ERROR BOUNDS OF MCMC FOR FUNCTIONS WITH UNBOUNDED STATIONARY VARIANCE

DANIEL RUDOLF AND NIKOLAUS SCHWEIZER

ABSTRACT. We prove explicit error bounds for Markov chain Monte Carlo (MCMC) methods to compute expectations of functions with unbounded stationary variance. We assume that there is a $p \in (1, 2)$ so that the functions have finite $L^p$-norm. For uniformly ergodic Markov chains we obtain error bounds with the optimal order of convergence $n^{1/p - 1}$ and if there exists a spectral gap we almost get the optimal order. Further, a burn-in period is taken into account and a recipe for choosing the burn-in is provided.

1. Introduction

Let $G \subseteq \mathbb{R}^d$ and let $\mathcal{B}(G)$ be the corresponding Borel $\sigma$-algebra. We study the problem of computing an expectation of a measurable function, say $f : G \to \mathbb{R}$, with respect to a probability measure $\pi$. Thus we want to know

$$E_{\pi}(f) = \int_G f(x) \, d\pi(x).$$

We assume that the variance $(E_{\pi}(f^2)) - (E_{\pi}(f))^2$ is not finite, but that there is a $p \in (1, 2)$ such that

$$\|f\|_p = \left( \int_G |f(x)|^p \, d\pi(x) \right)^{1/p} < \infty.$$ 

This is the case if $f$ has a singularity, e.g. $f(x) = |x|^{-1/2}$ and $\pi$ has a bounded strictly positive density over a compact convex set $G$ with $0 \in G$. Here, $|\cdot|$ denotes the Euclidean norm in $\mathbb{R}^d$. Another application is the computation of a second moment of $\pi$, say $f(x) = x^2$, when the fourth moment of $\pi$ is infinite due to heavy tails.

Our focus is on situations where $\pi$ and $f$ are complicated and thus Monte Carlo algorithms are applied. Often one does not have an i.i.d. sequence of random variables with distribution $\pi$. For instance, this is the case if $\pi$ is only known up to a normalizing constant. Such situations naturally arise in Bayesian Statistics and Statistical Physics. Markov Chain Monte Carlo (MCMC) methods are a popular approach for overcoming this problem. The objective of this paper is to provide explicit error bounds of MCMC methods for the computation of $E_{\pi}(f)$ when $E_{\pi}(f^2) = \|f\|_2^2$ is not finite. So far, the literature has largely focused on cases where $\|f\|_2$ is finite, see [Rud09, BC09, JO10, LNT1, Rud12, LMN13].

The basic idea of MCMC is to approximate $\pi$ using a Markov chain $(X_n)_{n \in \mathbb{N}}$ with transition kernel $K$ and initial distribution $\nu$. Here $\pi$ is the stationary and

---

D.R. was supported by the DFG-Priority Program 1324 and the DFG Research Training Group 1523. N.S. was supported by the DFG-Priority Program 1324.
limit distribution. Then, one approximates \( E_\pi(f) \) by
\[
S_{n,n_0}(f) = \frac{1}{n} \sum_{j=1}^{n} f(X_{j+n_0}),
\]
where \( n \) denotes the number of function evaluations and \( n_0 \) the burn-in. The burn-in is the number of steps needed to get sufficiently close to \( \pi \). For simplicity we assume that the Markov chain is reversible with respect to \( \pi \). To quantify convergence, we need further assumptions on the Markov chain. We begin with assuming that the Markov chain is uniformly ergodic. Then we relax this assumption to the existence of an (absolute) spectral gap. For reversible Markov chains, existence of such a spectral gap is equivalent to geometric ergodicity, see \[RR97\]. For \( p \in (1,2] \) we prove bounds on the absolute mean error
\[
e_1(S_{n,n_0}, f) = \mathbb{E} |S_{n,n_0}(f) - E_\pi(f)|,
\]
for functions \( f \), with \( \|f\|_p < \infty \). Here, \( \mathbb{E} \) denotes the expectation with respect to the distribution of the trajectory of the Markov chain. We consider this absolute error criterion since the root mean square error of \( S_{n,n_0} \) is not necessarily finite for \( p \in (1,2) \).

It is known that
\[
\inf_{A_n} \sup_{\|f\|_p \leq 1} e_1(A_n,f) \geq c \cdot n^{1/p-1},
\]
for some number \( c > 0 \), where the infimum is taken over all algorithms \( A_n \) that use at most \( n \) function evaluations of \( f \). For a proof of this fact follow the arguments in \[Nov88, Hei94\]. Therefore, our bounds cannot decay faster than \( n^{1/p-1} \), the optimal order of convergence.

Our absolute mean error bounds satisfy the following properties:

- For uniformly ergodic Markov chains we obtain the optimal order of convergence \( n^{1/p-1} \). For Markov chains with spectral gap we come arbitrarily close to the optimal order.
- We quantify the penalty that arises since the initial distribution \( \nu \) is not the stationary distribution. This penalty appears in our bounds through \( \log \|d\nu/d\pi\|_\infty \), where \( d\nu/d\pi \) denotes the density of \( \nu \) with respect to \( \pi \). This effect is controlled by the choice of the burn-in \( n_0 \). We provide a recipe for choosing \( n_0 \).

