

Nonasymptotic bounds on the estimation error of MCMC algorithms *

Krzysztof Łatuszyński, Błażej Miasojedow and Wojciech Niemiro

K. Łatuszyński

Department of Statistics

University of Warwick

CV4 7AL, Coventry, UK

e-mail: latuch@gmail.com

url: <http://www.warwick.ac.uk/go/klatuszynski>

B. Miasojedow

Institute of Applied Mathematics and Mechanics

University of Warsaw

Banacha 2, 02-097 Warszawa, Poland

e-mail: bmia@mimuw.edu.pl

W. Niemiro

Faculty of Mathematics and Computer Science

Nicolaus Copernicus University

Chopina 12/18, 87-100 Toruń, Poland

e-mail: wniem@mat.uni.torun.pl

Abstract: We address the problem of upper bounding the mean square error of MCMC estimators. Our analysis is non-asymptotic. 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 $\sigma_{\text{as}}^2(P, f)/n$, where $\sigma_{\text{as}}^2(P, f)$ is the CLT asymptotic variance. Next, we proceed to specific assumptions and give explicit computable bounds for geometrically and polynomially ergodic Markov chains. As a corollary we provide results on confidence estimation.

AMS 2000 subject classifications: Primary 60J05, 65C05; secondary 62F15.

Keywords and phrases: Mean square error, Computable bounds, Geometric ergodicity, Polynomial ergodicity, Drift conditions, Regeneration, Asymptotic variance, Confidence estimation.

1. Introduction

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

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

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

*Work partially supported by Polish Ministry of Science and Higher Education Grants No. N N201387234 and N N201 608740

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 (c.f. [Ros95a, JH01, JHCN06]). The aim of our paper is to derive non-asymptotic and explicit bounds on the mean square error,

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

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 $\sigma_{\text{as}}^2(P, f)/n$, where $\sigma_{\text{as}}^2(P, f)$ 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. Uniform ergodicity is stronger than 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 [LMN11].

The Subgeometric condition, considered in e.g. [DGM08], 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 [JVV86], further developed in [NP09]. 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 a benchmark example which is related to a simplified hierarchical Bayesian model.

The paper is organized 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 numerical example is presented in Section 6. Technical proofs are deferred to Sections 7, 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 [Ald87, Gil98, LP04, NP09] 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 [GO02, KLMM05] can be applied to derive confidence bounds. In the accompanying paper [LMN11] 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 e.g. [Mar96], [Sam00] and [KR08] (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 $\sup_{x,y \in \mathcal{X}} \|P^i(x, \cdot) - P^i(y, \cdot)\|_{\text{tv}}$ and when applied to our setting, they are well suited for *bounded* functionals of *uniformly* ergodic Markov chains, but can not be applied to geometrically ergodic chains. For details we refer to the original papers and the discussion in Section 3.5 of [Ada08].

For lazy reversible Markov chains, nonasymptotic mean square error bounds have been obtained for *bounded* target functions in [Rud09] 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 [Rud09] 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 than 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 [Clé01], [Ada08] and [DGM08] using regeneration techniques. These results apply e.g. to *geometrically* or *subgeometrically* ergodic Markov chains, however they also involve non-explicit 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 Cléménçon in [BC10] by regenerative approach and using truncation arguments. However they involve non-explicit constants and can not be directly applied to confidence estimation.

Recent work [JO10] address error estimates for MCMC algorithms under positive curvature condition. The positive curvature implies geometric ergodicity in the Wasserstein distance and bivariate drift conditions (c.f. [RR01]). Their approach appears to be applicable in different settings to ours and also rests on different notions, e.g. 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 [JO10] appear to be complementary.

Nonasymptotic rates of convergence of geometrically, polynomially and sub-geometrically ergodic Markov chains to their stationary distributions have been investigated in many papers [MT94, Ros95b, RT99, Ros02, JH04, For03, DMR04, Bax05, FM03b, DMS07, RR11] 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 the accuracy of estimation, because they allow us 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 [Num78] and Athreya and Ney [AN78].

2.1 Assumption (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

$$(2.2) \quad \mathbb{P}(\Gamma_{n-1} = 1 | X_{n-1} = x) = \beta \mathbb{I}(x \in J).$$

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

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

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

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

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

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

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}_ξ and \mathbb{E}_ξ 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 (2.2).

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

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

3. A General Inequality for the MSE

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

3.1 Theorem. *If Assumption 2.1 holds then*

$$(3.2) \quad \sqrt{\mathbb{E}_\xi(\hat{\theta}_n - \theta)^2} \leq \frac{\sigma_{\text{as}}(P, f)}{\sqrt{n}} \left(1 + \frac{C_0(P)}{n}\right) + \frac{C_1(P, f)}{n} + \frac{C_2(P, f)}{n},$$

where

$$(3.3) \quad \sigma_{\text{as}}^2(P, f) := \frac{\mathbb{E}_\nu \Xi(\bar{f})^2}{\mathbb{E}_\nu T},$$

$$(3.4) \quad C_0(P) := \mathbb{E}_\pi T - \frac{1}{2}$$

$$(3.5) \quad C_1(P, f) := \sqrt{\mathbb{E}_\xi \Xi(|\bar{f}|)^2},$$

$$(3.6) \quad C_2(P, f) = C_2(P, f, n) := \sqrt{\mathbb{E}_\xi \left(\mathbb{I}(T_1 < n) \sum_{i=n}^{T_{R(n)}-1} |\bar{f}|(X_i) \right)^2},$$

$$(3.7) \quad R(n) := \min\{r \geq 1 : T_r > n\}.$$

3.8 REMARK. The bound in Theorem 3.1 is meaningful only if $\sigma_{\text{as}}^2(P, f) < \infty$, $C_0(P) < \infty$, $C_1(P, f) < \infty$ 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 [BLL08] in a more general setting. For our purposes it is important to note that $\sigma_{\text{as}}^2(P, f)$ in Theorem 3.1 is indeed the *asymptotic variance* which appears in the CLT, that is

$$\sqrt{n} \left(\hat{\theta}_n - \theta \right) \rightarrow_d \text{N}(0, \sigma_{\text{as}}^2(P, f)).$$

