ii) of the current lemma, we obtain

$$\frac{\mathrm{d}\nu}{\mathrm{d}\pi} \le \frac{1}{\beta\pi(J)} \le \frac{K-\lambda}{\beta(1-\lambda)}.$$

Consequently,

$$\mathbb{E}_\nu V^\alpha(X_n) = \int_{\mathcal{X}} P^n V^\alpha(x)\frac{\mathrm{d}\nu}{\mathrm{d}\pi}\pi(\mathrm{d}x) \le \frac{K-\lambda}{\beta(1-\lambda)}\int_{\mathcal{X}} P^n V^\alpha(x)\pi(\mathrm{d}x)$$
$$= \frac{K-\lambda}{\beta(1-\lambda)}\pi\left(V^\alpha\right),$$

and (iii) results from (i). □

8. Proofs for Section 4 and 5

In the proofs for Section 4, we work under Assumption 4.1 and repeatedly use Corollary 7.5.

Proof of Theorem 4.2.

(i) Recall that $C_0(P) = \mathbb{E}_\pi T - \frac{1}{2}$, write

$$\mathbb{E}_\pi T \le 1 + \mathbb{E}_\pi \sum_{n=1}^{T-1} V(X_n)$$

and use Corollary 7.5. The proof of the alternative statement (i′) uses first (7.1) and then is the same.

(ii) Without loss of generality, assume that $\|\bar{f}\|_{V^{1/2}} = 1$. By Lemma 7.1, we then have

$$\begin{aligned}\sigma_{\text{as}}^2(P, f) &= \mathbb{E}_\nu\bigl(\Xi(\bar{f})\bigr)^2/\mathbb{E}_\nu T \le \mathbb{E}_\nu\bigl(\Xi\bigl(V^{1/2}\bigr)\bigr)^2/\mathbb{E}_\nu T \\ &= \mathbb{E}_\pi V(X_0) + 2\mathbb{E}_\pi \sum_{n=1}^{T-1} V^{1/2}(X_0)V^{1/2}(X_n) =: \mathrm{I} + \mathrm{II}.\end{aligned}$$

To bound the second term, we will use Corollary 7.5 with $V^{1/2}$ in place of V, which is legitimate because of (7.1).

$$\begin{aligned}\mathrm{II}/2 &= \mathbb{E}_\pi \sum_{n=1}^{T-1} V^{1/2}(X_0)V^{1/2}(X_n) = \mathbb{E}_\pi V^{1/2}(X_0)\mathbb{E}\left(\sum_{n=1}^{T-1} V^{1/2}(X_n)\middle| X_0\right) \\ &\le \mathbb{E}_\pi V^{1/2}(X_0)\left(\frac{\lambda^{1/2}}{1-\lambda^{1/2}}V^{1/2}(X_0) + \frac{K^{1/2}-\lambda^{1/2}-\beta}{\beta(1-\lambda^{1/2})}\right) \\ &= \frac{\lambda^{1/2}}{1-\lambda^{1/2}}\pi(V) + \frac{K^{1/2}-\lambda^{1/2}-\beta}{\beta(1-\lambda^{1/2})}\pi\bigl(V^{1/2}\bigr).\end{aligned}$$

Rearranging terms in $\mathrm{I} + \mathrm{II}$, we obtain

$$\sigma_{\text{as}}^2(P, f) \le \frac{1+\lambda^{1/2}}{1-\lambda^{1/2}}\pi(V) + \frac{2(K^{1/2}-\lambda^{1/2}-\beta)}{\beta(1-\lambda^{1/2})}\pi\bigl(V^{1/2}\bigr)$$

and the proof of (ii) is complete.