The main idea of proof is simple and adapted from \[Hei94\] Proposition 5.4. The key technique is to apply the interpolation theorem of Riesz-Thorin to the absolute mean error and root mean square error, viewed as operators.

1.1. Literature overview. By an ergodic theorem see \[MT09, Theorem 17.1.7, p. 427\] or \[AG11\], the MCMC method is well defined, i.e. for a \( \varphi \)-irreducible and Harris recurrent Markov chain
\[
\lim_{n \rightarrow \infty} S_{n,n_0}(f) = E_\pi(f)
\]
holds almost surely for any \( f \), with finite \( \|f\|_1 \). For \( p \in [2,\infty) \), the ergodic theorem is usually augmented by a central limit theorem, for a survey see \[Jon04\] and for estimating with confidence see \[FJ11\]. The corresponding limit theorems for \( p \in [0,2) \) have been less studied, see \[CMA13\] for recent results and an overview.

Under different convergence assumptions on the Markov chain, various bounds on the mean square error of \( S_{n,n_0} \) are known, see for example \[Rud09, BC09, JO10\].
In particular, for \( p \in [2, \infty] \) with finite \( \|f\|_p \) in [Rud12] Theorem 3.34 and Theorem 3.41] explicit error bounds of \( S_{n,n_0} \) are provided. However, no explicit error bounds are known for \( p \in (1, 2) \).

In [Hei94, Proposition 5.4] a simple Monte Carlo method where \( (X_n)_{n \in \mathbb{N}} \) is an i.i.d. sequence and \( X_n \) has distribution \( \pi \) is studied. A bound of the mean absolute error is provided, namely, for \( p \in (1, 2] \) holds

\[
e_1(S_{n,0}, f) \leq 2^{2/p-1} n^{1/p-1} \|f\|_p.
\]

A similar estimate follows by an inequality of v. Bahr and Esseen [vBE65, Theorem 2]. Their bounds work also for some dependent sequences of random variables, but their condition (2) is not satisfied for most Markov chains used for MCMC. Finally let us mention that a slight (nonlinear) modification of the simple Monte Carlo method does also give the optimal rate of the mean square error, for a discussion see [HN02].

1.2. Outline. The paper is organized as follows: In the next section we state and discuss our two main results. In Section 3 we provide two applications, the independent Metropolis-Hastings algorithm and the hit-and-run algorithm. Finally, Section 4 contains the proofs. For the reader’s convenience the theorem of Riesz-Thorin is stated in the appendix.

2. Main results

By \( L_2 = L_2(\pi) \) we denote the set of all square integrable functions with respect to \( \pi \). For \( f \in L_2 \) note that the transition kernel \( K \) induces the Markov operator

\[
Pf(x) = \int_G f(y) K(x, dy).
\]

By

\[
gap(P) = 1 - \|P - \mathbb{E}_\pi\|_{L_2 \to L_2}
\]

we denote the spectral gap of the Markov operator. Recall that the transition kernel is reversible with respect to \( \pi \).

Now we state the error bound for uniformly ergodic Markov chains.

**Theorem 1.** Let us assume that we have a Markov chain with transition kernel \( K \) and initial distribution \( \nu \). Let \( K \) be uniformly ergodic, i.e. for an \( \alpha \in [0, 1) \) and an \( M \in (0, \infty) \) it holds for \( \pi \)-almost all \( x \in G \) that

\[
\|K^n(x, \cdot) - \pi\|_{TV} \leq \alpha^n M,
\]

where \( \|\cdot\|_{TV} \) denotes the total variation distance. Further, assume that there exists \( \frac{d\nu}{d\pi} \) with finite \( \|\frac{d\nu}{d\pi}\|_{\infty} \), where \( \frac{d\nu}{d\pi} \) denotes the density of \( \nu \) with respect to \( \pi \). Let \( p \in (1, 2] \) and assume that \( n_0 \in \mathbb{N}_0 \) satisfies

\[
n_0 \geq \frac{\log(2 M \|\frac{d\nu}{d\pi} - 1\|_{\infty})}{1 - \alpha}.
\]

Then

\[
e_1(S_{n,n_0}, f) \leq \frac{4 \|f\|_p}{(n \cdot \gap(P))^{1-1/p}} + \frac{4 \|f\|_p}{(n \cdot (1 - \alpha))^{2-2/p}}.
\]
First, note that uniform ergodicity implies $\text{gap}(P) \geq 1 - \alpha$. The upper bound might be interpreted as follows: The burn-in $n_0$ is used to decrease the influence of the initial distribution. The number $n$ decreases the error of the averaging procedure. The leading term has the optimal order of convergence $n^{1/p - 1}$ as in (2). The spectral gap appears in the leading term and $1 - \alpha$ appears in the higher order term. Both quantities describe the price we have to pay for using Markov chains for approximate sampling. If one can sample with respect to the initial distribution, then $\alpha = 0$, $\text{gap}(P) = 1$ and $\nu = \pi$. Thus $n_0 = 0$ and the error bound is, up to a constant factor, the same as in (2).

Now we state the result for Markov chains which have a spectral gap.