Moreover,

$$\lim_{n \rightarrow \infty} n \mathbb{E}_\xi \left(\hat{\theta}_n - \theta \right)^2 = \sigma_{\text{as}}^2(P, f).$$

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

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

3.10 REMARK. In our related work [LMN11], we discuss a special case of the setting considered here, namely when regeneration times T_k are identifiable. This leads to $X_0 \sim \nu$ and an regenerative estimator of the form

$$(3.11) \quad \hat{\theta}_{T_{R(n)}} := \frac{1}{T_{R(n)}} \sum_{i=1}^{R(n)} \Xi_i(f) = \frac{1}{T_{R(n)}} \sum_{i=0}^{T_{R(n)}-1} f(X_i).$$

The estimator $\hat{\theta}_{T_{R(n)}}$ is somewhat easier to analyze and particularly well suited for uniformly ergodic chains, however its applicability is limited. We refer to [LMN11] for details.

Proof of Theorem 3.1. Recall $R(n)$ defined in (3.7) 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

$$(3.12) \quad \sqrt{\mathbb{E}_\xi (\hat{\theta}_n - \theta)^2} \leq \frac{1}{n} \left(\sqrt{\mathbb{E}_\xi \mathcal{Z}^2} + \sqrt{\mathbb{E}_\xi (\mathcal{O}_1 - \mathcal{O}_2)^2} \right).$$

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

$$\begin{aligned} C(P, f) &= \left(\mathbb{E}_\xi \left\{ \left(\sum_{i=0}^{T-1} \bar{f}(X_i) - \sum_{i=n}^{T_{R(n)}-1} \bar{f}(X_i) \right) \mathbb{I}(T \geq n) \right. \right. \\ &\quad \left. \left. + \left(\sum_{i=0}^{T-1} \bar{f}(X_i) - \sum_{i=n}^{T_{R(n)}-1} \bar{f}(X_i) \right) \mathbb{I}(T < n) \right\}^2 \right)^{\frac{1}{2}} \\ &\leq \left(\mathbb{E}_\xi \left(\sum_{i=0}^{T-1} |\bar{f}(X_i)| + \sum_{i=n}^{T_{R(n)}-1} |\bar{f}(X_i)| \mathbb{I}(T < n) \right)^2 \right)^{\frac{1}{2}} \\ &\leq \sqrt{\mathbb{E}_\xi \left(\sum_{i=0}^{T-1} |\bar{f}(X_i)| \right)^2} + \sqrt{\mathbb{E}_\xi \left(\sum_{i=n}^{T_{R(n)}-1} |\bar{f}(X_i)| \mathbb{I}(T < n) \right)^2} \\ (3.13) \quad &= C_1(P, f) + C_2(P, f). \end{aligned}$$

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:

$$(3.14) \quad \mathbb{E}_\nu \left(\sum_{i=0}^{T_{R(n)}-1} \bar{f}(X_i) \right)^2 \leq \sigma_{\text{as}}^2(P, f) (n + 2C_0(P)).$$

Once this is proved, it is easy to see that

$$\begin{aligned}
\mathbb{E}_\xi \mathcal{Z}^2 &\leq \sum_{j=1}^n \mathbb{E}_\xi (\mathcal{Z}^2 | T_1 = j) \mathbb{P}_\xi(T_1 = j) \\
&= \sum_{j=1}^n \mathbb{E}_\nu \left(\sum_{i=0}^{T_{R(n-j)}-1} \bar{f}(X_i) \right)^2 \mathbb{P}_\xi(T_1 = j) \\
&\leq \sum_{j=1}^n \sigma_{\text{as}}^2(P, f) (n - j + 2C_0(P)) \mathbb{P}_\xi(T_1 = j) \\
&\leq \sigma_{\text{as}}^2(P, f) (n + 2C_0(P)),
\end{aligned}$$

consequently $\sqrt{\mathbb{E}_\xi \mathcal{Z}^2} \leq \sqrt{n} \sigma_{\text{as}}(P, f) (1 + C_0(P)/n)$ and the conclusion will follow by recalling (3.12) and (3.13).

We are therefore left with the task of proving (3.14). This is essentially a statement about sums of i.i.d. random variables. Indeed,

$$(3.15) \quad \sum_{i=0}^{T_{R(n)}-1} \bar{f}(X_i) = \sum_{k=1}^{R(n)} \Xi_k(\bar{f})$$

and all the blocks Ξ_k (including $\Xi = \Xi_1$) are i.i.d. under \mathbb{P}_ν . By the Kac theorem ([MT93] or [Num02]) 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 $\text{Var}_\nu \Xi(\bar{f}) = \sigma_{\text{as}}^2(P, f) \mathbb{E}_\nu T$. 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

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

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

$$(3.16) \quad \mathbb{E}_\nu \left(\sum_{k=1}^{R(n)} \Xi_k(\bar{f}) \right)^2 \leq \sigma_{\text{as}}^2(P, f) \mathbb{E}_\nu T_{R(n)} = \sigma_{\text{as}}^2(P, f) (n + \mathbb{E}_\nu \Delta(n)).$$

We now focus attention on bounding the “mean overshoot” $\mathbb{E}_\nu \Delta(n)$. Under \mathbb{P}_ν , the cumulative sums $T = T_1 < T_2 < \dots < T_k < \dots$ form a (nondelayed) renewal process in discrete time. Let us invoke the following elegant theorem of Lorden ([Lor70], Theorem 1):

$$(3.17) \quad \mathbb{E}_\nu \Delta(n) \leq \frac{\mathbb{E}_\nu T^2}{\mathbb{E}_\nu T}.$$

By Lemma 7.3 with $g \equiv 1$ from section 7 we obtain:

$$(3.18) \quad \mathbb{E}_\nu \Delta(n) \leq 2 \mathbb{E}_\pi T - 1$$

Hence substituting (3.18) into (3.16) and taking into account (3.15) we obtain (3.14) and complete the proof. \square

4. Geometrically Ergodic Chains

In this section we upper bound constants $\sigma_{\text{as}}^2(P, f), C_0(P), C_1(P, f), C_2(P, f)$, 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 [RR04] or [MT93] for the definition and further details. The assumption below is the same as in [Bax05]. Specifically, let J be the small set which appears in Assumption 2.1.