(iii) The proof is similar to that of (ii) but more delicate, because we now cannot use Lemma 7.1. First, write

$$\begin{aligned}\mathbb{E}_x\bigl(\Xi\bigl(V^{1/2}\bigr)\bigr)^2 &= \mathbb{E}_x\left(\sum_{n=0}^{T-1} V^{1/2}(X_n)\right)^2 = \mathbb{E}_x\left(\sum_{n=0}^{\infty} V^{1/2}(X_n)\mathbb{I}(n<T)\right)^2 \\ &= \mathbb{E}_x \sum_{n=0}^{\infty} V(X_n)\mathbb{I}(n<T) + 2\mathbb{E}_x \sum_{n=0}^{\infty}\sum_{j=n+1}^{\infty} V^{1/2}(X_n)V^{1/2}(X_j)\mathbb{I}(j<T) \\ &=: \mathrm{I} + \mathrm{II}.\end{aligned}$$

The first term can be bounded directly using Corollary 7.5 applied to V.

$$\mathrm{I} = \mathbb{E}_x \sum_{n=0}^{\infty} V(X_n)\mathbb{I}(n < T) \le \frac{1}{1-\lambda} V(x) + \frac{K-\lambda-\beta}{\beta(1-\lambda)}.$$

To bound the second term, first condition on X_n and apply Corollary 7.5 to $V^{1/2}$, then again apply this corollary to V and to $V^{1/2}$.

$$\begin{aligned}
\mathrm{II}/2 &= \mathbb{E}_x \sum_{n=0}^{\infty} V^{1/2}(X_n)\mathbb{I}(n < T)\mathbb{E}\left(\sum_{j=n+1}^{\infty} V^{1/2}(X_j)\mathbb{I}(j < T)\middle| X_n\right) \\
&\le \mathbb{E}_x \sum_{n=0}^{\infty} V^{1/2}(X_n)\mathbb{I}(n < T)\left(\frac{\lambda^{1/2}}{1-\lambda^{1/2}} V^{1/2}(X_n) + \frac{K^{1/2}-\lambda^{1/2}-\beta}{\beta(1-\lambda^{1/2})}\right) \\
&= \frac{\lambda^{1/2}}{1-\lambda^{1/2}} \mathbb{E}_x \sum_{n=0}^{\infty} V(X_n)\mathbb{I}(n < T) + \frac{K^{1/2}-\lambda^{1/2}-\beta}{\beta(1-\lambda^{1/2})} \mathbb{E}_x \sum_{n=0}^{\infty} V^{1/2}(X_n)\mathbb{I}(n < T) \\
&\le \frac{\lambda^{1/2}}{1-\lambda^{1/2}} \left(\frac{1}{1-\lambda} V(x) + \frac{K-\lambda-\beta}{\beta(1-\lambda)}\right) \\
&\quad + \frac{K^{1/2}-\lambda^{1/2}-\beta}{\beta(1-\lambda^{1/2})} \left(\frac{1}{1-\lambda^{1/2}} V^{1/2}(x) + \frac{K^{1/2}-\lambda^{1/2}-\beta}{\beta(1-\lambda^{1/2})}\right).
\end{aligned}$$

Finally, rearranging terms in $\mathrm{I} + \mathrm{II}$, we obtain

$$\begin{aligned}
\mathbb{E}_x \bigl(\Xi\bigl(V^{1/2}\bigr)\bigr)^2 &\le \frac{1}{(1-\lambda^{1/2})^2} V(x) + \frac{2(K^{1/2}-\lambda^{1/2}-\beta)}{\beta(1-\lambda^{1/2})^2} V^{1/2}(x) \\
&\quad + \frac{\beta(K-\lambda-\beta) + 2(K^{1/2}-\lambda^{1/2}-\beta)^2}{\beta^2(1-\lambda^{1/2})^2},
\end{aligned}$$

which is tantamount to the desired result.

(iv) The proof of (iii) applies the same way. □

Proof of Proposition 4.5. For (i) and (ii) Assumption 4.1 or respectively drift condition (7.1) implies that $\pi V = \pi P V \le \lambda(\pi V - \pi(J)) + K\pi(J)$ and the result follows immediately.