**Theorem 2.** Let us assume that we have a Markov chain with transition kernel $K$ and initial distribution $\nu$. Let $\text{gap}(P) > 0$ and further assume that there exists $\frac{d\nu}{d\pi}$ with finite $\|\frac{d\nu}{d\pi}\|_\infty$, where $\frac{d\nu}{d\pi}$ denotes the density of $\nu$ with respect to $\pi$. Let $\delta \in (0, 1], p \in (1 + \delta, 2]$ and assume that $n_0 \in \mathbb{N}_0$ satisfies

$$n_0 \geq \frac{\log(64 \|\frac{d\nu}{d\pi} - 1\|_\infty) + \log \delta^{-1}}{\delta \cdot \text{gap}(P)}.$$

Then

$$e_1(S_{n,n_0}, f) \leq \frac{4 \|f\|_p}{(n \cdot \text{gap}(P))^{1 - \frac{1}{2p}}} + \frac{4 \|f\|_p}{(n \cdot \text{gap}(P))^{2 - \frac{2(1 + \delta)}{p}}}.$$

Let us interpret the result. The burn-in $n_0$ is used to decrease the dependence on the initial distribution and $n$ denotes the sample size of the average procedure. The convergence of the Markov chain is captured by the spectral gap. However, an additional parameter $\delta \in (0, 1]$ appears. This parameter measures a minimal integrability and provides a relation between integrability and convergence of the Markov chain. If a bound on the root mean square error could be derived from $\text{gap}(P) > 0$ for $f \in L_2$, then we would also obtain a bound on the absolute mean error for $\delta = 0$.

For $\delta$ close to zero, the rate of convergence in the error bound is arbitrarily close to optimal. But we pay a price. Namely, the burn-in $n_0$ increases for decreasing $\delta$. There is obviously a trade-off and one might ask for an optimal $\delta$. After some computations by hand one can guess that

$$\hat{\delta} = \frac{\sqrt{p - 1}}{\sqrt{p}} \left( \frac{\log(64 \|\frac{d\nu}{d\pi} - 1\|_\infty)}{(8\epsilon^{-1})^{p/(p-1)} \log(8\epsilon^{-1})} \right)^{1/2},$$

is a good choice for $\delta$ to achieve an error estimate smaller than $\epsilon$. We justify this heuristic $\hat{\delta}$ as follows: For different values of $p$ and $\epsilon$ we numerically compute $\delta^*$ which minimizes the total size of the Markov chain sample $N(\delta) = n(\delta) + n_0(\delta)$ which is needed to obtain an estimate with error $\epsilon$ from Theorem 2. In Table 1 and Table 2 one can see that $\hat{\delta}$ and $\delta^*$ have the same behavior for decreasing $\epsilon \in (0, 1]$ and decreasing $p \in (1, 2]$. Furthermore, the numbers $N(\hat{\delta})$ and $N(\delta^*)$ are quite close. For $p$ close to 1 (Table 1) we also see that the penalty for the lack of integrability leads to a drastic increase in $N(\delta^*)$. This is not surprising, since for $p$ close to 1 also Theorem 1 and 2 exhibit similar behavior. However, for $p$ not too far away from 2, the total size of the Markov chain sample $N(\delta^*)$ and $N(\hat{\delta})$ is reasonable.
Table 1. Required size of the Markov chain sample $N(\delta)$ as prescribed by Theorem 2 for different values of $p$ and $\delta \in \{\delta^*, \hat{\delta}\}$. The parameters are $\|f\|_p = 1$, gap$(P) = 0.01$, $\|\frac{d\nu}{d\pi} - 1\|_{tv} = 10^{30}$ and $\varepsilon = 0.1$.

| $p$   | $\delta^*$ | $N(\delta^*)$ | $\hat{\delta}$ | $N(\hat{\delta})$ |
|-------|-------------|----------------|------------------|---------------------|
| 1.1   | $6.54 \cdot 10^{-10}$ | $5.47 \cdot 10^{19}$ | $4.21 \cdot 10^{-11}$ | $5.47 \cdot 10^{19}$ |
| 1.3   | $3.78 \cdot 10^{-4}$   | $1.01 \cdot 10^{9}$   | $1.48 \cdot 10^{-4}$   | $1.03 \cdot 10^{9}$   |
| 1.5   | $6.58 \cdot 10^{-3}$   | $9.17 \cdot 10^{6}$   | $3.30 \cdot 10^{-3}$   | $9.80 \cdot 10^{6}$   |

Table 2. Required size of the Markov chain sample $N(\delta)$ as prescribed by Theorem 2 for different values of $\varepsilon$ and $\delta \in \{\delta^*, \hat{\delta}\}$. The parameters are $\|f\|_p = 1$, gap$(P) = 0.01$, $\|\frac{d\nu}{d\pi} - 1\|_{tv} = 10^{30}$ and $p = 1.3$.

| $\varepsilon$ | $\delta^*$ | $N(\delta^*)$ | $\hat{\delta}$ | $N(\hat{\delta})$ |
|---------------|------------|----------------|-----------------|------------------|
| 0.01          | $1.97 \cdot 10^{-6}$ | $1.91 \cdot 10^{13}$ | $8.15 \cdot 10^{-7}$ | $1.91 \cdot 10^{13}$ |
| 0.2           | $1.78 \cdot 10^{-3}$   | $6.20 \cdot 10^{7}$   | $7.23 \cdot 10^{-4}$   | $6.59 \cdot 10^{7}$   |
| 0.5           | $1.11 \cdot 10^{-2}$   | $2.57 \cdot 10^{6}$   | $6.08 \cdot 10^{-3}$   | $2.87 \cdot 10^{6}$   |

3. Applications

Now we illustrate our error bounds in two settings. We apply Theorem 1 to the independent Metropolis-Hastings algorithm and Theorem 2 to the hit-and-run algorithm.