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

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

In many papers conditions similar to Assumption 4.1 have been established for realistic MCMC algorithms in statistical models of practical relevance [HG98, FM00, FMRR03, JH04, JJ10, RH10]. 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.

4.2 Theorem. *If Assumptions 2.1 and 4.1 hold and f is such that $\|\bar{f}\|_{V^{\frac{1}{2}}} := \sup_x |\bar{f}(x)|/V^{\frac{1}{2}}(x) < \infty$, then*

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

$$(ii) \quad \frac{\sigma_{\text{as}}^2(P, f)}{\|\bar{f}\|_{V^{\frac{1}{2}}}^2} \leq \frac{1 + \lambda^{\frac{1}{2}}}{1 - \lambda^{\frac{1}{2}}} \pi(V) + \frac{2(K^{\frac{1}{2}} - \lambda^{\frac{1}{2}} - \beta)}{\beta(1 - \lambda^{\frac{1}{2}})} \pi(V^{\frac{1}{2}}),$$

$$(iii) \quad \frac{C_1(P, f)^2}{\|\bar{f}\|_{V^{\frac{1}{2}}}^2} \leq \frac{1}{(1 - \lambda^{\frac{1}{2}})^2} \xi(V) + \frac{2(K^{\frac{1}{2}} - \lambda^{\frac{1}{2}} - \beta)}{\beta(1 - \lambda^{\frac{1}{2}})^2} \xi(V^{\frac{1}{2}}) \\ + \frac{\beta(K - \lambda - \beta) + 2(K^{\frac{1}{2}} - \lambda^{\frac{1}{2}} - \beta)^2}{\beta^2(1 - \lambda^{\frac{1}{2}})^2}.$$

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

4.3 REMARK. A related result to Theorem 4.2 combined with Theorem 3.1 is Proposition 2 of [FM03a] where the L^p norm of $\hat{\theta}_n$ for $p \geq 2$ is controlled under similar assumptions. Specialized to $p = 2$ the bound of [FM03a] is of order $n^{-1}\beta^{-3}(1 - \lambda)^{-4}$ compared to ours $n^{-1}\beta^{-1}(1 - \lambda)^{-1}$.

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

$$(i') \quad C_0(P) \leq \frac{\lambda^{\frac{1}{2}}}{1 - \lambda^{\frac{1}{2}}} \pi(V^{\frac{1}{2}}) + \frac{K^{\frac{1}{2}} - \lambda^{\frac{1}{2}} - \beta}{\beta(1 - \lambda^{\frac{1}{2}})} + \frac{1}{2}.$$

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

4.5 Proposition. *Under Assumptions 2.1 and 4.1,*

$$(i) \quad \pi(V^{\frac{1}{2}}) \leq \pi(J) \frac{K^{\frac{1}{2}} - \lambda^{\frac{1}{2}}}{1 - \lambda^{\frac{1}{2}}} \leq \frac{K^{\frac{1}{2}} - \lambda^{\frac{1}{2}}}{1 - \lambda^{\frac{1}{2}}}, \\ (ii) \quad \pi(V) \leq \pi(J) \frac{K - \lambda}{1 - \lambda} \leq \frac{K - \lambda}{1 - \lambda}, \\ (iii) \quad \text{if } \xi(V^{\frac{1}{2}}) \leq \frac{K^{\frac{1}{2}}}{1 - \lambda^{\frac{1}{2}}} \quad \text{then } \xi P^n(V^{\frac{1}{2}}) \leq \frac{K^{\frac{1}{2}}}{1 - \lambda^{\frac{1}{2}}}, \\ (iv) \quad \text{if } \xi(V) \leq \frac{K}{1 - \lambda} \quad \text{then } \xi P^n(V) \leq \frac{K}{1 - \lambda}, \\ (v) \quad \|\bar{f}\|_{V^{\frac{1}{2}}} \text{ can be related to } \|f\|_{V^{\frac{1}{2}}} \text{ by} \\ \|\bar{f}\|_{V^{\frac{1}{2}}} \leq \|f\|_{V^{\frac{1}{2}}} \left[1 + \frac{\pi(J)(K^{\frac{1}{2}} - \lambda^{\frac{1}{2}})}{(1 - \lambda^{\frac{1}{2}}) \inf_{x \in \mathcal{X}} V^{\frac{1}{2}}(x)} \right] \\ \leq \|f\|_{V^{\frac{1}{2}}} \left[1 + \frac{K^{\frac{1}{2}} - \lambda^{\frac{1}{2}}}{1 - \lambda^{\frac{1}{2}}} \right].$$

4.6 REMARK. 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 $V^{\frac{1}{2}}(x) \leq K^{\frac{1}{2}}/(1 - \lambda^{\frac{1}{2}})$ and $V(x) \leq K/(1 - \lambda)$, 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, 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 .

4.7 REMARK. 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 $\sigma_{\text{as}}^2(P, f), C_0(P), C_1(P, f), C_2(P, f)$, 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 [JR02, DFMS04, DGM08, MT93].

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

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

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 e.g. [FM00, JT03, JR07].

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.