(iii) and (iv) by induction: $\xi P^{n+1} V = \xi P^n (P V) \le \xi P^n (\lambda V + K) \le \lambda K/(1-\lambda) + K = K/(1-\lambda)$.

(v) We compute:

$$\begin{aligned}
\|\bar{f}\|_V &= \sup_{x\in\mathcal{X}} \frac{|f(x) - \pi f|}{V(x)} \le \sup_{x\in\mathcal{X}} \frac{|f(x)| + |\pi f|}{V(x)} \le \|f\|_V + \sup_{x\in\mathcal{X}} \frac{\pi((|f|/V)V)}{V(x)} \\
&\le \sup_{x\in\mathcal{X}} \left(\|f\|_V \left[1 + \frac{\pi V}{V(x)}\right]\right) \le \|f\|_V \left[1 + \frac{\pi(J)(K-\lambda)}{(1-\lambda)\inf_{x\in\mathcal{X}} V(x)}\right].
\end{aligned}$$

□

In the proofs for Section 5, we work under Assumption 5.1 and repeatedly use Lemma 7.3 or Corollary 7.4.

Proof of Theorem 5.2.

(i) Recall that $C_0(P) = \mathbb{E}_\pi T - \frac{1}{2}$ and write

$$\mathbb{E}_\pi T \le 1 + \mathbb{E}_\pi \sum_{i=1}^{T-1} V^{2\alpha-1}(X_n) = 1 + \int_{\mathcal{X}} \mathbb{E}_x \sum_{i=1}^{T-1} V^{2\alpha-1}(X_n)\pi(\mathrm{d}x).$$

From Lemma 7.3 with V, α and $\eta = \alpha$, we have

$$\begin{aligned} C_0(P) &\le -\frac{1}{2} + 1 + \int_{\mathcal{X}} \left(\frac{V^\alpha(x) - 1}{\alpha(1-\lambda)} + \frac{K^\alpha - 1}{\beta\alpha(1-\lambda)} + \frac{1}{\beta} - 1 \right) \pi(\mathrm{d}x) \\ &= \frac{1}{\alpha(1-\lambda)} \pi\left(V^\alpha\right) + \frac{K^\alpha - 1 - \beta}{\beta\alpha(1-\lambda)} + \frac{1}{\beta} - \frac{1}{2}. \end{aligned}$$

(ii) Without loss of generality, we can assume that $\|\bar{f}\|_{V^{(3/2)\alpha-1}} = 1$. By Lemma 7.1, we have

$$\begin{aligned} \sigma^2_{\mathrm{as}}(P, f) &= \mathbb{E}_\nu\left(\Xi(\bar{f})\right)^2 / \mathbb{E}_\nu T \le \mathbb{E}_\nu\left(\Xi\left(V^{(3/2)\alpha-1}\right)\right)^2 / \mathbb{E}_\nu T \\ &= \mathbb{E}_\pi V(X_0)^{3\alpha-2} + 2\mathbb{E}_\pi \sum_{n=1}^{T-1} V^{(3/2)\alpha-1}(X_0) V^{(3/2)\alpha-1}(X_n) =: \mathrm{I} + \mathrm{II}. \end{aligned}$$

To bound the second term, we will use Lemma 7.3 with V, α and $\eta = \frac{\alpha}{2}$.

$$\begin{aligned} \mathrm{II}/2 &= \mathbb{E}_\pi \sum_{n=1}^{T-1} V^{(3/2)\alpha-1}(X_0) V^{(3/2)\alpha-1}(X_n) = \mathbb{E}_\pi V^{(3/2)\alpha-1}(X_0) \mathbb{E}\left(\sum_{n=1}^{T-1} V^{(3/2)\alpha-1}(X_n) \middle| X_0 \right) \\ &\le \mathbb{E}_\pi V^{(3/2)\alpha-1}(X_0) \left(\frac{V^{\alpha/2}(X_0) - 1}{\alpha/2(1-\lambda)} + \frac{K^{\alpha/2} - 1}{\beta\alpha/2(1-\lambda)} + \frac{1}{\beta} - 1 \right) \\ &= \frac{2}{\alpha(1-\lambda)} \pi\left(V^{2\alpha-1}\right) + \left(\frac{2K^{\alpha/2} - 2 - 2\beta}{\alpha\beta(1-\lambda)} + \frac{1}{\beta} - 1 \right) \pi\left(V^{(3/2)\alpha-1}\right). \end{aligned}$$