3.1. Independent Metropolis-Hastings algorithm. Let $\pi$ be a probability measure with density $\rho$. We use the independent Metropolis-Hastings algorithm for approximate sampling with respect to $\pi$. Let $\nu$ be the initial and proposal distribution and denote its density by $q$. Then, for $x \in G$ and $A \in B(G)$ the transition kernel is given by

$$M_{\rho,q}(x,A) = \int_A \alpha(x,y)q(y)\,dy + 1_A(x) \left( 1 - \int_G \alpha(x,y)q(y)\,dy \right),$$

with $\alpha(x,y) = \min\left\{1, \frac{\rho(y)\alpha(x)}{\rho(x)\alpha(y)}\right\}$. It is well known that $M_{\rho,q}$ is reversible with respect to $\pi$. For $\gamma \in (0,1]$ and $\kappa \geq \gamma$ let us define

$$R_{\gamma,\kappa} = \left\{ \rho: G \to (0,\infty) \mid \gamma \leq \frac{\rho(x)}{\rho(y)} \leq \kappa \right\}.$$ 

By [MT96, Theorem 2.1] we have for $\rho \in R_{\gamma,\kappa}$ a criterion for uniform ergodicity, i.e.

$$\|M_{\rho,q}^n(x,\cdot) - \pi\|_{tv} \leq (1 - \gamma)^n. \quad (6)$$

This leads to the following application of Theorem 1.
Corollary 3. For $\rho \in \mathbb{R}$, let $(X_n)_{n \in \mathbb{N}}$ be a Markov chain with transition kernel $M_{\rho,q}$ and initial distribution $\nu$. Let $p \in (1,2]$ and assume that $n_0 \in \mathbb{N}$ satisfies

$$n_0 \geq \gamma^{-1} \log(2\kappa).$$

Then

$$e_1(S_{n,n_0}, f) \leq 4 \frac{\|f\|_p}{(n\gamma)^{1-1/p}} + 4 \frac{\|f\|_p}{(n\gamma)^{2-2/p}}.$$

Proof. By (6) the spectral gap is bounded from below by $\gamma$. Together with $\frac{d\nu}{d\pi}(x) = q(x) \rho(x) \leq \kappa$ this implies the error bound. 

The result shows that the Metropolis-Hastings algorithm, in a suitable setting, can be used to compute expectations of functions which do not necessarily have finite stationary variance.

3.2. Hit-and-run algorithm. Let $G \subset \mathbb{R}^d$ be a measurable set with $0 < \text{vol}_d(G) < \infty$, where $\text{vol}_d$ denotes the Lebesgue measure. We assume that $G$ is given by a membership oracle, i.e. we are able to evaluate $1_G(x)$ for any $x \in \mathbb{R}^d$. The goal is to compute

$$E_G(f) = \frac{1}{\text{vol}_d(G)} \int_G f(x)dx$$

for a function $f: G \to \mathbb{R}$ with finite $\|f\|_p$ for $p \in (1,2)$. We impose some additional structure on possible sets $G$. Namely, let

$$G_{r,d} = \{G \subset \mathbb{R}^d \mid G \text{ is convex, } B_d \subset G \subset rB_d\},$$

where $r \geq 1$ and $rB_d = \{x \in \mathbb{R}^d \mid |x| \leq r\}$ is the Euclidean ball with radius $r$. Note that here the input of an approximation scheme is given by a tuple $(f,G)$ with $G \in G_{r,d}$. The hit-and-run algorithm gives a suitable tool to construct a Markov chain to approximate a sample with uniform distribution in $G \in G_{r,d}$.

A transition of the hit-and-run algorithm from $x \in G$ works as follows:

1. Choose a direction, say $\theta$, uniformly distributed on the sphere $\partial B_d$.
2. Choose the next state, say $y \in G$, uniformly distributed in $G \cap \{x + \theta r : r \in \mathbb{R}\}$.

After randomly choosing a direction $\theta$ one samples the next state $y \in G$ uniformly on the line segment determined by the current state $x$ and the direction $\theta$. The resulting hit-and-run algorithm is reversible with respect to the uniform distribution on $G$.

It is known that the hit-and-run algorithm induces a positive semidefinite Markov operator (on $L_2$), say $H$, see [RU13]. Lovász and Vempala prove in [LV06, Theorem 4.2, p. 993] a lower bound on the conductance, which implies by the positive semidefiniteness and by the Cheeger inequality, see [LS88], a lower bound on the spectral gap.

Proposition 4. Let $G \in G_{r,d}$. Then, the Markov operator $H$ of the hit-and-run algorithm satisfies $\text{gap}(H) \geq 2^{-51}(dr)^{-2}$.