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

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

$$(ii) \quad \frac{\sigma_{\text{as}}^2(P, f)}{\|\bar{f}\|_{V^{\frac{3}{2}\alpha-1}}^2} \leq \pi(V^{3\alpha-2}) + \frac{4}{\alpha(1-\lambda)}\pi(V^{2\alpha-1})$$

$$\begin{aligned}
& +2 \left(\frac{2K^{\frac{\alpha}{2}} - 2 - 2\beta}{\alpha\beta(1-\lambda)} + \frac{1}{\beta} - 1 \right) \pi(V^{\frac{3}{2}\alpha-1}), \\
(iii) \quad \frac{C_1(P, f)^2}{\|f\|_{V^{\frac{3}{2}\alpha-1}}^2} & \leq \frac{1}{(2\alpha-1)(1-\lambda)} \xi(V^{2\alpha-1}) + \frac{4}{\alpha^2(1-\lambda)^2} \xi(V^\alpha) \\
& + \left(\frac{8K^{\frac{\alpha}{2}} - 8 - 8\beta}{\alpha^2\beta(1-\lambda)^2} + \frac{4-4\beta}{\alpha\beta(1-\lambda)} \right) \xi(V^{\frac{\alpha}{2}}) \\
& + \frac{\alpha(1-\lambda) + 4}{\alpha\beta(1-\lambda)} \frac{K^{2\alpha-1} - 1 - \beta}{(2\alpha-1)\beta(1-\lambda)} \\
& + \frac{4(K^\alpha - 1 - \beta)}{\alpha^2\beta(1-\lambda)^2} + 2 \left(\frac{2K^{\frac{\alpha}{2}} - 2 - 2\beta}{\alpha\beta(1-\lambda)} + \frac{1}{\beta} \right)^2 \\
& - 2 \left(\frac{2K^{\frac{\alpha}{2}} - 2 - 2\beta}{\alpha\beta(1-\lambda)} + \frac{1}{\beta} \right). \\
(iv) \quad \frac{C_2(P, f)^2}{\|f\|_{V^{\frac{3}{2}\alpha-1}}^2} & \leq \frac{1}{(2\alpha-1)\beta^{\frac{2\alpha-1}{\alpha}}(1-\lambda)} \left(\frac{K-\lambda}{1-\lambda} \right)^{\frac{4\alpha-2}{\alpha}} + \frac{4(K-\lambda)^2}{\alpha^2\beta(1-\lambda)^4} \\
& + \left(\frac{8K^{\frac{\alpha}{2}} - 8 - 8\beta}{\alpha^2\beta(1-\lambda)^2} + \frac{4-4\beta}{\alpha\beta(1-\lambda)} \right) \frac{K-\lambda}{\sqrt{\beta}(1-\lambda)} \\
& + \frac{\alpha(1-\lambda) + 4}{\alpha\beta(1-\lambda)} \frac{K^{2\alpha-1} - 1 - \beta}{(2\alpha-1)\beta(1-\lambda)} \\
& + \frac{4(K^\alpha - 1 - \beta)}{\alpha^2\beta(1-\lambda)^2} + 2 \left(\frac{2K^{\frac{\alpha}{2}} - 2 - 2\beta}{\alpha\beta(1-\lambda)} + \frac{1}{\beta} \right)^2 \\
& - 2 \left(\frac{2K^{\frac{\alpha}{2}} - 2 - 2\beta}{\alpha\beta(1-\lambda)} + \frac{1}{\beta} \right).
\end{aligned}$$

5.3 REMARK. A counterpart of Theorem 5.2 parts (i–iii) for $\frac{1}{2} < \alpha \leq \frac{2}{3}$ and functions s.t. $\|f\|_{V^{\alpha-\frac{1}{2}}} < \infty$ 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 $\eta \leq \alpha$. The following Proposition gives some simple complementary bounds.

5.4 Proposition. *Under Assumptions 2.1 and 5.1,*

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

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

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

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

6. Example

The simulation experiments described below are designed to compare the bounds proved in this paper with actual errors of MCMC estimation. We use a very simple example similar as [JH01, Example 2]. Assume that $y = (y_1, \dots, y_t)$ is an i.i.d. sample from the normal distribution $N(\mu, \kappa^{-1})$, 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 non-informative (improper) prior $p(\mu, \kappa) = p(\mu)p(\kappa) \propto \kappa^{-1}$ (in [JH01] a different prior is considered). The posterior density is

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

where

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

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. Now, the Gibbs sampler consists of drawing samples intermittently from both the conditionals. Start with some (μ_0, κ_0) . Then, for $i = 1, 2, \dots$,

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

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

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

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

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

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

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

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

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

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

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

6.3 Proposition (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*

$$PV(\mu) \leq \begin{cases} \lambda V(\mu) & \text{for } |\mu| > a; \\ K & \text{for } |\mu| \leq a, \end{cases}$$

provided that $a > \sqrt{t/(t-3)}$. The quantities λ and K are given by

$$\lambda = \frac{1}{t-2} \left(\frac{2t-3}{1+a^2} + 1 \right),$$

$$K = 2 + \frac{a^2+2}{t-2}.$$

Moreover,

$$\pi(V) = \frac{2t-3}{t-3}.$$

Proof. It is enough to use the fact that

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

and some simple algebra. Analogously, $\pi(V) = \mathbb{E}_\pi \mu^2 + 1$. \square

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

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

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

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

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

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

$$p_{\min}(\mu) = \begin{cases} p(\mu|a) & \text{if } p(\mu|a) \leq p(\mu|0); \\ p(\mu|0) & \text{if } p(\mu|a) > p(\mu|0). \end{cases}$$

Now it is enough to solve the inequality, say, $p(\mu|a) \leq p(\mu|0)$ with respect to μ . Elementary computation shows that this inequality is fulfilled iff $\mu \leq h(a)$. The formula for β follows from (6.2) and from the fact that

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

□

6.5 REMARK. It is interesting to compare the asymptotic behavior of the constants in Propositions 6.3 and 6.4 for $a \rightarrow \infty$. We can immediately see that $\lambda^2 \rightarrow 1/(t-2)$ and $K^2 \sim a^2/(t-2)$. Slightly more tedious computation reveals that $h(a) \sim \text{const} \cdot a^{1/(t+1)}$ and consequently $\beta \sim \text{const} \cdot a^{-t/(t+1)}$.