The proof of (ii) is complete.

(iii) The proof is similar to that of (ii) but more delicate, because we now cannot use Lemma 7.1. Write

$$\begin{aligned} \mathbb{E}_x\left(\Xi\left(V^{(3/2)\alpha-1}\right)\right)^2 &= \mathbb{E}_x \left(\sum_{n=0}^{T-1} V^{(3/2)\alpha-1}(X_n) \right)^2 = \mathbb{E}_x \left(\sum_{n=0}^{\infty} V^{(3/2)\alpha-1}(X_n) \mathbb{I}(n < T) \right)^2 \\ &= \mathbb{E}_x \sum_{n=0}^{\infty} V^{3\alpha-2}(X_n) \mathbb{I}(n < T) \end{aligned}$$

$$
\begin{aligned}
&+ 2\mathbb{E}_x \sum_{n=0}^{\infty} \sum_{j=n+1}^{\infty} V^{(3/2)\alpha-1}(X_n) V^{(3/2)\alpha-1}(X_j)\mathbb{I}(j < T) \\
&=: \mathrm{I} + \mathrm{II}.
\end{aligned}
$$

The first term can be bounded directly using Corollary 7.4 with $\eta = 2\alpha - 1$

$$
\mathrm{I} = \mathbb{E}_x \sum_{n=0}^{\infty} V^{3\alpha-2}(X_n)\mathbb{I}(n < T) \le \frac{V^{2\alpha-1}(x)}{(2\alpha-1)(1-\lambda)} + \frac{K^{2\alpha-1} - 1 - \beta}{(2\alpha-1)\beta(1-\lambda)} + \frac{1}{\beta}.
$$

To bound the second term, first condition on X_n and use Corollary 7.4 with $\eta = \frac{\alpha}{2}$ then again use Corollary 7.4 with $\eta = \alpha$ and $\eta = \frac{\alpha}{2}$.

$$
\begin{aligned}
\mathrm{II}/2 &= \mathbb{E}_x \sum_{n=0}^{\infty} V^{(3/2)\alpha-1}(X_n)\mathbb{I}(n < T)\mathbb{E}\left(\sum_{j=n+1}^{\infty} V^{(3/2)\alpha-1}(X_j)\mathbb{I}(j < T) \Big| X_n \right) \\
&\le \mathbb{E}_x \sum_{n=0}^{\infty} V^{(3/2)\alpha-1}(X_n)\mathbb{I}(n < T)\left(\frac{2V^{\alpha/2}(X_n)}{\alpha(1-\lambda)} + \frac{2K^{\alpha/2} - 2 - 2\beta}{\alpha\beta(1-\lambda)} + \frac{1}{\beta} - 1 \right) \\
&= \frac{2}{\alpha(1-\lambda)} \mathbb{E}_x \sum_{n=0}^{\infty} V(X_n)^{2\alpha-1}\mathbb{I}(n < T) \\
&\quad + \left(\frac{2K^{\alpha/2} - 2 - 2\beta}{\alpha\beta(1-\lambda)} + \frac{1}{\beta} - 1 \right) \mathbb{E}_x \sum_{n=0}^{\infty} V^{(3/2)\alpha-1}(X_n)\mathbb{I}(n < T) \\
&\le \frac{2}{\alpha(1-\lambda)} \left(\frac{1}{\alpha(1-\lambda)} V^{\alpha}(x) + \frac{K^{\alpha} - 1 - \beta}{\alpha\beta(1-\lambda)} + \frac{1}{\beta} \right) \\
&\quad + \left(\frac{2K^{\alpha/2} - 2 - 2\beta}{\alpha\beta(1-\lambda)} + \frac{1}{\beta} - 1 \right)\left(\frac{2V^{\alpha/2}(x)}{\alpha(1-\lambda)} + \frac{2K^{\alpha/2} - 2 - 2\beta}{\alpha\beta(1-\lambda)} + \frac{1}{\beta} \right).
\end{aligned}
$$