Now we can apply Theorem 2.
Corollary 5. Let $\nu$ be the uniform distribution on $B_d$. Let $G \in \mathcal{G}_{r, d}$ and assume that we have a Markov chain with transition kernel, given by the hit-and-run algorithm, and initial distribution $\nu$. Let $\delta \in (0, 1]$, $p \in [1 + \delta, 2]$ and assume that $n_0 \in \mathbb{N}_0$ satisfies

$$n_0 \geq 2.3 \cdot 10^{15} (dr)^2 \delta^{-1} (d \log r + \log \delta^{-1} + 4.2).$$

Then

$$e_1(S_{n,n_0}, (f, G)) \leq \frac{4 \|f\|_p (6.8 \cdot 10^7 dr)^{2(1 - \frac{1 + \delta}{p})}}{n^{1 - \frac{2 + \delta}{p}}} + \frac{4 \|f\|_p (6.8 \cdot 10^7 dr)^{4(1 - \frac{1 + \delta}{p})}}{n^{2 - \frac{2 + \delta}{p}}}.$$

Note that the constants in the error bound are rather large. Thus the bound is not useful for simulations. However, it shows that the error behaves only polynomially with respect to the dimension $d$.

4. Auxiliary results and proofs

Let $(X_n)_{n \in \mathbb{N}}$ be a Markov chain with transition kernel $K$ and initial distribution $\nu$. We assume that the transition kernel is reversible with respect to $\pi$. Thus $\pi$ is a stationary distribution.

By (4) the transition kernel induces the Markov operator acting on signed measures $\nu$ on $(G, \mathcal{B}(G))$. If $\nu$ is absolutely continuous with respect to $\pi$, then also $\nu P$ is absolutely continuous with respect to $\pi$. In particular, $\frac{d(\nu P)}{dx} = P(\frac{d\nu}{dx})$, for details we refer to [Rud12, Lemma 3.9].

For $p \in [1, \infty]$ we denote by $L_p = L_p(\pi)$ the space of all functions $f : G \to \mathbb{R}$ with $\|f\|_p < \infty$. Note that $\|P\|_{L_p \to L_p} = 1$ and that $P : L_2 \to L_2$ is self-adjoint (because of reversibility).

Next we provide some relations between different convergence properties of reversible Markov chains. If the Markov chain is uniformly ergodic, i.e., if it satisfies (1) for an $\alpha \in [0, 1)$ and an $M \in (0, \infty)$, then

$$\|P^n - \mathbb{E}_\pi\|_{L_p \to L_p} \leq \alpha^n 2M, \quad n \in \mathbb{N},$$

for $p \in [1, \infty]$. This implies $\text{gap}(P) \geq 1 - \alpha$. For more details we refer to [Rud12, Proposition 3.24]. Further, by the Riesz-Thorin theorem (see Proposition 11) it follows for $p \in [2, \infty)$ that $\text{gap}(P) > 0$ implies

$$\|P^n - \mathbb{E}_\pi\|_{L_p \to L_p} \leq 2^{1 - \frac{1}{p}} \|P - \mathbb{E}_\pi\|_{L_2 \to L_2}^{2n/p} = 2^{1 - \frac{1}{p}} (1 - \text{gap}(P))^{2n/p}, \quad n \in \mathbb{N}.$$

Now we define a generalized error term of $S_{n,n_0}$ with parameter $p \in [1, 2]$ for the computation of $\mathbb{E}_\pi(f)$. Let

$$e_p(S_{n,n_0}, f) := (\mathbb{E}|S_{n,n_0}(f) - \mathbb{E}_\pi(f)|^p)^{1/p}.$$

Note that for $p = 1$ this is the absolute mean error and for $p = 2$ we have the root mean square error. The expectation in the definition of the error is taken with respect to the distribution, say $\mu_{\nu,K}$, of the trajectory $X_1, \ldots, X_{n+n_0}$.
4.1. Proof of Theorem \[ \text{[1]} \] We prove that under the assumptions of Theorem \[ \text{[1]} \] the inequality
\[
\sup_{\|f\| \leq 1} e_p(S_{n,n_0}, f) \leq 2^{2/p-1} \left( 1 + 2 \alpha^{n_0} M \left\| \frac{d\nu}{d\pi} - 1 \right\|_\infty \right)^{2/p-1} \\
\times \left( \frac{2^{1-1/p}}{(n \cdot \text{gap}(P))^{1-1/p}} + \left( \frac{4 M \left\| \frac{d\nu}{d\pi} - 1 \right\|_\infty \alpha^{n_0}}{n^2(1-\alpha)^2} \right)^{1-1/p} \right)
\] holds. From this upper bound the assertion of the theorem follows immediately by \[ \log \alpha^{-1} \geq 1 - \alpha \] for \( \alpha \in [0,1] \) and the choice of the burn-in \( n_0 \).

First we state two auxiliary inequalities. By the next proposition we have an upper bound on the root mean square error for \( f \in L_2 \), see [Rud12, Theorem 3.34]. The next lemma states that the absolute mean error is bounded for \( f \in L_1 \).

