Contents

Abstract

We prove explicit, i.e. non-asymptotic, error bounds for Markov chain Monte Carlo methods. The problem is to compute the expectation of a function f with respect to a measure π. Different convergence properties of Markov chains imply different error bounds. For uniformly ergodic and reversible Markov chains we prove a lower and an upper error bound with respect to $\|f\|_2$. If there exists an L_2-spectral gap, which is a weaker convergence property than uniform ergodicity, then we show an upper error bound with respect to $\|f\|_p$ for $p > 2$. Usually a burn-in period is an efficient way to tune the algorithm. We provide and justify a recipe how to choose the burn-in period. The error bounds are applied to the problem of integration with respect to a possibly unnormalized density. More precise, we consider integration with respect to log-concave densities and integration over convex bodies. By the use of the Metropolis algorithm based on a ball walk and the hit-and-run algorithm it is shown that both problems are polynomial tractable.

Acknowledgements. I want to thank Aicke Hinrichs, Thomas Müller-Gronbach, Erich Novak, Klaus Ritter and Mario Ullrich for several suggestions and helpful hints. This research was supported by DFG Priority Program 1324 and the DFG Research Training Group 1523.

2010 *Mathematics Subject Classification*: Primary 60J05; Secondary 60J10, 37A25.

Key words and phrases: Markov chain Monte Carlo, spectral gap, uniform ergodicity, geometric ergodicity, burn-in, error bounds, Metropolis algorithm, ball walk, mean square error, log-concave density, hit-and-run algorithm.

Received 17.12.2010; revised version 24.7.2011.

1. Introduction and results

In numerous applications one wants to compute the expectation of a function $f\colon D \to \mathbb{R}$ with respect to a probability measure π defined on a measurable space $(D, \mathfrak{D})$. The goal is to approximate

$$S(f) = \int_D f(x)\,\pi(\mathrm{d}x), \tag{1.1}$$

where we assume that it is not possible to sample π directly with reasonable cost. In other words, we assume that there is no random number generator which generates a sample with respect to π reasonably fast. This might happen if the available information on π is incomplete or one has a complicated measurable space. However, many applications have in common that one knows enough about π to design a Markov chain which approximates the desired distribution. Hence we assume that we cannot sample π directly, but we can run a Markov chain to get close to π.

Let us briefly illustrate such problems:

- Let $A \subset \mathbb{R}^d$ be an arbitrary *convex body* ([1]). Suppose that we can sample the uniform distribution on

  $$A \cap \ell \quad \text{for an arbitrary line } \ell.$$

  The goal is to simulate the uniform distribution on A, say μ_A. For a complicated A it might be impossible to generate a uniformly distributed sample with reasonable cost. But the hit-and-run algorithm (see Section 4.2) provides a Markov chain which has the limit distribution μ_A.
- Let $D \subset \mathbb{R}^d$ be a convex body. Suppose that $f\colon D \to \mathbb{R}$ is an integrable function with respect to π_ϱ, where ϱ is an unnormalized positive density and

  $$\pi_\varrho(A) = \frac{\int_A \varrho(x)\,\mathrm{d}x}{\int_D \varrho(x)\,\mathrm{d}x}, \quad A \subset D.$$

  The goal is to approximate

  $$S(f, \varrho) = \int_D f(x)\,\pi_\varrho(\mathrm{d}x) = \frac{\int_D f(x)\varrho(x)\,\mathrm{d}x}{\int_D \varrho(x)\,\mathrm{d}x}.$$

  By the Metropolis algorithm based on the ball walk (see Section 4.1) one can construct a Markov chain which has the limit distribution π_ϱ. It might be impossible to sample π_ϱ directly, in particular if ϱ is a complicated density function.

([1]) A convex body is a bounded convex set with non-empty interior.

One can ask the following questions. How does the error of numerical integration based on Markov chains behave? And, how long does the Markov chain need to get close to the limit distribution?

The thesis deals with the first question and, because of the close relation, touches briefly the second one. The Markov chain Monte Carlo method for approximating the expectation plays a crucial role in computer science, in statistical physics, in statistics, and in financial mathematics (see e.g. [GRS96, Mar99, Liu08, Dia09, BGJM11]). Suppose that the function $f\colon D \to \mathbb{R}$ is given by an oracle which provides function values of f. The goal is to approximate $S(f)$. The integral simplifies to a sum if the state space D is finite, such that

$$S(f) = \sum_{x \in D} f(x)\, \pi(x). \tag{1.2}$$

We assume that the distribution π can be simulated by a Markov chain $(X_n)_{n\in\mathbb{N}}$ with transition kernel K and initial distribution ν. The distribution π is the limit distribution, in particular it is stationary, i.e.

$$\pi(A) = \int_D K(x, A)\, \pi(\mathrm{d}x), \quad A \in \mathfrak{D}.$$

Under weak assumptions on the Markov chain we find that after sufficiently many steps $m \geq n_0$, the distribution of X_m is close to π. The number n_0 determines the number of steps to get close to π; it is called the burn-in or the warm up period. Afterwards, we approximate $S(f)$ by

$$S_{n,n_0}(f) = \frac{1}{n} \sum_{j=1}^{n} f(X_{j+n_0}).$$

It is well known that an ergodic theorem ([2]) holds which says that

$$\lim_{n\to\infty} S_{n,n_0}(f) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} f(X_{j+n_0}) = \int_D f(x)\, \pi(\mathrm{d}x) = S(f) \quad \text{almost surely.}$$

This means that the algorithm is well defined but does not imply an error bound. It is a qualitative rather than a quantitative result. We study the mean square error of S_{n,n_0}. For a function f, integrable with respect to π, it is given by

$$e_\nu(S_{n,n_0}, f) = (\mathbf{E}_{\nu,K} |S_{n,n_0}(f) - S(f)|^2)^{1/2},$$

where $\mathbf{E}_{\nu,K}$ denotes the expectation of a Markov chain with transition kernel K and initial distribution ν.