So after gathering the terms

$$
\begin{aligned}
&\mathbb{E}_x\big(\Xi\big(V^{(3/2)\alpha-1}\big)\big)^2 \\
&\quad \le \frac{1}{(2\alpha-1)(1-\lambda)} V^{2\alpha-1}(x) + \frac{4}{\alpha^2(1-\lambda)^2} V^{\alpha}(x) + \frac{\alpha(1-\lambda) + 4}{\alpha\beta(1-\lambda)} \\
&\qquad + \left(\frac{8K^{\alpha/2} - 8 - 8\beta}{\alpha^2\beta(1-\lambda)^2} + \frac{4 - 4\beta}{\alpha\beta(1-\lambda)} \right) V^{\alpha/2}(x) + \frac{K^{2\alpha-1} - 1 - \beta}{(2\alpha-1)\beta(1-\lambda)} \\
&\qquad + \frac{4(K^{\alpha} - 1 - \beta)}{\alpha^2\beta(1-\lambda)^2} + 2\left(\frac{2K^{\alpha/2} - 2 - 2\beta}{\alpha\beta(1-\lambda)} + \frac{1}{\beta} \right)^2 - 2\left(\frac{2K^{\alpha/2} - 2 - 2\beta}{\alpha\beta(1-\lambda)} + \frac{1}{\beta} \right).
\end{aligned} \tag{8.1}
$$

(iv) Recall that $C_2(P, f)^2 = \mathbb{E}_\xi(\sum_{i=n}^{T_{R(n)}-1} |\bar{f}(X_i)|\mathbb{I}(T < n))^2$ and we have

$$\begin{aligned}
&\mathbb{E}_\xi\left(\sum_{i=n}^{T_{R(n)}-1} \left|\bar{f}(X_i)\right|\mathbb{I}(T < n)\right)^2 \\
&\quad = \sum_{j=1}^{n} \mathbb{E}_\xi\left(\left(\sum_{i=n}^{T_{R(n)}-1} \left|\bar{f}(X_i)\right|\mathbb{I}(T < n)\right)^2 \middle| T = j\right)\mathbb{P}_\xi(T = j) \\
&\quad \le \sum_{j=1}^{n} \mathbb{E}_\nu\left(\sum_{i=n-j}^{T_{R(n-j)}-1} \left|\bar{f}(X_i)\right|\right)^2 \mathbb{P}_\xi(T = j) \\
&\quad = \sum_{j=1}^{n} \mathbb{E}_{\nu P^{n-j}}\left(\sum_{i=0}^{T-1} \left|\bar{f}(X_i)\right|\right)^2 \mathbb{P}_\xi(T = j).
\end{aligned} \tag{8.2}$$

Since

$$\mathbb{E}_{\nu P^{n-j}}\left(\sum_{i=0}^{T-1} \left|\bar{f}(X_i)\right|\right)^2 = \nu P^{n-j}\left(\mathbb{E}_x\left(\sum_{i=0}^{T-1} \left|\bar{f}(X_i)\right|\right)^2\right)$$

and $|\bar{f}| \le V^{(3/2)\alpha-1}$ we put (8.1) into (8.2) and apply Lemma 7.6 to complete the proof. □