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 $\mu_0, \dots, \mu_i, \dots$ is a zero-mean martingale, so $\bar{f} = f$ and

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

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

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

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 1 shows how our bounds on $\sigma_{\text{as}}(P, f)$ 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 $\pi(V^{1/2}) \leq (\pi V)^{1/2}$). 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).

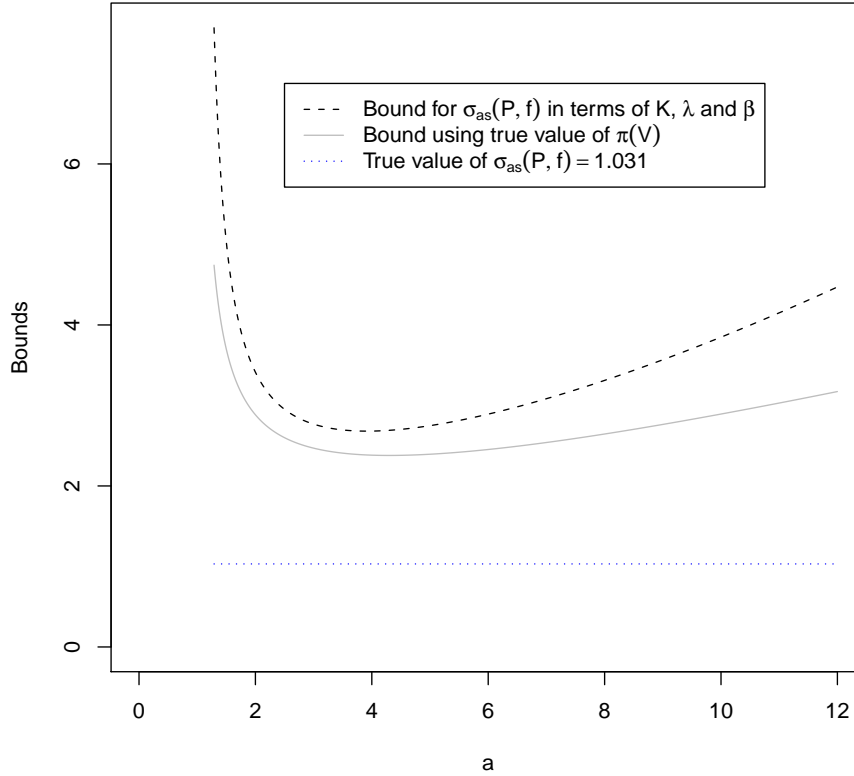


Figure 1. Bounds for the root asymptotic variance $\sigma_{as}(P, f)$ as functions of a .

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 $\sigma_{as}(P, f) = 1.031$. In Table 1 below we summarize the analogous bounds for three values of t .

t	$\sigma_{as}(P, f)$	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 1. Values of $\sigma_{as}(P, f)$ vs. bounds of Theorem 4.2 (ii) combined with Proposition 4.5 (i) and (ii) for different 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 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 $n \geq 10$).

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

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 (3.2) with “true” values of $\sigma_{\text{as}}(P, f)$ and $C_0(P), C_1(P, f), C_2(P, f)$ given by (3.3)-(3.6). Column (b) obtains if we replace those constants by their bounds given in Theorem 4.2 but use the true value of πV . Finally, the bounds involving only λ, K, β are in column (c).

n	\sqrt{n} RMSE	Bound (3.2)		
		(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
10000	1.03	1.04	2.45	2.75
50000	1.03	1.04	2.41	2.71

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

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. [Bax05, RT99, JH04]). However, at present, drift conditions remain the main and most universal tool for proving computable bounds for 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. [Rud09, MN07] and references therein).

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 4.1 is the special case of 5.1, with $\alpha = 1$. Assumption (4.1) implies

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

because by Jensen's inequality $PV^{\frac{1}{2}}(x) \leq \sqrt{PV(x)}$. Whereas for $\alpha < 1$, Lemma 3.5 of [JR02] for all $\eta \leq 1$ yields

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

The following lemma is a well-known fact which appears e.g. in [Num02] (for bounded g). The proof for nonnegative function g is the same.

7.3 Lemma. *If $g \geq 0$ 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 generalized Kac Lemma,

7.4 Lemma (Theorem 10.0.1 of [MT93]). *If $\pi(|f|) < \infty$, then*

$$(7.5) \quad \begin{aligned} \pi(f) &= \int_J \sum_{i=1}^{\tau(J)} f(X_i)\pi(dx), & \text{where} \\ \tau(J) &:= \min\{n > 0 : X_n \in J\}. \end{aligned}$$

The following Lemma is related to other calculations in the drift conditions setting, e.g. [Bax05, LT96, DMR04, Ros02, For03, DGM08].

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

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

7.7 Corollary. *For $\mathbb{E}_x \sum_{n=0}^{T-1} V^{\alpha+\eta-1}(X_n)$ we need to add an additional term $V^{\alpha+\eta-1}(x)$. So from Lemma 7.6 we have*

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

In the case of geometric drift, the second inequality in Lemma 7.6 can be replaced by a slightly better bound. For $\alpha = \eta = 1$, the first inequality in Lemma 7.6 entails the following.

7.8 Corollary. *If Assumptions 2.1 and 4.1 hold then*

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

Proof of Lemma 7.6. The proof is given for $\eta = 1$, because for $\eta < 1$ it is identical and the constants can be obtained from (7.2).