The main topic of the thesis is the presentation of old and new explicit error bounds for computing the expectation by Markov chain Monte Carlo. These bounds are in terms

([2]) Suppose that $(D, \mathfrak{D})$ is countably generated. Let the Markov chain $(X_n)_{n\in\mathbb{N}}$ be φ-irreducible (φ is a non-trivial σ-finite measure, for all $A \in \mathfrak{D}$ and for all $x \in D$ there exists an $n \in \mathbb{N}$ such that $\varphi(A) > 0$ implies $K^n(x, A) > 0$). We assume that π is a stationary distribution. Furthermore for all $A \in \mathfrak{D}$ and for all $x \in D$ we have $\Pr(X_n \in A \text{ infinitely often} \mid X_1 = x) = 1$. Then $\lim_{n\to\infty} S_{n,n_0}(f) = S(f)$ almost surely. For a proof of this fact see [MT09, Theorem 17.1.7, p. 427]. For a simple approach to a similar ergodic theorem we refer to [AG10]. For a central limit theorem and fixed-width asymptotics of Markov chain Monte Carlo see [Gey92, JHCN06].

of the $\|\cdot\|_p$-norm of the integrand f,

$$\|f\|_p = \begin{cases} \left(\int_D |f(x)|^p \, \pi(\mathrm{d}x) \right)^{1/p}, & p \in [2, \infty), \\ \pi\text{-}\operatorname*{ess\,sup}_{x \in D} |f(x)|, & p = \infty. \end{cases}$$

The kernel K of the Markov chain determines the Markov operator

$$Pf(x) = \int_D f(y) \, K(x, \mathrm{d}y), \quad x \in D,$$

and $S(f) = \int_D f(x) \, \pi(\mathrm{d}x)$ can be considered as an operator mapping into the constant functions. If P is self-adjoint acting on L_2 then the Markov chain is called reversible. The asymptotic error is completely known if the underlying Markov chain is reversible, the initial distribution has a bounded density with respect to π and one has $\|P^j - S\|_{L_1 \to L_1} \leq M\alpha^j$ for an $\alpha \in [0, 1)$ and $M < \infty$ (Corollary 3.37). One obtains

$$\lim_{n \to \infty} n \cdot \sup_{\|f\|_2 \leq 1} e_\nu(S_{n,n_0}, f)^2 = \frac{1 + \Lambda}{1 - \Lambda} \leq \frac{2}{1 - \Lambda}, \tag{1.3}$$

and

$$\lim_{n_0 \to \infty} \sup_{\|f\|_2 \leq 1} e_\nu(S_{n,n_0}, f)^2 = \frac{1 + \Lambda}{n(1 - \Lambda)} - \frac{2\Lambda(1 - \Lambda^n)}{n^2(1 - \Lambda)^2} \leq \frac{2}{n(1 - \Lambda)}, \tag{1.4}$$

where $\Lambda = \sup\{\alpha \mid \alpha \in \mathbf{spec}(P - S)\}$. Similar asymptotic estimates are shown in [Sok97, Mat99, Bré99, Mat04, RR08]. However, we want to have explicit error bounds. The desired error estimate should behave asymptotically as described in (1.3) and (1.4). For Λ close to 1 the right hand sides of the equalities of the asymptotic error can be very well estimated by $\frac{2}{1-\Lambda}$ and $\frac{2}{n(1-\Lambda)}$. The main goal is to prove non-asymptotic, explicit error bounds with respect to $\|f\|_p$ of the form

$$\sup_{\|f\|_p \leq 1} e_\nu(S_{n,n_0}, f)^2 \leq \frac{2}{n(1 - \Lambda)} + \frac{C_{\nu,p} \, \gamma^{n_0}}{n^2(1 - \gamma)^2},$$

where $C_{\nu,p}$ and $\gamma < 1$ should be known. If the initial distribution ν of the Markov chain is the stationary one, say π, then the influence of the initial part resulting from ν should vanish, i.e. $C_{\nu,p} = 0$. We give more details in the following.

First we consider the special case where the state space is finite. Let the cardinality of D be astronomically large, say for example $|D| = 10^{30}$, so that an exact computation of the sum (1.2) might be practically impossible. Suppose that we have a Markov chain with transition matrix P and initial distribution ν. All definitions, such as stationarity, irreducibility, aperiodicity and all relevant facts of Markov chains on finite state spaces are provided in Section 2.1. The Markov chain is reversible if the transition matrix $P = (p(x, y))_{x,y \in D}$ is such that, for a probability measure π,

$$\pi(x)p(x, y) = \pi(y)p(y, x), \quad x, y \in D.$$

If the Markov chain is reversible, then let us define

$$\beta = \|P - S\|_{\ell_2 \to \ell_2} = \max\{\beta_1, |\beta_{|D|-1}|\},$$

where β_1 is the second largest and $\beta_{|D|-1}$ the smallest eigenvalue of P. We consider reversible and ergodic Markov chains, i.e. β, the second largest absolute eigenvalue of P,

is less than 1. Hence also β_1, the second largest eigenvalue of P, is less than 1. Section 2.2 contains the first error estimate. The explicit error bound is developed with respect to the ℓ_2-norm of the integrand $f \in \mathbb{R}^D$. For

$$C = \sqrt{\|1/\pi\|_\infty}\, \|\nu/\pi - 1\|_2$$

we find in Theorem 2.20 that

$$\sup_{\|f\|_2 \le 1} e_\nu(S_{n,n_0}, f)^2 \le \frac{2}{n(1-\beta_1)} + \frac{2C\beta^{n_0}}{n^2(1-\beta)^2}. \tag{1.5}$$

Obviously C is 0 if ν is π and the asymptotic estimates of (1.3) and (1.4) hold true. However, the factor $\|1/\pi\|_\infty$ is unsatisfactory for an extension to general state spaces. Furthermore we also provide a lower bound of the error (see Remark 2.24). In Section 2.3 we suggest a choice of the burn-in. The main result is as follows.

THEOREM 2.25. *Suppose that*

$$n_0 = \max\left\{ \left\lceil \frac{\log(\sqrt{\|1/\pi\|_\infty}\, \|\nu/\pi - 1\|_2)}{\log(\beta^{-1})} \right\rceil, 0 \right\}.$$

Then

$$\frac{1+\beta_1}{n(1-\beta_1)} - \frac{4}{n^2(1-\beta)^2} \le \sup_{\|f\|_2 \le 1} e_\nu(S_{n,n_0}, f)^2 \le \frac{2}{n(1-\beta)} + \frac{2}{n^2(1-\beta)^2}.$$

The suggestion of the burn-in is optimal in the following sense. For $\eta > 0$ let the number of steps $N = n + n_0$ of the Markov chain be large enough, let $\beta = \beta_1$ and assume that C and β obey an additional less restrictive condition. Then the burn-in n_{opt}, which minimizes the upper error bound of (1.5), satisfies $n_{\mathrm{opt}} \in [n_0, (1+\eta)n_0]$.

In many examples an estimate for β is available. In Section 2.4 we consider some illustrating examples where all eigenvalues and eigenvectors are known, so that the exact error is computable. Then we compare the lower and upper estimates with the exact error. It turns out that the estimates are sharp depending on the available information on the eigenvalues. Similar estimates can be found in [Ald87] and [NP09]. However, the suggestion of the burn-in and the lower bound seem to be new.