**Proposition 6.** Under the assumptions of Theorem \[ \text{[1]} \] we have
\[
\sup_{\|f\| \leq 1} e_2(S_{n,n_0}, f) \leq \frac{\sqrt{2}}{(n \cdot \text{gap}(P))^{1/2}} + \frac{2 M^{1/2} \left\| \frac{d\nu}{d\pi} - 1 \right\|_\infty^{1/2} \alpha^{n_0/2}}{n(1-\alpha)}.
\]

**Lemma 7.** Under the assumptions of Theorem \[ \text{[1]} \] we have
\[
\sup_{\|f\| \leq 1} e_1(S_{n,n_0}, f) \leq 2 + 4 \alpha^{n_0} M \left\| \frac{d\nu}{d\pi} - 1 \right\|_\infty.
\]

**Proof.** By \( P_j^{j+n_0} \left( \frac{d\nu}{d\pi} \right) = 1 + (P_j^{j+n_0} - E_\pi) \left( \frac{d\nu}{d\pi} - 1 \right) \) and \[ \text{[7]} \] we have
\[
e_1(S_{n,n_0}, f) \\
\leq \frac{1}{n} \sum_{j=1}^n \mathbb{E} |f(X_{j+n_0}) - E_\pi(f)| \\
= \frac{1}{n} \sum_{j=1}^n \int_G |f(x) - E_\pi(f)| \frac{d(\nu P_j^{j+n_0})}{d\pi}(x) d\pi(x) \\
= \frac{1}{n} \sum_{j=1}^n \int_G |f(x) - E_\pi(f)| P_j^{j+n_0} \left( \frac{d\nu}{d\pi} \right)(x) d\pi(x) \\
= \frac{1}{n} \sum_{j=1}^n \left( \|f - E_\pi(f)\|_1 + \int_G |f(x) - E_\pi(f)| (P_j^{j+n_0} - E_\pi)(\frac{d\nu}{d\pi} - 1)(x) d\pi(x) \right) \\
\leq 2 \|f\|_1 \frac{1}{n} \sum_{j=1}^n \left( 1 + 2 M \alpha^{j+n_0} \left\| \frac{d\nu}{d\pi} - 1 \right\|_\infty \right) \\
\leq 2 \|f\|_1 \left( 1 + 2 \alpha^{n_0} M \left\| \frac{d\nu}{d\pi} - 1 \right\|_\infty \right).
\]

Now we prove \[ \text{[3]} \]. We consider the linear operator \( T: L_p(\pi) \to L_p(\mu_{\nu,K}) \) defined by
\[
T(f) = S_{n,n_0}(f) - E_\pi(f).
\]
Recall that $\mu_{\nu,K}$ denotes the distribution of the sample trajectory. Further, note that

$$
\|T\|_{L_p(\nu) \to L_p(\mu_{\nu,K})} = \sup_{\|f\|_{L_p} \leq 1} e_p(S_{n,n_0}, f).
$$

By Proposition $\text{6}$ and Lemma $\text{7}$ we obtain

$$
\|T\|_{L_1(\nu) \to L_1(\mu_{\nu,K})} \leq M_1 \quad \text{and} \quad \|T\|_{L_2(\nu) \to L_2(\mu_{\nu,K})} \leq M_2,
$$

with

$$
M_1 = 2 + 4\alpha^{n_0} M \left\| \frac{d\nu}{d\pi} - 1 \right\|_{\infty},
$$

$$
M_2 = \frac{\sqrt{2}}{(n \cdot \text{gap}(P))^{1/2}} + \frac{2 M^{1/2} \left\| \frac{d\nu}{d\pi} - 1 \right\|_{\infty}^{1/2} \alpha^{n_0/2}}{n(1-\alpha)}.
$$

The application of the Riesz-Thorin theorem (see Proposition $\text{11}$) leads to

$$
\|T\|_{L_p(\nu) \to L_p(\mu_{\nu,K})} \leq M_1^{1-\theta} M_2^\theta
$$

with $\theta = 2 - 2/p$. The assertion of $\text{[11]}$ follows by $(x + y)^r \leq x^r + y^r$ for $x, y \geq 0$ and $r \in [0,1]$.

4.2. Proof of Theorem $\text{2}$. We prove that for any $\delta \in (0,1]$ and $p \in [1 + \delta, 2]$ by the assumptions of Theorem $\text{2}$

$$
\begin{equation}
\sup_{\|f\|_{L_p} \leq 1} e_1(S_{n,n_0}, f) \leq \left( 2 + 4(1 - \text{gap}(P)) \frac{2^{\frac{n_0}{p}}}{\delta} \left\| \frac{d\nu}{d\pi} - 1 \right\|_{\infty} \right)^{\frac{2 + \delta}{p} - 1} \times \left( \frac{2^{1-\frac{n_0}{p}}}{n \cdot \text{gap}(P)^{1-\frac{n_0}{p}}} + \frac{64 \frac{1+\delta}{\delta} \left\| \frac{d\nu}{d\pi} - 1 \right\|_{\infty} (1 - \text{gap}(P))^{2^{\frac{n_0}{p}}} \right)^{1-\frac{n_0}{p}}.
\end{equation}
$$

From this upper bound and the choice

$$
n_0 \geq \frac{1 + \delta}{2\delta} \cdot \frac{\log(\frac{32(1+\delta)}{\delta} \left\| \frac{d\nu}{d\pi} - 1 \right\|_{\infty})}{\log(1 - \text{gap}(P))^{-1}}.
$$

the assertion of the theorem follows by taking $\log(1 - \text{gap}(P))^{-1} \geq \text{gap}(P)$ and $\delta \in (0,1]$ into account.