Proof of Proposition 5.4. For (i) see Lemma 7.6. For (ii), we compute:

$$\begin{aligned}
\|\bar{f}\|_{V^\eta} &= \sup_{x\in\mathcal{X}} \frac{|f(x) - \pi f|}{V^\eta(x)} \le \sup_{x\in\mathcal{X}} \frac{|f(x)| + |\pi f|}{V^\eta(x)} \le \|f\|_{V^\eta} + \sup_{x\in\mathcal{X}} \frac{\pi((|f|/V^\eta)V^\eta)}{V^\eta(x)} \\
&\le \sup_{x\in\mathcal{X}}\left(\|f\|_{V^\eta}\left[1 + \frac{\pi V^\eta}{V^\eta(x)}\right]\right) \le \|f\|_{V^\eta}\left(1 + \pi\left(V^\eta\right)\right).
\end{aligned}$$
□

Acknowledgements

We thank Gersende Fort and Jacek Wesołowski for insightful comments on earlier versions of this work. We also gratefully acknowledge the help of Agnieszka Perduta and comments of three referees and an Associate Editor that helped improve the paper.

Work partially supported by Polish Ministry of Science and Higher Education Grants No. N N201387234 and N N201 608740. Krzysztof Łatuszyński was also partially supported by EPSRC.

References

[1] Adamczak, R. (2008). A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.* **13** 1000–1034. MR2424985

[2] Aldous, D. (1987). On the Markov chain simulation method for uniform combinatorial distributions and simulated annealing. *Probab. Engrg. Inform. Sci.* **1** 33–46.

[3] Athreya, K.B. and Ney, P. (1978). A new approach to the limit theory of recurrent Markov chains. *Trans. Amer. Math. Soc.* **245** 493–501. MR0511425

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

[5] Bednorz, W., Łatuszyński, K. and Latała, R. (2008). A regeneration proof of the central limit theorem for uniformly ergodic Markov chains. *Electron. Commun. Probab.* **13** 85–98. MR2386065

[6] Bertail, P. and Clémençon, S. (2010). Sharp bounds for the tails of functionals of Markov chains. *Theory Probab. Appl.* **54** 1–19.

[7] Clémençon, S.J.M. (2001). Moment and probability inequalities for sums of bounded additive functionals of regular Markov chains via the Nummelin splitting technique. *Statist. Probab. Lett.* **55** 227–238. MR1867526

[8] Davison, A.C. (2003). *Statistical Models. Cambridge Series in Statistical and Probabilistic Mathematics* **11**. Cambridge: Cambridge Univ. Press. MR1998913

[9] 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

[10] Douc, R., Guillin, A. and Moulines, E. (2008). Bounds on regeneration times and limit theorems for subgeometric Markov chains. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** 239–257. MR2446322

[11] Douc, R., Moulines, E. and Rosenthal, J.S. (2004). Quantitative bounds on convergence of time-inhomogeneous Markov chains. *Ann. Appl. Probab.* **14** 1643–1665. MR2099647

[12] Douc, R., Moulines, E. and Soulier, P. (2007). Computable convergence rates for sub-geometric ergodic Markov chains. *Bernoulli* **13** 831–848. MR2348753

[13] Fort, G. (2003). Computable bounds for V-geometric ergodicity of Markov transition kernels. Preprint, Université Joseph Fourier, Grenoble, France.

[14] Fort, G. and Moulines, E. (2000). V-subgeometric ergodicity for a Hastings–Metropolis algorithm. *Statist. Probab. Lett.* **49** 401–410. MR1796485

[15] Fort, G. and Moulines, E. (2003). Convergence of the Monte Carlo expectation maximization for curved exponential families. *Ann. Statist.* **31** 1220–1259. MR2001649

[16] Fort, G. and Moulines, E. (2003). Polynomial ergodicity of Markov transition kernels. *Stochastic Process. Appl.* **103** 57–99. MR1947960

[17] Fort, G., Moulines, E., Roberts, G.O. and Rosenthal, J.S. (2003). On the geometric ergodicity of hybrid samplers. *J. Appl. Probab.* **40** 123–146. MR1953771