After the study of Markov chains on finite state spaces let us introduce the general state space setting. Assume that the measurable space $(D, \mathfrak{D})$ is given. Then the desired expectation becomes an integral (see (1.1)). Suppose we have a Markov chain with transition kernel K and initial distribution ν. Let us recall that the transition kernel K defines the Markov operator

$$Pf(x) = \int_D f(y)\, K(x, \mathrm{d}y), \quad x \in D,$$

and $S(f) = \int_D f(x)\, \pi(\mathrm{d}x)$ can be considered as an operator mapping into the constant functions. It is well known that reversibility of K is equivalent to self-adjointness of P acting on L_2. In Section 3.1 we provide all definitions such as stationarity and reversibility in detail. Furthermore it contains all relevant convergence properties of Markov chains. Mainly the two convergence properties of Definitions 3.14 and 3.10 are essential:

- Let $\alpha \in [0,1)$ and $M < \infty$. The Markov chain is called L_1-exponentially convergent with (α, M) if
$$\|P^j - S\|_{L_1 \to L_1} \leq M\alpha^j, \quad j \in \mathbb{N}.$$
For reversible Markov chains L_1-exponential convergence with $(\alpha, 2M)$ is equivalent to π-a.e. uniform ergodicity with (α, M), see Proposition 3.24.
- The Markov operator has an L_2-spectral gap if
$$\beta = \|P - S\|_{L_2 \to L_2} < 1,$$
where the gap is given by $1 - \beta$. The existence of an L_2-spectral gap implies an exponential convergence of P^j to S with respect to the L_2-operator norm for $j \to \infty$.

Section 3.2 contains the error estimates for S_{n,n_0}. We explain the main results in the following. Let Λ be the largest element of the spectrum of $P - S$ acting on L_2, i.e.
$$\Lambda = \sup\{\alpha \mid \alpha \in \mathbf{spec}(P - S)\}.$$
Suppose that the Markov chain is reversible and L_1-exponentially convergent with (α, M). Furthermore assume that there exists a bounded density $\frac{d\nu}{d\pi}$ of the initial distribution ν with respect to π. For
$$C = M\left\|\frac{d\nu}{d\pi} - 1\right\|_\infty$$
we show in Theorem 3.34 that the error obeys
$$\sup_{\|f\|_2 \leq 1} e_\nu(S_{n,n_0}, f)^2 \leq \frac{2}{n(1-\Lambda)} + \frac{2C\alpha^{n_0}}{n^2(1-\alpha)^2}.$$
Note that the error bound is of the same form as for finite state spaces except for the fact that the α of the L_1-exponential convergence appears. If the transition kernel is reversible one has $\Lambda \leq \beta$ and in Proposition 3.24 it is shown that $\beta \leq \alpha$. Hence one can further estimate the leading term of the error bound by using $(1-\Lambda)^{-1} \leq (1-\alpha)^{-1}$. Then a reasonable choice of the burn-in can be obtained by the same arguments as for finite state spaces. In Section 3.3 we also justify the choice of the burn-in by numerical experiments, which confirm the theoretical result.

THEOREM 3.45(i). *Suppose that we have a Markov chain which is reversible with respect to π and L_1-exponentially convergent with (α, M). Let*
$$n_0 = \max\left\{\left\lceil \frac{\log(M\|\frac{d\nu}{d\pi} - 1\|_\infty)}{\log(\alpha^{-1})} \right\rceil, 0\right\}.$$
Then
$$\sup_{\|f\|_2 \leq 1} e_\nu(S_{n,n_0}, f)^2 \leq \frac{2}{n(1-\alpha)} + \frac{2}{n^2(1-\alpha)^2}.$$

The condition that the Markov chain is L_1-exponentially convergent with (α, M) is rather restrictive. This motivates the study of Markov chains with a weaker convergence property, namely we assume that there is an L_2-spectral gap. Let us provide the main result.

THEOREM 3.45(ii). *Suppose that we have a Markov chain with Markov operator P which has an L_2-spectral gap, $1-\beta > 0$. For $p \in (2,\infty]$ let $n_0(p)$ be the smallest natural number (including zero) which is greater than or equal to*

$$\frac{1}{\log(\beta^{-1})} \cdot \begin{cases} \dfrac{p}{2(p-2)} \log\left(\dfrac{32p}{p-2} \left\| \dfrac{d\nu}{d\pi} - 1 \right\|_{\frac{p}{p-2}} \right), & p \in (2,4), \\ \log\left(64 \left\| \dfrac{d\nu}{d\pi} - 1 \right\|_2 \right), & p \in [4,\infty]. \end{cases}$$

Then

$$\sup_{\|f\|_p \le 1} e_\nu(S_{n,n_0(p)}, f)^2 \le \frac{2}{n(1-\beta)} + \frac{2}{n^2(1-\beta)^2}.$$

For further details we refer to Section 3.3. There we justify the choice for the burn-in $n_0(p)$ by numerical experiments. Briefly summarized, by weakening the convergence property we get an estimate of the error for a smaller class of functions. Namely, we prove an error bound for integrands f which satisfy $\|f\|_p < \infty$ where $p > 2$.

The last chapter deals with applications. The problem of integration with respect to log-concave densities is the following. For a function $f\colon D \to \mathbb{R}$ and a convex body $D \subset \mathbb{R}^d$ the goal is to approximate

$$S(f,\varrho) = \frac{\int_D f(x)\varrho(x)\,\mathrm{d}x}{\int_D \varrho(x)\,\mathrm{d}x},$$

where ϱ is an unnormalized density. The problem is linear in f but not in ϱ. Suppose that the domain D is the d-dimensional Euclidean unit ball B^d. Furthermore assume that ϱ is log-concave and $\log \varrho$ is Lipschitz continuous with Lipschitz constant L. Hence we consider the class of densities

$$\mathcal{R}^{\mathrm{L}}(B^d) = \{\varrho > 0 \mid \varrho \text{ is log-concave, } |\log \varrho(x) - \log \varrho(y)| \le \mathrm{L}\|x-y\|_{\mathrm{E}}\},$$

where $\|\cdot\|_{\mathrm{E}}$ denotes the Euclidean norm. We analyze the Metropolis algorithm based on a δ ball walk (see Algorithm 1 on page 80 and for the Procedure Ball-Walk see page 79). The algorithm generates the desired sample. The sample, say $(X^\delta_{n_0+1}, \dots, X^\delta_{n_0+n})$, is used to compute

$$S^\delta_{n,n_0}(f,\varrho) = \frac{1}{n}\sum_{j=1}^{n} f(X^\delta_{j+n_0}).$$

For an adapted $\delta = \min\{(d+1)^{-1}, \mathrm{L}^{-1}\}$ Mathé and Novak showed in [MN07] that the Markov chain which is defined by the Metropolis algorithm based on a δ ball walk has an L_2-spectral gap. This result is used to get an explicit error bound. We state the result for the unit ball and for simplicity we consider integrands f with $\|f\|_3 \le 1$. For

$$n_0 \asymp d\,\mathrm{L}\max\{d, \mathrm{L}^2\}$$

the error obeys

$$\sup_{\|f\|_3 \le 1,\, \varrho \in \mathcal{R}^{\mathrm{L}}(B^d)} e(S^\delta_{n,n_0}, (f,\varrho)) \prec \sqrt{\frac{d}{n}}\max\{\sqrt{d}, \mathrm{L}\} + \frac{d}{n}\max\{d, \mathrm{L}^2\},$$

where $d \in \mathbb{N}$ and $\mathrm{L} \geq 0$ ([3]). The geometry of the unit ball is essential for the estimate of the L_2-spectral gap of [MN07], since the ball walk might get stuck on domains which have corners. However, the results of Section 4.1 are slightly more general. There we treat balls with arbitrary radius $r > 0$ and the result is with respect to $\|f\|_p$ for $p > 2$. We refer to Theorem 4.8 for the details. The number of function evaluations to obtain an error smaller than ε is polynomially bounded in the dimension d and the Lipschitz constant L. Hence the problem of integration with respect to a log-concave density is tractable (see Novak and Woźniakowski [NW08, NW10]).