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

$$\begin{aligned} H(x) &:= \mathbb{E}_x \sum_{n=0}^S V^\alpha(X_n), \text{ for } x \in \mathcal{X}, \\ \tilde{H} &:= \sup_{x \in J} \mathbb{E}_x \left(\sum_{n=1}^{S_1} V^\alpha(X_n) \middle| \Gamma_0 = 0 \right) = \sup_{x \in J} \int Q(x, dy) H(y). \end{aligned}$$

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

For $x \notin J$ Assumption 5.1 yields

$$(7.9) \quad V^\alpha(x) \leq \frac{V(x) - PV(x)}{1-\lambda} \leq \frac{V(x) - 1}{1-\lambda}.$$

Since $H(x) = V^\alpha(x)$ for $x \in J$, we have

$$(7.10) \quad V^\alpha(x) \leq \max \left\{ V^\alpha(x), \frac{V(x) - 1}{1 - \lambda} \right\} \leq \frac{V(x) - \lambda}{1 - \lambda} \quad \text{for all } x,$$

and we show by induction that for every $k \geq 0$ and $x \in \mathcal{X}$

$$(7.11) \quad \mathbb{E}_x \sum_{n=0}^{S \wedge k} V^\alpha(X_n) \leq \frac{V(x) - \lambda}{1 - \lambda}.$$

Assume that (7.11) holds for $k - 1$, i.e.

$$\mathbb{E}_x \sum_{n=0}^{S \wedge (k-1)} V^\alpha(X_n) = \mathbb{E}_x \sum_{n=0}^{k-1} V^\alpha(X_n) \mathbb{I}(S \geq n) \leq \frac{V(x) - \lambda}{1 - \lambda}.$$

For $x \in J$ no proof is needed. For $x \notin J$ applying P to both sides yields

$$P \mathbb{E}_x \sum_{n=0}^{k-1} V^\alpha(X_n) \mathbb{I}(S \geq n) = \mathbb{E}_x \sum_{n=1}^k V^\alpha(X_n) \mathbb{I}(S \geq n) \leq \frac{PV(x) - \lambda}{1 - \lambda}.$$

We bound the RHS using Assumption 5.1 to obtain

$$\mathbb{E}_x \sum_{n=1}^k V^\alpha(X_n) \mathbb{I}(S \geq n) \leq \frac{V(x) - \lambda}{1 - \lambda} - V^\alpha(x)$$

and consequently we obtained (7.11) for all k . Hence for all x

$$(7.12) \quad H(x) = \mathbb{E}_x \sum_{n=0}^S V^\alpha(X_n) \leq \frac{V(x) - \lambda}{1 - \lambda}.$$

Next, from Assumption 5.1 we obtain $PV(x) = (1 - \beta)QV(x) + \beta\nu V \leq K$ for $x \in J$, so $QV(x) \leq (K - \beta)/(1 - \beta)$ and, taking into account (7.12),

$$(7.13) \quad \tilde{H} \leq \frac{(K - \beta)/(1 - \beta) - \lambda}{1 - \lambda} = \frac{K - \lambda - \beta(1 - \lambda)}{(1 - \lambda)(1 - \beta)}.$$

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

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

by (7.13). For $x \notin J$ we have to add one more term and note that the above calculation also applies.

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

The extra term is equal to $H(x) - V^\alpha(x)$ and we can use (7.12) to bound it. Finally we obtain

$$(7.14) \quad \mathbb{E}_x \sum_{n=1}^{T-1} V^\alpha(X_n) \leq \frac{V(x) - \lambda - (1-\lambda)V^\alpha(x)}{1-\lambda} \mathbb{I}(x \notin J) + \frac{K-\lambda}{\beta(1-\lambda)} - 1.$$

□

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

(i) for all $\eta \leq \alpha$

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

(ii)

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

(iii) for all $n \geq 0$ and $\eta \leq \alpha$

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

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 (7.5). From the proof of Lemma 7.6

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

And by Lemma 7.4 we obtain

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

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

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

Consequently

$$\begin{aligned} \mathbb{E}_\nu V^\alpha(X_n) &= \int_{\mathcal{X}} P^n V^\alpha(x) \frac{d\nu}{d\pi} \pi(dx) \leq \frac{K-\lambda}{\beta(1-\lambda)} \int_{\mathcal{X}} P^n V^\alpha(x) \pi(dx) \\ &= \frac{K-\lambda}{\beta(1-\lambda)} \pi(V^\alpha), \end{aligned}$$

and (iii) results from (i). \square

8. Proofs for Section 4 and 5

In the proofs for Section 4 we work under Assumption 4.1 and repeatedly use Corollary 7.8.

Proof of Theorem 4.2. (i) Recall that $C_0(P) = \mathbb{E}_\pi T - \frac{1}{2}$, write

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

and use Corollary 7.8. The proof of the alternative statement (i') is the same.

(ii) Without loss of generality we can assume that $\|\bar{f}\|_{V^{\frac{1}{2}}} = 1$. By Lemma 7.3 we then have

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

To bound the second term we will use Corollary 7.8 with $V^{\frac{1}{2}}$ in place of V , which is legitimate because of (7.1).

$$\begin{aligned} \text{II}/2 &= \mathbb{E}_\pi \sum_{n=1}^{T-1} V^{\frac{1}{2}}(X_0) V^{\frac{1}{2}}(X_n) = \mathbb{E}_\pi V^{\frac{1}{2}}(X_0) \mathbb{E} \left(\sum_{n=1}^{T-1} V^{\frac{1}{2}}(X_n) \mid X_0 \right) \\ &\leq \mathbb{E}_\pi V^{\frac{1}{2}}(X_0) \left(\frac{\lambda^{\frac{1}{2}}}{1-\lambda^{\frac{1}{2}}} V^{\frac{1}{2}}(X_0) + \frac{K^{\frac{1}{2}} - \lambda^{\frac{1}{2}} - \beta}{\beta(1-\lambda^{\frac{1}{2}})} \right) \\ &= \frac{\lambda^{\frac{1}{2}}}{1-\lambda^{\frac{1}{2}}} \pi(V) + \frac{K^{\frac{1}{2}} - \lambda^{\frac{1}{2}} - \beta}{\beta(1-\lambda^{\frac{1}{2}})} \pi(V^{\frac{1}{2}}). \end{aligned}$$

Rearranging terms in I + II, we obtain

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

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 7.3. First write

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

The first term can be bounded directly using Corollary 7.8 applied to V .