[18] Gelfand, A.E. and Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85** 398–409. MR1141740

[19] Gillman, D. (1998). A Chernoff bound for random walks on expander graphs. *SIAM J. Comput.* **27** 1203–1220. MR1621958

[20] Glynn, P.W. and Ormoneit, D. (2002). Hoeffding's inequality for uniformly ergodic Markov chains. *Statist. Probab. Lett.* **56** 143–146. MR1881167

[21] Hobert, J.P. and Geyer, C.J. (1998). Geometric ergodicity of Gibbs and block Gibbs samplers for a hierarchical random effects model. *J. Multivariate Anal.* **67** 414–430. MR1659196

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

[23] Jarner, S.F. and Roberts, G.O. (2007). Convergence of heavy-tailed Monte Carlo Markov chain algorithms. *Scand. J. Statist.* **34** 781–815. MR2396939

[24] Jarner, S.F. and Tweedie, R.L. (2003). Necessary conditions for geometric and polynomial ergodicity of random-walk-type Markov chains. *Bernoulli* **9** 559–578. MR1996270

[25] Jerrum, M.R., Valiant, L.G. and Vazirani, V.V. (1986). Random generation of combinatorial structures from a uniform distribution. *Theoret. Comput. Sci.* **43** 169–188. MR0855970
[26] Johnson, A.A. and Jones, G.L. (2007). Gibbs sampling for a Bayesian hierarchical version of the general linear mixed model. Preprint, available at arXiv:0712.3056v1.
[27] Johnson, A.A. and Jones, G.L. (2010). Gibbs sampling for a Bayesian hierarchical general linear model. *Electron. J. Stat.* **4** 313–333. MR2645487
[28] Jones, G.L., Haran, M., Caffo, B.S. and Neath, R. (2006). Fixed-width output analysis for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **101** 1537–1547. MR2279478
[29] Jones, G.L. and Hobert, J.P. (2001). Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statist. Sci.* **16** 312–334. MR1888447
[30] Jones, G.L. and Hobert, J.P. (2004). Sufficient burn-in for Gibbs samplers for a hierarchical random effects model. *Ann. Statist.* **32** 784–817. MR2060178
[31] Joulin, A. and Ollivier, Y. (2010). Curvature, concentration and error estimates for Markov chain Monte Carlo. *Ann. Probab.* **38** 2418–2442. MR2683634
[32] Kontorovich, L. and Ramanan, K. (2008). Concentration inequalities for dependent random variables via the martingale method. *Ann. Probab.* **36** 2126–2158. MR2478678
[33] Kontoyiannis, I., Lastras-Montano, L.A. and Meyn, S.P. (2005). Relative entropy and exponential deviation bounds for general Markov chains. In *IEEE, International Symposium on Information Theory* 1563–1567. Adelaide, Australia: IEEE.
[34] Łatuszyński, K., Miasojedow, B. and Niemiro, W. (2012). Nonasymptotic bounds on the mean square error for MCMC estimates via renewal techniques. In *Monte Carlo and Quasi-Monte Carlo 2010. Springer Proceedings in Mathematics and Statistics*. **23** 539–555. Berlin: Springer.
[35] Łatuszyński, K. and Niemiro, W. (2011). Rigorous confidence bounds for MCMC under a geometric drift condition. *J. Complexity* **27** 23–38. MR2745298
[36] León, C.A. and Perron, F. (2004). Optimal Hoeffding bounds for discrete reversible Markov chains. *Ann. Appl. Probab.* **14** 958–970. MR2052909
[37] Lorden, G. (1970). On excess over the boundary. *Ann. Math. Statist.* **41** 520–527. MR0254981
[38] Lund, R.B. and Tweedie, R.L. (1996). Geometric convergence rates for stochastically ordered Markov chains. *Math. Oper. Res.* **21** 182–194. MR1385873
[39] Marchev, D. and Hobert, J.P. (2004). Geometric ergodicity of van Dyk and Meng's algorithm for the multivariate Student's t model. *J. Amer. Statist. Assoc.* **99** 228–238. MR2054301
[40] Marton, K. (1996). A measure concentration inequality for contracting Markov chains. *Geom. Funct. Anal.* **6** 556–571. MR1392329
[41] Mathé, P. and Novak, E. (2007). Simple Monte Carlo and the Metropolis algorithm. *J. Complexity* **23** 673–696. MR2372022
[42] Meyn, S.P. and Tweedie, R.L. (1993). *Markov Chains and Stochastic Stability*. *Communications and Control Engineering Series*. London: Springer London Ltd. MR1287609
[43] Meyn, S.P. and Tweedie, R.L. (1994). Computable bounds for geometric convergence rates of Markov chains. *Ann. Appl. Probab.* **4** 981–1011. MR1304770
[44] Mykland, P., Tierney, L. and Yu, B. (1995). Regeneration in Markov chain samplers. *J. Amer. Statist. Assoc.* **90** 233–241. MR1325131
[45] Niemiro, W. and Pokarowski, P. (2009). Fixed precision MCMC estimation by median of products of averages. *J. Appl. Probab.* **46** 309–329. MR2535815
[46] Nummelin, E. (1978). A splitting technique for Harris recurrent Markov chains. *Z. Wahrsch. Verw. Gebiete* **43** 309–318. MR0501353
[47] Nummelin, E. (2002). MC's for MCMC'ists. *International Statistical Review* **70** 215–240.
[48] Papaspiliopoulos, O. and Roberts, G. (2008). Stability of the Gibbs sampler for Bayesian hierarchical models. *Ann. Statist.* **36** 95–117. MR2387965