The problem of integration on a convex body is as follows. Let $A \subset \mathbb{R}^d$ be a convex body. The goal is to compute

$$S(f, A) = \frac{1}{\mathbf{vol}_d(A)} \int_A f(A)\,\mathrm{d}x,$$

where $\mathbf{vol}_d(A)$ denotes the d-dimensional volume of A. In other words the goal is to approximate the expectation of f with respect to the uniform distribution, say μ_A, on A. The problem is linear in f but not in A. Let $B^d \subset A \subset rB^d$ where rB^d is the Euclidean ball with radius r around 0. We assume that there is an oracle $\mathrm{Or}_A(\ell)$ which returns a uniform distributed state on $A \cap \ell$ for an arbitrary line ℓ. Hence we consider state spaces from the class

$$\mathcal{S}_d(r) = \{A \subset \mathbb{R}^d \text{ convex} \mid B^d \subset A \subset rB^d\}$$

and we assume that $\mathrm{Or}_A(\ell)$ is available for any $A \in \mathcal{S}_d(r)$. We analyze the hit-and-run algorithm, see Algorithm 2 on page 85 and for the Procedure Hit-and-Run see page 84. It generates the desired sample, say $(X^{\mathrm{har}}_{n_0+1}, \ldots, X^{\mathrm{har}}_{n+n_0})$. Afterwards we compute

$$S^{\mathrm{har}}_{n,n_0}(f, A) = \frac{1}{n} \sum_{j=1}^{n} f(X^{\mathrm{har}}_{j+n_0}).$$

The Markov chain generated by the hit-and-run algorithm has the right stationary distribution (see Lemma 4.10 or [Smi84]). A result of Lovász and Vempala presented in [LV06] provides an estimate of $1 - \beta$. Hence there exists an L_2-spectral gap and we can apply the error bound of Theorem 3.45 (ii). For simplicity suppose that $\|f\|_3 \leq 1$. For

$$n_0 \asymp d^3 r^2 \log(r)$$

the error obeys

$$\sup_{\|f\|_3 \leq 1,\, A \in \mathcal{S}_d(r)} e(S^{\mathrm{har}}_{n,n_0}, (f, A)) \prec \frac{dr}{\sqrt{n}} + \frac{d^2 r^2}{n}.$$

For the general result with respect to $\|f\|_p$ with $p > 2$ we refer to Theorem 4.12. The number of function evaluations to obtain an error smaller than ε is polynomially bounded in the dimension d and the radius r. Hence the problem of integration on a convex body is tractable (see [NW08, NW10]).

([3]) We use the notation $\prec$ and $\asymp$ as follows. Let $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ be positive sequences. We write $a_n \prec b_n$ if there exists an absolute constant c such that $a_n \leq c b_n$ for all $n \in \mathbb{N}$. We write $a_n \asymp b_n$ if $a_n \prec b_n$ and $b_n \prec a_n$.

2. Finite state spaces

In the following we study the mean square error of Markov chain Monte Carlo methods on finite state spaces. In Section 2.1 the basic definitions and properties of Markov chains on finite state spaces are stated. The estimate of the mean square error is shown in Section 2.2. We suggest and justify a recipe how to choose the burn-in. Afterwards the error bound is applied to illustrating examples and finally we discuss how the results fit into the published literature.

2.1. Markov chains. In this section the basics of Markov chains on finite state spaces are provided. Let D be a finite set and $\mathcal{P}(D)$ the power set of D, so that the measurable state space $(D, \mathcal{P}(D))$ is given.

Definition 2.1 (Markov chain). A sequence $(X_n)_{n\in\mathbb{N}}$ of random variables on a probability space $(\Omega, \mathcal{F}, \Pr)$ mapping into $(D, \mathcal{P}(D))$ is called a *Markov chain with transition matrix* $P = (p(x,y))_{x,y\in D}$ if for all $n \in \mathbb{N}$, all $x, y \in D$ and all $x_1, \dots, x_{n-1}$ with

$$\Pr(X_1 = x_1, \dots, X_{n-1} = x_{n-1}, X_n = x) > 0$$

one has

$$\Pr(X_{n+1} = y \mid X_1 = x_1, \dots, X_{n-1} = x_{n-1}, X_n = x) = \Pr(X_{n+1} = y \mid X_n = x) = p(x,y).$$

All entries of the transition matrix P are non-negative and the rows sum up to 1. For $x, y \in D$ the value $p(x,y)$ is the probability of jumping from state x to state y in a single step of the chain. The distribution

$$\nu(x) = \Pr(X_1 = x), \quad x \in D,$$

is called the *initial distribution.*

Suppose that we have a transition matrix P and a probability measure ν. Any transition matrix P has a *random mapping representation* [LPW09, Proposition 1.5, p. 7]. A random mapping representation of P on state space D is a function $\Phi\colon D \times [0,1] \to D$ which satisfies

$$\Pr(\Phi(x, Z) = y) = p(x,y), \quad x, y \in D,$$

where $Z : (\Omega, \mathcal{F}, \Pr) \to ([0,1], \mathcal{B}([0,1]))$ is a uniformly distributed random variable and $\mathcal{B}([0,1])$ denotes the Borel σ-algebra. Then a Markov chain can be constructed as follows. If $(Z_n)_{n\in\mathbb{N}}$ is a sequence of i.i.d. random variables with uniform distribution, and X_1 has a distribution ν, then it is easy to see that $(X_n)_{n\in\mathbb{N}}$ defined by

$$X_n = \Phi(X_{n-1}, Z_n), \quad n \geq 2,$$

is a Markov chain with transition matrix P and initial distribution ν.