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

To bound the second term, first condition on X_n and apply Corollary 7.8 to $V^{\frac{1}{2}}$, then again apply this corollary to V and to $V^{\frac{1}{2}}$.

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

Finally, rearranging terms in I + II, we obtain

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

which is tantamount to the desired result.

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

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

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

(v) We compute:

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

□

In the proofs for Section 5 we work under Assumption 5.1 and repeatedly use Lemma 7.6 or Corollary 7.7.

Proof of Theorem 5.2. (i) Recall that $C_0(P) = \mathbb{E}_\pi T - \frac{1}{2}$ and write

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

From Lemma 7.6 with V , α and $\eta = \alpha$ we have

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

(ii) Without loss of generality we can assume that $\|\bar{f}\|_{V^{\frac{3}{2}\alpha-1}} = 1$. By Lemma 7.3 we have

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

To bound the second term we will use Lemma 7.6 with V , α and $\eta = \frac{\alpha}{2}$.

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

The proof of (ii) is complete.

(iii) The proof is similar to that of (ii) but more delicate, because we now cannot use Lemma 7.3. Write

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

The first term can be bounded directly using Lemma 7.6 with $\eta = 2\alpha - 1$

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

To bound the second term, first condition on X_n and use Lemma 7.6 with $\eta = \frac{\alpha}{2}$ then again use Lemma 7.6 with $\eta = \alpha$ and $\eta = \frac{\alpha}{2}$.

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

So after gathering the terms

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

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

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

Since

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

and $|\bar{f}| \leq V^{\frac{3}{2}\alpha-1}$ we put (8.1) into (8.2) and apply Lemma 7.15 to complete the proof. \square

Proof of Proposition 5.4. (i) See Lemma 7.15.

(ii) We compute:

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

\square

Acknowledgements

We thank Gersende Fort and Jacek Wesolowski for insightful comments on earlier versions of this work. We also gratefully acknowledge the help of Agnieszka Perduta.

References

- [Ada08] R. Adamczak. A tail inequality for suprema of unbounded empirical processes with applications to markov chains. *Electronic Journal of Probability*, 13:1000–1034, 2008.
- [Ald87] D. Aldous. On the Markov chain simulation method for uniform combinatorial distributions and simulated annealing. *Probability in the Engineering and Informational Sciences*, 1(01):33–46, 1987.
- [AN78] K.B. Athreya and P. Ney. A new approach to the limit theory of recurrent Markov chains. *Transactions of the American Mathematical Society*, 245(Nov):493–501, 1978.
- [Bax05] P.H. Baxendale. Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *The Annals of Applied Probability*, 15(1B):700–738, 2005.
- [BC10] P. Bertail and S. Cléménçon. Sharp bounds for the tail of functionals of Markov chains. *Theory of Probability and its Applications*, 54(3):1–19, 2010.
- [BLL08] W. Bednorz, K. Latuszyński, and R. Latała. A regeneration proof of the central limit theorem for uniformly ergodic markov chains. *Electronic Communications in Probability*, 13:85–98, 2008.
- [Clé01] SJM Cléménçon. Moment and probability inequalities for sums of bounded additive functionals of regular Markov chains via the Nummelin splitting technique. *Statistics & probability letters*, 55(3):227–238, 2001.
- [DFMS04] R. Douc, G. Fort, E. Moulines, and P. Soulier. Practical drift conditions for subgeometric rates of convergence. *Annals of Applied Probability*, pages 1353–1377, 2004.
- [DGM08] R. Douc, A. Guillin, and E. Moulines. Bounds on regeneration times and limit theorems for subgeometric Markov chains. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 44(2):239–257, 2008.
- [DMR04] R. Douc, E. Moulines, and J.S. Rosenthal. Quantitative bounds on convergence of time-inhomogeneous Markov Chains. *Annals of Applied Probability*, pages 1643–1665, 2004.
- [DMS07] Randal Douc, Eric Moulines, and Philippe Soulier. Computable convergence rates for sub-geometric ergodic Markov chains. *Bernoulli*, 13(3):831–848, 2007.
- [FM00] G. Fort and E. Moulines. V-subgeometric ergodicity for a Hastings-Metropolis algorithm. *Statistics & probability letters*, 49(4):401–410, 2000.