First we state two auxiliary inequalities with parameters $p_1 \in [1,2]$ and $p_2 \in (2,4]$. By the next proposition we have an upper bound on the root mean square error for $f \in L_{p_2}$, see [Rud12 Theorem 3.41]. The next lemma states that the absolute mean error is bounded for $f \in L_{p_1}$.

**Proposition 8.** Under the assumptions of Theorem $\text{2}$ we have

$$
\sup_{\|f\|_{L_p} \leq 1} e_2(S_{n,n_0}, f) \leq \frac{\sqrt{2}}{(n \cdot \text{gap}(P))^{1/2}} + \frac{8\sqrt{p_2} \left\| \frac{d\nu}{d\pi} - 1 \right\|_{\infty}^{1/2} (1 - \text{gap}(P))^{n_0 \frac{2-p_2}{p_2}}}{n \cdot \text{gap}(P)}.
$$
Lemma 9. Under the assumptions of Theorem 3 we have
\[
\sup_{\|f\|_{p_1} \leq 1} e_1(S_{n,n_0}, f) \leq 2 + 4 \left\| \frac{d\nu}{d\pi} - 1 \right\|_\infty (1 - \text{gap}(P))^{2 \frac{p_1 - 1}{p_1} n_0}.
\]

Proof. By the same steps as in the proof of Lemma 4 we obtain the assertion with Hölder’s inequality and (8).

By the last two inequalities we can apply similar interpolation arguments as in the proof of Theorem 1. We obtain the following:

Lemma 10. Let \( p_1 \in [1, 2], p_2 \in (2, 4) \) and \( p \in [p_1, p_2] \). Then, under the assumptions of Theorem 3 we have
\[
\sup_{\|f\|_{p} \leq 1} e_q(S_{n,n_0}, f) \leq M_1 \frac{p_2}{p_2 - p_1} \left( \frac{p_2}{p_1} - 1 \right) \cdot M_2 \frac{p_1}{p_2} \left( 1 - \frac{p_2}{p_1} \right)
\]
with
\[
q = 1 + \frac{p_2(p - p_1)}{p_2(p + p_1) - 2pp_1} \in [1, 2],
\]
\[
M_1 = 2 + 4 \left\| \frac{d\nu}{d\pi} - 1 \right\|_\infty (1 - \text{gap}(P))^{2 \frac{p_1 - 1}{p_1} n_0},
\]
\[
M_2 = \frac{\sqrt{2}}{(n \cdot \text{gap}(P))^{1/2}} + \frac{8\sqrt{p_2}}{\sqrt{p_2} - 2} \left\| \frac{d\nu}{d\pi} - 1 \right\|_\infty (1 - \text{gap}(P))^{n_0 \frac{p_2 - 2}{p_2}}
\]

Proof. We consider the linear operator \( T: L_p(\pi) \to L_q(\mu, \nu) \) defined by (10). Note that
\[
\|T\|_{L_p(\pi) \to L_q(\mu, \nu)} = \sup_{\|f\|_{p} \leq 1} e_q(S_{n,n_0}, f).
\]
By Lemma 4 and Proposition 5 we have
\[
\|T\|_{L_{p_1}(\pi) \to L_{1}(\mu, \nu)} \leq M_1 \quad \text{and} \quad \|T\|_{L_{p_2}(\pi) \to L_{2}(\mu, \nu)} \leq M_2.
\]
By the Riesz-Thorin theorem (see Proposition 11) \( \|T\|_{L_p(\pi) \to L_q(\mu, \nu)} \leq M_1^{1-\theta} M_2^{\theta} \) holds for \( \theta \in [0, 1] \) satisfying \( p^{-1} = (1 - \theta)p_1^{-1} + \theta p_2^{-1} \) and \( q^{-1} = 1 - \frac{\theta}{2} \).

Note that for \( q \in [1, 2] \) we have \( e_1(S_{n,n_0}, f) \leq e_q(S_{n,n_0}, f) \). Thus, the proof of (11) follows by an application of Lemma 10 with \( p_1 = 1 + \delta \) and \( p_2 = 2(1 + \delta) \).

Appendix A. Riesz–Thorin interpolation theorem

Let \((G, \mathcal{G}, \pi)\) and \((\Omega, \mathcal{F}, \mu)\) be probability spaces. Let \( p \in [1, \infty) \) and let \( L_p(\pi) \) be the space of \( \mathcal{G}\)-measurable functions \( g: G \to \mathbb{R} \) with \( \|g\|_{p, \pi} = \left( \int_G |g(x)|^p \, d\pi(x) \right)^{1/p} < \infty \) and let \( L_p(\mu) \) be the space of \( \mathcal{F}\)-measurable functions \( f: \Omega \to \mathbb{R} \) with \( \|f\|_{p, \mu} = \left( \int_{\Omega} |f(x)|^p \, d\mu(x) \right)^{1/p} < \infty \). In the following we formulate a version of the theorem of Riesz–Thorin. For details we refer to [BSSS, Chapter 4: Corollary 1.8, Exercise 5, Corollary 2.3].
Proposition 11 (Riesz-Thorin theorem). Let $1 \leq p_k \leq q_k \leq \infty$ for $k = 1, 2$. We assume that $\theta \in [0, 1]$ and

$$\frac{1}{p} = \frac{1 - \theta}{p_1} + \frac{\theta}{p_2}, \quad \frac{1}{q} = \frac{1 - \theta}{q_1} + \frac{\theta}{q_2}.$$  