In the following we assume that we have a Markov chain $(X_n)_{n\in\mathbb{N}}$ with transition matrix P and initial distribution ν. The expectation $\mathbf{E}_{\nu,P}$ is taken with respect to the joint distribution of $(X_n)_{n\in\mathbb{N}}$, say $W_{\nu,P}$, which is defined on $(D^{\mathbb{N}}, \sigma(\mathcal{A}))$ where

$$D^{\mathbb{N}} = \{\omega = (x_1, x_2, \dots) \mid x_i \in D \text{ for all } i \geq 1\},$$
$$\mathcal{A} = \bigcup_{k\in\mathbb{N}} \{A_1 \times \cdots \times A_k \times D \times \cdots \mid A_i \in \mathcal{P}(D),\ i = 1, \dots, k\}.$$

For $k \in \mathbb{N}$ one has, with $A_1 \times \cdots \times A_k \subset D^k$,

$$W_{\nu,P}(A_1 \times \cdots \times A_k \times D \times \cdots) = \sum_{x_1 \in A_1} \cdots \sum_{x_k \in A_k} \Pr(X_1 = x_1, \ldots, X_k = x_k).$$

If the initial distribution ν is δ_x, the point mass at $x \in D$, we say that the Markov chain *starts at* x. By

$$Pf(x) = \sum_{y \in D} f(y)\, p(x,y) = \sum_{y \in D} f(y) \Pr(X_2 = y \mid X_1 = x) = \mathbf{E}_{\delta_x, P}[f(X_2)]$$

one has the expectation of $f \in \mathbb{R}^D$ after a single step of the chain which starts at $x \in D$. The probability to get from x to y in $k \geq 2$ steps is

$$\Pr(X_{k+1} = y \mid X_1 = x) = \sum_{x_2 \in D} \sum_{x_3 \in D} \cdots \sum_{x_k \in D} p(x,x_2)p(x_2,x_3)\ldots p(x_k,y) = p^k(x,y),$$

where $P^k = (p^k(x,y))_{x,y \in D}$ denotes the kth power of P. Then

$$P^k f(x) = \sum_{y \in D} f(y)\, p^k(x,y) = \sum_{y \in D} f(y) \Pr(X_{k+1} = y \mid X_1 = x) = \mathbf{E}_{\delta_x, P}[f(X_{k+1})]$$

is the expectation after k steps of the Markov chain which starts at x. Similarly we consider the application of P to a probability measure ν, i.e.

$$\nu P(x) = \sum_{y \in D} p(y,x)\, \nu(y) = \sum_{y \in D} \Pr(X_2 = x \mid X_1 = y)\, \nu(y) = \Pr(X_2 = x).$$

This is the distribution which arises after a single transition where the initial state is chosen by ν. The distribution which arises after $k \geq 1$ steps is given by

$$\nu P^k(x) = \sum_{y \in D} p^k(y,x)\, \nu(y) = \sum_{y \in D} \Pr(X_{k+1} = x \mid X_1 = y)\, \nu(y) = \Pr(X_{k+1} = x).$$

In the following we present properties of transition matrices.

Definition 2.2 (irreducibility, aperiodicity, periodicity). A transition matrix P is called *irreducible* if for all $x, y \in D$ there exists a $k \in \mathbb{N}$ such that

$$p^k(x,y) > 0, \quad \text{where} \quad P^k = (p^k(x,y))_{x,y \in D}.$$

A transition matrix P is called *aperiodic* if for all $x \in D$,

$$\gcd(\{k \in \mathbb{N} \mid p^k(x,x) > 0\}) = 1,$$

where gcd denotes the greatest common divisor. If P is not aperiodic we call it *periodic*.

If a transition matrix is irreducible, then one can reach every state y from every state x in finitely many steps. Aperiodicity ensures that the number of steps to return to an arbitrary state is not in $\{m, 2m, 3m, \ldots\}$ for $m > 1$.

Definition 2.3 (stationarity). Let π be a probability measure on D. Then π is called a *stationary distribution* of a transition matrix P if

$$\pi P(x) = \pi(x), \quad x \in D.$$

If the initial distribution of a Markov chain with transition matrix P is a stationary one, say π, then after a single transition the same distribution as the initial one appears,

i.e.

$$\Pr(X_1 = x) = \pi(x) = \pi P(x) = \Pr(X_2 = x), \quad x \in D.$$

Definition 2.4 (reversibility). Let π be a probability measure on D. A transition matrix P is called *reversible with respect to* π if

$$\pi(x)p(x,y) = \pi(y)p(y,x), \quad x,y \in D.$$

If a transition matrix P is reversible with respect to a probability measure π, then π is a stationary distribution (see [LPW09, Proposition 1.19, p. 14]). If the initial distribution of a Markov chain with transition matrix P is π, then reversibility with respect to π is equivalent to

$$\Pr(X_1 = x, X_2 = y) = \Pr(X_1 = y, X_2 = x), \quad x,y \in D.$$

Definition 2.5 (lazy version). Let P be a transition matrix and let I be the identity matrix. Then we call

$$\widetilde{P} = \tfrac{1}{2}(I + P)$$

the *lazy version of* P.

Let π be a stationary distribution of a transition matrix P. Then π is also stationary with respect to $\widetilde{P}$. If P is irreducible, reversible with respect to π and periodic, then one can pass over to the lazy version $\widetilde{P}$ and find that $\widetilde{P}$ is irreducible, reversible with respect to π and aperiodic.

A Markov chain is called irreducible, periodic, aperiodic and reversible with respect to π if the corresponding transition matrix is irreducible, periodic, aperiodic and reversible with respect to π, respectively.

Let us state some well known implications of the different properties. For proofs or more details see [Bré99, Str05, LPW09]. For every transition matrix there exists a stationary distribution and if the matrix is irreducible then there exists a unique stationary distribution, which is positive ([LPW09, Proposition 1.14, p. 12 and Corollary 1.17, p. 14]). Note that if ξ is an eigenvalue of a transition matrix P, then $|\xi| \leq 1$ ([LPW09, Lemma 12.1(i), p. 153]). Furthermore, for irreducible transition matrices 1 is a simple eigenvalue ([LPW09, Lemma 12.1(ii), p. 153]). If the Markov chain is aperiodic and irreducible, then -1 is not an eigenvalue of P ([LPW09, Lemma 12.1(iii), p. 153] or [Str05, Theorem 5.1.14, p. 113]). These eigenvalue results are also known as consequences of the Perron–Frobenius Theorem (see [Sen06]).

In the following we always assume that the Markov chains are irreducible, aperiodic and reversible with respect to a probability measure π. Hence π is a stationary distribution. Aperiodicity is not essential. For a Markov chain with periodic transition matrix P and initial distribution ν we may consider a *lazy Markov chain*, i.e. a chain with aperiodic transition matrix $\widetilde{P}$, the lazy version of P, and initial distribution ν.