[49] Roberts, G.O. and Rosenthal, J.S. (2001). Small and pseudo-small sets for Markov chains. *Stoch. Models* **17** 121–145. MR1853437

[50] Roberts, G.O. and Rosenthal, J.S. (2004). General state space Markov chains and MCMC algorithms. *Probab. Surv.* **1** 20–71. MR2095565

[51] Roberts, G.O. and Rosenthal, J.S. (2011). Quantitative non-geometric convergence bounds for independence samplers. *Methodol. Comput. Appl. Probab.* **13** 391–403. MR2788864

[52] Roberts, G.O. and Tweedie, R.L. (1999). Bounds on regeneration times and convergence rates for Markov chains. *Stochastic Process. Appl.* **80** 211–229. MR1682243

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

[54] Rosenthal, J.S. (1995). Rates of convergence for Gibbs sampling for variance component models. *Ann. Statist.* **23** 740–761. MR1345197

[55] Rosenthal, J.S. (2002). Quantitative convergence rates of Markov chains: A simple account. *Electron. Commun. Probab.* **7** 123–128 (electronic). MR1917546

[56] Roy, V. and Hobert, J.P. (2010). On Monte Carlo methods for Bayesian multivariate regression models with heavy-tailed errors. *J. Multivariate Anal.* **101** 1190–1202. MR2595301

[57] Rudolf, D. (2009). Explicit error bounds for lazy reversible Markov chain Monte Carlo. *J. Complexity* **25** 11–24. MR2475305

[58] Samson, P.M. (2000). Concentration of measure inequalities for Markov chains and Φ-mixing processes. *Ann. Probab.* **28** 416–461. MR1756011

[59] Tan, A. and Hobert, J.P. (2009). Block Gibbs sampling for Bayesian random effects models with improper priors: Convergence and regeneration. *J. Comput. Graph. Statist.* **18** 861–878. MR2598033

[60] Tierney, L. (1994). Markov chains for exploring posterior distributions. *Ann. Statist.* **22** 1701–1762. With discussion and a rejoinder by the author. MR1329166

Received August 2011 and revised January 2012