Let $T$ be a linear operator from $L_{p_1}(\pi)$ to $L_{q_1}(\mu)$ and at the same time from $L_{p_2}(\pi)$ to $L_{q_2}(\mu)$ with

$$\|T\|_{L_{p_1}(\pi) \to L_{q_1}(\mu)} \leq M_1, \quad \|T\|_{L_{p_2}(\pi) \to L_{q_2}(\mu)} \leq M_2.$$  

Then

$$\|T\|_{L_p(\pi) \to L_q(\mu)} \leq M_1^{1-\theta} M_2^\theta \tag{12}$$

and if $T$ is a positive operator, i.e. $f \geq 0$ implies $Tf \geq 0$, then (12) holds for all $1 \leq p_k, q_k \leq \infty$ with $k = 1, 2$.

References

[AG11] S. Asmussen and P. W. Glynn, A new proof of convergence of mcmc via the ergodic theorem, Statistics & Probability Letters 81 (2011), no. 10, 1482–1485.

[BC09] A. Belloni and V. Chernozhukov, On the computational complexity of MCMC-based estimators in large samples, Ann. Statist. 37 (2009), no. 4, 2011–2055.

[BS88] C. Bennett and R. Sharpley, Interpolation of operators, Pure and Applied Mathematics, vol. 129, Academic Press Inc., Boston, MA, 1988.

[CMA13] P. Cattiaux and M. Manou-Abi, Limit theorems for some functionals with heavy tails of a discrete time Markov chain, ESAIM: Probability and Statistics 18 (2013), 1–8.

[FJ11] M. Flegal and G. Jones, Implementing MCMC: Estimating with confidence, in Handbook of Markov chain Monte Carlo, Chapman & Hall (2011), 175–197.

[Hei94] S. Heinrich, Random approximation in numerical analysis, Proceedings of the Conference “Functional Analysis” Essen (1994), 123–171.

[HN02] S. Heinrich and E. Novak, Optimal summation and integration by deterministic, randomized, and quantum algorithms, Monte Carlo and Quasi-Monte Carlo Methods 2000, K.-T. Fang, F. J. Hickernell, H. Niederreiter (Eds.), Springer Verlag (2002), 50–62.

[JO10] A. Joulin and Y. Ollivier, Curvature, concentration and error estimates for Markov chain Monte Carlo, Ann. Probab. 38 (2010), no. 6, 2418–2442.

[Jon04] G. Jones, On the Markov chain central limit theorem, Probab. Surv. 1 (2004), 299–320.

[LNM13] K. Latuszynski, B. Miasojedow, and W. Niemiro, Nonasymptotic bounds on the estimation error of mcmc algorithms, Bernouilli 19 (2013), no. 5A, 2033–2066.

[LN11] K. Latuszynski and W. Niemiro, Rigorous confidence bounds for MCMC under a geometric drift condition, J. Complexity 27 (2011), no. 1, 23–38.

[LS88] G. Lawler and A. Sokal, Bounds on the $L^2$ spectrum for Markov chains and Markov processes: a generalization of Cheeger’s inequality, Trans. Amer. Math. Soc. 309 (1988), no. 2, 557–580.

[LV06] L. Lovász and S. Vempala, Hit-and-run from a corner, SIAM J. Comput. 35 (2006), no. 4, 985–1005.

[MT96] K. Mengersen and R. Tweedie, Rates of convergence of the Hastings and Metropolis algorithms, Ann. Statist. 24 (1996), no. 1, 101–121.

[MT09] S. Meyn and R. Tweedie, Markov chains and stochastic stability, second ed., Cambridge University Press, 2009.

[Nov88] E. Novak, Deterministic and Stochastic Error Bounds in Numerical Analysis, Lecture Notes in Mathematics 1349, Springer: Berlin, Heidelberg, New York, London, Paris and Tokyo, 1988.

[RR97] G. Roberts and J. Rosenthal, Geometric ergodicity and hybrid Markov chains, Electron. Comm. Probab. 2 (1997), no. 2, 13–25.

[RU13] D. Rudolf and M. Ullrich, Positivity of hit-and-run and related algorithms, Electron. Commun. Probab. 18 (2013), 1–8.

[Rud09] D. Rudolf, Explicit error bounds for lazy reversible Markov chain Monte Carlo, J. Complexity 25 (2009), no. 1, 11–24.
[Rud12] ______, *Explicit error bounds for Markov chain Monte Carlo*, Dissertationes Math. **485** (2012), 93 pp.

[vBE65] B. v. Bahr and C. Esseen, *Inequalities for the $r$th absolute moment of a sum of random variables*, $1 \leq r \leq 2$, Ann. Math. Statist. **36** (1965), 299–303.

**Inst. für Math., Universität Jena, Ernst-Abbe-Platz 2, 07743 Jena, Germany**

*E-mail address: daniel-rudolf@uni-jena.de*

**Saarland University, Dep. of Math., Postfach 151150, 66041 Saarbrücken, Germany**

*E-mail address: schweizer@math.uni-sb.de*