Let us define a weighted inner product for $f, g \in \mathbb{R}^D$ by

$$\langle f, g \rangle = \sum_{x \in D} f(x)g(x)\,\pi(x)$$

and let $\|f\|_2 = \langle f, f \rangle^{1/2}$. By considering the inner product it is easy to see that reversibility is equivalent to P being self-adjoint. Applying the spectral theorem for self-adjoint

transition matrices and the fact that the Markov chain is irreducible one finds that P has real eigenvalues

$$1 = \beta_0 > \beta_1 \geq \cdots \geq \beta_{|D|-1} \geq -1.$$

If the transition matrix is aperiodic, then $\beta_{|D|-1} > -1$. There exists a basis of orthonormal eigenfunctions (vectors) $\{u_0, u_1, \ldots, u_{|D|-1}\}$, i.e. for $i, j \in \{0, \ldots, |D|-1\}$ one has

$$Pu_i = \beta_i u_i, \qquad \langle u_i, u_j \rangle = \delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

Clearly, $u_0(x) = 1$ for all $x \in D$ and $S(u_i) = \langle u_i, u_0 \rangle = 0$ for $i \in \{1, \ldots, |D|-1\}$. By the spectral structure of the transition matrix one has

$$P^k = (p^k(x,y))_{x,y \in D} = \sum_{i=0}^{|D|-1} \beta_i^k \, (u_i(x) u_i(y) \, \pi(y))_{x,y \in D} \tag{2.1}$$

(see [Bré99, p. 203] for details).

For $p \in [1, \infty]$ let

$$\|f\|_p = \begin{cases} \Big(\sum\limits_{x \in D} |f(x)|^p \pi(x) \Big)^{1/p}, & p \in [1, \infty), \\ \sup\limits_{x \in D} |f(x)|, & p = \infty. \end{cases}$$

The weighted vector space $\ell_p = \ell_p(D, \pi)$ is defined to be the normed space $(\mathbb{R}^D, \|\cdot\|_p)$. Furthermore let

$$\ell_p^0 = \ell_p^0(D, \pi) = \{f \in \ell_p \mid S(f) = 0\}.$$

Then

$$\ell_2 = \ell_2^0 \oplus (\ell_2^0)^\perp \quad \text{with} \quad (\ell_2^0)^\perp = \{f \in \mathbb{R}^D \mid f \equiv c, \ c \in \mathbb{R}\} = \mathrm{Eig}(P, 1),$$

where $\mathrm{Eig}(P, 1)$ is the eigenspace of P with eigenvalue 1. Note that for the next well known result it is not assumed that the transition matrix is reversible with respect to π.

LEMMA 2.6. *Let $p \in [1, \infty]$ and $f \in \mathbb{R}^D$. For any transition matrix P with stationary distribution π one obtains*

$$\|Pf\|_p \leq \|f\|_p \quad \textit{and} \quad \|P\|_{\ell_p \to \ell_p} = 1.$$

Proof. By the Jensen inequality (J) ([1]) and stationarity (stat.) one has

$$\begin{aligned} \sum_{x \in D} |Pf(x)|^p \pi(x) &\leq \sum_{x \in D} \Big(\sum_{y \in D} |f(y)| p(x,y) \Big)^p \pi(x) \\ &\underset{\text{(J)}}{\leq} \sum_{x \in D} \sum_{y \in D} |f(y)|^p p(x,y) \, \pi(x) \underset{\text{(stat.)}}{=} \sum_{x \in D} |f(x)|^p \pi(x). \end{aligned}$$

([1]) Let $(D, \mathfrak{D}, \mu)$ be a probability space. For any convex function $h \colon \mathbb{R} \to \mathbb{R}$ and for any function $f \colon D \to \mathbb{R}$ that is integrable with respect to μ, the Jensen inequality is $h(\int_D f \, \mathrm{d}\mu) \leq \int_D (h \circ f) \, \mathrm{d}\mu$.

If $p = \infty$ then

$$\|Pf\|_\infty = \sup_{x\in D} |Pf(x)| \le \sup_{x\in D} \sum_{y\in D} |f(y)|p(x,y) = \|f\|_\infty.$$

Since $\|Pf\|_p \le \|f\|_p$ and $Pu_0 = u_0$ with $\|u_0\|_p = 1$ we have $\|P\|_{\ell_p\to\ell_p} = 1$. ∎

Let us briefly explain how to quantify the difference of two distributions. For any measure ν let

$$\left\|\frac{\nu}{\pi}\right\|_2 = \left(\sum_{x\in D}\left(\frac{\nu(x)}{\pi(x)}\right)^2 \pi(x)\right)^{1/2}.$$

If ν is a probability measure on D, then the quantity $\|\nu/\pi - 1\|_2$ is related to the χ^2-contrast, defined as follows.

Definition 2.7 (χ^2-contrast). The *χ^2-contrast* of a distribution ν and a positive distribution μ is defined by

$$\chi^2(\nu,\mu) = \sum_{x\in D} \frac{(\nu(x)-\mu(x))^2}{\mu(x)}.$$

The χ^2-contrast is not symmetric and therefore no distance. By a simple calculation one obtains

$$\chi^2(\nu,\pi) = \left\|\frac{\nu}{\pi} - 1\right\|_2^2.$$

The functional S can be interpreted as an operator which maps into the constant functions, so one can view

$$S = (\pi(y))_{x,y\in D}$$

as a matrix. Furthermore let

$$\beta = \max\{\beta_1, |\beta_{|D|-1}|\}$$

be the second largest absolute value of the eigenvalues of P. Now we state a property of the matrix $P - S$.

Lemma 2.8. *Let P be a reversible transition matrix with respect to π. Then*

$$\|P^n - S\|_{\ell_2\to\ell_2} = \|P^n\|_{\ell_2^0\to\ell_2^0} = \beta^n, \quad n \in \mathbb{N}. \tag{2.2}$$

Proof. The self-adjointness of P implies that $\|P\|_{\ell_2^0\to\ell_2^0} = \max\{\beta_1, |\beta_{|D|-1}|\} = \beta$, and consequently $\|P^n\|_{\ell_2^0\to\ell_2^0} = \beta^n$. From

$$\begin{aligned}\|P^n - S\|_{\ell_2\to\ell_2} &= \sup_{\|f\|_2\le 1} \|(P^n - S)f\|_2 = \sup_{\|f\|_2\le 1} \|P^n(f - S(f))\|_2 \\ &\le \sup_{\|g\|_2\le 1,\ S(g)=0} \|P^n g\|_2 = \|P^n\|_{\ell_2^0\to\ell_2^0}\end{aligned}$$

and

$$\begin{aligned}\|P^n\|_{\ell_2^0\to\ell_2^0} &= \sup_{\|g\|_2\le 1,\ S(g)=0} \|P^n g\|_2 = \sup_{\|g\|_2\le 1,\ S(g)=0} \|P^n g - S(g)\|_2 \\ &\le \sup_{\|f\|_2\le 1} \|(P^n - S)f\|_2 = \|P^n - S\|_{\ell_2\to\ell_2}\end{aligned}$$

claim (2.2) follows. ∎

This section is finished by stating a well known fact which shows that νP^k converges to π for increasing k exponentially fast if $\beta < 1$.