- [FM03a] G. Fort and E. Moulines. Convergence of the Monte Carlo expectation maximization for curved exponential families. *The Annals of Statistics*, 31(4):1220–1259, 2003.
- [FM03b] G. Fort and E. Moulines. Polynomial ergodicity of Markov transition kernels. *Stochastic Process. Appl.*, 103(1):57–99, 2003.
- [FMRR03] G. Fort, E. Moulines, GO Roberts, and JS Rosenthal. On the geometric ergodicity of hybrid samplers. *Journal of Applied Probability*, pages 123–146, 2003.
- [For03] G. Fort. Computable bounds for V-geometric ergodicity of Markov transition kernels. *Preprint, Université Joseph Fourier, Grenoble, France*, 2003.
- [Gil98] D. Gillman. A Chernoff bound for random walks on expander graphs. *SIAM Journal on Computing*, 27:1203–1220, 1998.
- [GO02] P.W. Glynn and D. Ormoneit. Hoeffding’s inequality for uniformly ergodic Markov chains. *Statistics & probability letters*, 56(2):143–146, 2002.
- [HG98] J.P. Hobert and C.J. Geyer. Geometric ergodicity of Gibbs and block Gibbs samplers for a hierarchical random effects model. *Journal of Multivariate Analysis*, 67(2):414–430, 1998.
- [JH01] G.L. Jones and J.P. Hobert. Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statistical Science*, pages 312–334, 2001.
- [JH04] G.L. Jones and J.P. Hobert. Sufficient burn-in for Gibbs samplers for a hierarchical random effects model. *The Annals of Statistics*, 32(2):784–817, 2004.
- [JHCN06] G.L. Jones, M. Haran, B.S. Caffo, and R. Neath. Fixed-width output analysis for Markov chain Monte Carlo. *Journal of the American Statistical Association*, 101(476):1537–1547, 2006.
- [JJ10] A.A. Johnson and G.L. Jones. Gibbs sampling for a Bayesian hierarchical general linear model. *Electronic Journal of Statistics*, 4:313–333, 2010.
- [JO10] A. Joulin and Y. Ollivier. Curvature, concentration and error estimates for Markov chain Monte Carlo. *The Annals of Probability*, 38(6):2418–2442, 2010.
- [JR02] S.F. Jarner and G.O. Roberts. Polynomial convergence rates of Markov chains. *The Annals of Applied Probability*, 12(1):224–247, 2002.
- [JR07] Søren F. Jarner and Gareth O. Roberts. Convergence of heavy-tailed Monte Carlo Markov chain algorithms. *Scand. J. Statist.*, 34(4):781–815, 2007.
- [JT03] S.F. Jarner and R.L. Tweedie. Necessary conditions for geometric and polynomial ergodicity of random-walk-type Markov chains. *Bernoulli*, 9(4):559–578, 2003.
- [JVV86] M.R. Jerrum, L.G. Valiant, and V.V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Science*, 43:169–188, 1986.

- [KLMM05] I. Kontoyiannis, L.A. Lastras-Montano, and S.P. Meyn. Relative entropy and exponential deviation bounds for general Markov chains. In *IEEE, International Symposium on Information Theory*, pages 1563–1567. IEEE, 2005.
- [KR08] L.A. Kontorovich and K. Ramanan. Concentration inequalities for dependent random variables via the martingale method. *The Annals of Probability*, 36(6):2126–2158, 2008.
- [LMN11] K. Latuszyński, B. Miasojedow, and W. Niemiro. Nonasymptotic bounds on the mean square error for MCMC estimates via renewal techniques. *to appear, MCQMC 2010 Proceedings, Arxiv preprint arXiv:1101.5837*, 2011.
- [Lor70] G. Lorden. On excess over the boundary. *The Annals of Mathematical Statistics*, 41(2):520–527, 1970.
- [LP04] C.A. León and F. Perron. Optimal Hoeffding bounds for discrete reversible Markov chains. *Annals of Applied Probability*, pages 958–970, 2004.
- [LT96] R.B. Lund and R.L. Tweedie. Geometric convergence rates for stochastically ordered Markov chains. *Mathematics of operations research*, pages 182–194, 1996.
- [Mar96] K. Marton. A measure concentration inequality for contracting Markov chains. *Geometric and Functional Analysis*, 6(3):556–571, 1996.
- [MN07] P. Mathé and E. Novak. Simple Monte Carlo and the metropolis algorithm. *Journal of Complexity*, 23(4-6):673–696, 2007.
- [MT93] S.P. Meyn and R.L Tweedie. *Markov chains and stochastic stability*. Springer London et al., 1993.
- [MT94] Sean P. Meyn and R. L. Tweedie. Computable bounds for geometric convergence rates of Markov chains. *Ann. Appl. Probab.*, 4(4):981–1011, 1994.
- [NP09] W. Niemiro and P. Pokarowski. Fixed precision MCMC estimation by median of products of averages. *Journal of Applied Probability*, 46(2):309–329, 2009.
- [Num78] E. Nummelin. A splitting technique for Harris recurrent Markov chains. *Probability Theory and Related Fields*, 43(4):309–318, 1978.
- [Num02] E. Nummelin. MC’s for MCMC’ists. *International Statistical Review*, 70(2):215–240, 2002.
- [RH10] V. Roy and J.P. Hobert. On Monte Carlo methods for Bayesian multivariate regression models with heavy-tailed errors. *Journal of Multivariate Analysis*, 101(5):1190–1202, 2010.
- [Ros95a] J.S. Rosenthal. Minorization Conditions and Convergence Rates for Markov Chain Monte Carlo. *Journal of the American Statistical Association*, 90(430):558–566, 1995.
- [Ros95b] J.S. Rosenthal. Rates of convergence for Gibbs sampling for variance component models. *The Annals of Statistics*, 23(3):740–761, 1995.
- [Ros02] J.S. Rosenthal. Quantitative convergence rates of Markov chains: A simple account. *Elec. Comm. Probab*, 7(13):123–128, 2002.

- [RR01] G.O. Roberts and J.S. Rosenthal. Small and pseudo-small sets for Markov chains. *Stochastic Models*, 17(2):121–145, 2001.
- [RR04] G.O. Roberts and J.S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probability Surveys*, 1:20–71, 2004.
- [RR11] G.O. Roberts and J.S. Rosenthal. Quantitative Non-Geometric Convergence Bounds for Independence Samplers. *Methodology and Computing in Applied Probability*, 13:391–403, 2011.
- [RT99] G.O. Roberts and R.L. Tweedie. Bounds on regeneration times and convergence rates for Markov chains. *Stochastic Processes and Their Applications*, 80(2):211–230, 1999.
- [Rud09] D. Rudolf. Explicit error bounds for lazy reversible Markov chain Monte Carlo. *Journal of Complexity*, 25(1):11–24, 2009.
- [Sam00] P.M. Samson. Concentration of measure inequalities for Markov chains and Φ -mixing processes. *The Annals of Probability*, 28(1):416–461, 2000.