COROLLARY 2.9. *Let P be a transition matrix and ν a probability measure on D. Let P be reversible with respect to π. Then*

$$\left\| \frac{\nu P^k}{\pi} - 1 \right\|_2 \leq \beta^k \left\| \frac{\nu}{\pi} - 1 \right\|_2, \quad k \in \mathbb{N}.$$

Proof. Indeed,

$$\left\| \frac{\nu P^k}{\pi} - 1 \right\|_2 \underset{\text{(rev.)}}{=} \left\| P^k \left(\frac{\nu}{\pi} \right) - 1 \right\|_2 = \left\| P^k \left(\frac{\nu}{\pi} - 1 \right) \right\|_2 \leq \beta^k \left\| \frac{\nu}{\pi} - 1 \right\|_2. \blacksquare$$

2.2. Error bounds. In this section explicit error bounds are proven. Let us repeat the idea of Markov chain Monte Carlo. Suppose we have a Markov chain $(X_n)_{n\in\mathbb{N}}$ with transition matrix P and initial distribution ν, where π is a stationary distribution, and we compute

$$S_{n,n_0}(f) = \frac{1}{n} \sum_{j=1}^{n} f(X_{j+n_0})$$

as an approximation for $S(f) = \sum_{x\in D} f(x)\,\pi(x)$. The error is measured in the mean square sense, i.e.

$$e_\nu(S_{n,n_0}, f) = (\mathbf{E}_{\nu,P} |S_{n,n_0}(f) - S(f)|^2)^{1/2}.$$

Now let us present a helpful result.

LEMMA 2.10. *Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with transition matrix P and initial distribution ν. Then for $i, j \in \mathbb{N}$ with $j \leq i$,*

$$\mathbf{E}_{\nu,P}[f(X_i) f(X_j)] = \sum_{x\in D} P^j (f P^{i-j} f)(x)\, \nu(x). \tag{2.3}$$

Moreover, if π is a stationary distribution and $\nu = \pi$ then

$$\mathbf{E}_{\pi,P}[f(X_i) f(X_j)] = \langle f, P^{i-j} f \rangle. \tag{2.4}$$

Proof. The calculation

$$\begin{aligned}
\mathbf{E}_{\nu,P}[f(X_i) f(X_j)] &= \sum_{x_1\in D} \cdots \sum_{x_i\in D} f(x_j) f(x_i)\, p(x_{i-1}, x_i) \cdots p(x_1, x_2) \nu(x_1) \\
&= \sum_{x_1\in D} \cdots \sum_{x_j\in D} f(x_j) P^{i-j} f(x_j)\, p(x_{j-1}, x_j) \cdots p(x_1, x_2) \nu(x_1) \\
&= \sum_{x\in D} P^j (f P^{i-j} f)(x)\, \nu(x)
\end{aligned}$$

proves (2.3) and by using $\pi P(x) = \pi(x)$ one gets (2.4). $\blacksquare$

In the following a special case of the S_{n,n_0} method is considered. In this case the initial distribution is stationary, thus, the distribution after a single transition does not change. Hence it is not necessary to do any burn-in, i.e. $n_0 = 0$. Afterwards the error representation of the special case is set in relation to the error where the initial distribution might differ from a stationary one. The techniques which are used are adapted from [Rud09] and [Rud10].

In the following $S_{n,0}$ is always denoted by S_n. Let us start with a result stated in [BD06, Proposition 2.1, p. 3].

PROPOSITION 2.11. *Let $f \in \mathbb{R}^D$ and let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with transition matrix P and initial distribution π. Let P be reversible with respect to π. Then*

$$e_\pi(S_n, f)^2 = \frac{1}{n^2} \sum_{k=1}^{|D|-1} a_k^2 \, W(n, \beta_k), \tag{2.5}$$

where

$$a_k = \langle f, u_k \rangle \quad \text{and} \quad W(n, \beta_k) = \frac{n(1-\beta_k^2) - 2\beta_k(1-\beta_k^n)}{(1-\beta_k)^2}.$$

Proof. Let us consider $g = f - S(f) \in \mathbb{R}^D$. The error obeys

$$\begin{aligned} e_\pi(S_n, f)^2 &= \mathbf{E}_{\pi,P}\left|\frac{1}{n}\sum_{j=1}^n g(X_j)\right|^2 = \frac{1}{n^2}\mathbf{E}_{\pi,P}\left|\sum_{j=1}^n g(X_j)\right|^2 \\ &= \frac{1}{n^2}\sum_{j=1}^n \mathbf{E}_{\pi,P}\left[g(X_j)^2\right] + \frac{2}{n^2}\sum_{j=1}^{n-1}\sum_{i=j+1}^{n} \mathbf{E}_{\pi,P}\left[g(X_j)g(X_i)\right]. \end{aligned}$$

By using the orthonormal basis $\{u_0, u_1, \ldots, u_{|D|-1}\}$ we have $g(x) = \sum_{k=1}^{|D|-1} a_k u_k(x)$. For $j \le i$ one obtains

$$\begin{aligned} \mathbf{E}_{\pi,P}\left[g(X_i)g(X_j)\right] &= \sum_{k=1}^{|D|-1}\sum_{l=1}^{|D|-1} a_k a_l \, \mathbf{E}_{\pi,P}\left[u_k(X_i)u_l(X_j)\right] \\ &\underset{(2.4)}{=} \sum_{k=1}^{|D|-1}\sum_{l=1}^{|D|-1} a_k a_l \, \langle u_k, P^{i-j} u_l \rangle \\ &= \sum_{k=1}^{|D|-1}\sum_{l=1}^{|D|-1} a_k a_l \, \beta_l^{i-j} \langle u_k, u_l \rangle = \sum_{k=1}^{|D|-1} a_k^2 \, \beta_k^{i-j}. \end{aligned}$$

The last two equalities follow from the orthonormality of the basis of the eigenvectors. Altogether this gives

$$\begin{aligned} e_\pi(S_n, f)^2 &= \frac{1}{n^2}\sum_{k=1}^{|D|-1} a_k^2 \Big[n + 2\sum_{j=1}^{n-1}\sum_{i=j+1}^{n} \beta_k^{i-j} \Big] \\ &= \frac{1}{n^2}\sum_{k=1}^{|D|-1} a_k^2 \left[n + 2\frac{(n-1)\beta_k - n\beta_k^2 + \beta_k^{n+1}}{(1-\beta_k)^2} \right] = \frac{1}{n^2}\sum_{k=1}^{|D|-1} a_k^2 \, W(n, \beta_k). \quad \blacksquare \end{aligned}$$

Let us consider $W(n, \beta_k)$ to simplify and interpret Proposition 2.11.

LEMMA 2.12. *For all $n \in \mathbb{N}$ and $k \in \{1, \ldots, |D| - 1\}$,*

$$W(n, \beta_k) \le W(n, \beta_1) \le \frac{2n}{1-\beta_1}. \tag{2.6}$$

Proof. We will show that the mapping $x \mapsto W(n, x)$ is increasing on $[-1, 1)$, so that $W(n, \beta_k) \le W(n, \beta_1)$. For $i \in \{0, \ldots, n\}$ we have

$$x^{n-i} \le 1 \;\Leftrightarrow\; (1 - x^i)\, x^{n-i} \le 1 - x^i \;\Leftrightarrow\; x^{n-i} + x^i \le 1 + x^n.$$

This implies

$$x^j + x^{j+1} + x^{n-j-1} + x^{n-j} \le 2(1+x^n), \quad j \in \{0,\dots,n-1\},$$

and

$$(1+x)\sum_{j=0}^{n-1} x^j = \frac{1}{2}\sum_{j=0}^{n-1}(x^j + x^{j+1} + x^{n-j-1} + x^{n-j}) \le n(1+x^n).$$

Now

$$\frac{dW}{dx}(n,x) = -2\frac{(1+x)\sum_{j=0}^{n-1} x^j - n(1+x^n)}{(1-x)^2} \ge 0$$

and the first inequality of the assertion is proven. Since

$$W(n,x) \le \begin{cases} \dfrac{n(1+x)-2xn}{1-x} \le \dfrac{2n}{1-x}, & x \in [-1,0], \\ \dfrac{n(1+x)}{1-x} \le \dfrac{2n}{1-x}, & x \in (0,1), \end{cases}$$

the proof is complete. ∎

An explicit formula for the error is established if the initial state is chosen by a stationary distribution. Let us consider the maximal error of S_n for f which satisfies $\|f\|_2 \le 1$.

COROLLARY 2.13. *Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with transition matrix P and initial distribution π. Let P be reversible with respect to π. Then*

$$\sup_{\|f\|_2\le 1} e_\pi(S_n,f)^2 = \frac{1+\beta_1}{n(1-\beta_1)} - \frac{2\beta_1(1-\beta_1^n)}{n^2(1-\beta_1)^2} \le \frac{2}{n(1-\beta_1)}. \tag{2.7}$$

Proof. The individual error of f is

$$\begin{aligned} e_\pi(S_n,f)^2 &\underset{(2.5)}{=} \frac{1}{n^2}\sum_{k=1}^{|D|-1} a_k^2\, W(n,\beta_k) \le \frac{\|f\|_2^2}{n^2}\max_{k=1,\dots,|D|-1} W(n,\beta_k) \\ &\underset{(2.6)}{=} \frac{\|f\|_2^2}{n^2} W(n,\beta_1) = \frac{1+\beta_1}{n(1-\beta_1)}\|f\|_2^2 - \frac{2\beta_1(1-\beta_1^n)}{n^2(1-\beta_1)^2}\|f\|_2^2, \end{aligned}$$

where a_k is chosen as in Proposition 2.11 and therefore $\sum_{k=1}^{|D|-1} a_k^2 \le \|f\|_2^2$. From the preceding analysis of the individual error we have for $\|f\|_2 \le 1$ the right upper error bound. Now we consider $f = u_1$, where $\|u_1\|_2 = 1$. By applying (2.5) we have

$$e_\pi(S_n,u_1)^2 = \frac{1+\beta_1}{n(1-\beta_1)} - \frac{2\beta_1(1-\beta_1^n)}{n^2(1-\beta_1)^2}.$$

Thus the equality of (2.7) is proven, and (2.6) yields the inequality. ∎

In Corollary 2.13 an explicit error bound with respect to $\|\cdot\|_2$ is shown. Notice that the first part of (2.7) is an equality, which means that the integration error is completely known if the initial distribution is stationary.

Suppose that the distribution π can be simulated directly, i.e. we can apply a Monte Carlo method with an i.i.d. sample. Then an i.i.d. sequence $(X_n)_{n\in\mathbb{N}}$, where every X_n is

distributed with respect to π, is a Markov chain with transition matrix $S = (\pi(y))_{x,y\in D}$ and initial distribution π. In this setting one has

$$e_\pi(S_n, f)^2 = \frac{1}{n}\|f - S(f)\|_2^2.$$

This corresponds to $\beta_i = 0$ for all $i > 0$. In some artificial cases other Markov chain Monte Carlo methods can do better; for example, if there is a Markov chain where $\beta_i < 0$ and the target is to approximate $S(u_i)$ or if all eigenvalues are smaller than zero. A simple transition matrix which satisfies this eigenvalue condition is given by

$$P = \begin{pmatrix} 0 & \frac{1}{|D|-1} & \cdots & \frac{1}{|D|-1} \\ \frac{1}{|D|-1} & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{1}{|D|-1} \\ \frac{1}{|D|-1} & \cdots & \frac{1}{|D|-1} & 0 \end{pmatrix},$$

where $\pi(x) = 1/|D|$ for all $x \in D$ (see [FHY92, Remark 3, p. 617]). It turns out that $\beta_1 = \cdots = \beta_{|D|-1} = -\frac{1}{|D|-1}$. For large $|D|$ it is unfortunately not possible to construct a transition matrix where β_1 is close to -1.

Proposition 2.14. *Let P be an irreducible transition matrix. Then*

$$\beta_1 \geq -\frac{1}{|D|-1}.$$

Proof. Since $\beta_0 = 1$ one has

$$1 + \sum_{i=1}^{|D|-1} \beta_i = \sum_{i=0}^{|D|-1} \beta_i = \operatorname{trace}(P) = \sum_{x\in D} p(x,x) \geq 0.$$

Then

$$-1 \leq \sum_{i=1}^{|D|-1} \beta_i \leq (|D|-1)\beta_1. \quad \blacksquare$$

The error estimates under the assumption that the initial distribution is stationary seem to be restrictive. If we could sample π directly we would approximate $S(f)$ by Monte Carlo with an i.i.d. sample. However, even if it is possible it might happen that the sampling procedure is computationally expensive. It can be reasonable to generate only the initial state by sampling from π and afterwards run a Markov chain with stationary distribution π. Perfect sampling might be helpful for the construction of such direct sampling procedures (see [PW96, Häg02]).

In the following we consider the case where the initial distribution is not necessarily stationary. Let ν be a distribution on D and $k \in \mathbb{N}$. Then we define

$$d_k(x) = \sum_{y\in D} \frac{\nu(y)}{\pi(y)}(p^k(x,y) - \pi(y)) = P^k\left(\frac{\nu}{\pi}\right)(x) - 1 = (P^k - S)\left(\frac{\nu}{\pi} - 1\right)(x), \quad x \in D.$$

If P is reversible with respect to π, then we obtain

$$\|d_k\|_2 = \left\|\frac{\nu P^k}{\pi} - 1\right\|_2, \quad k \in \mathbb{N},$$

thus d_k determines the difference between νP^k and the stationary distribution π. Additionally by the spectral representation of P^k (see (2.1)) one obtains

$$d_k(x) = \sum_{i=1}^{|D|-1} \beta_i^k \sum_{y\in D} u_i(y)\nu(y)\, u_i(x) = \sum_{i=1}^{|D|-1} \beta_i^k \left\langle \frac{\nu}{\pi}, u_i \right\rangle u_i(x), \quad x \in D. \tag{2.8}$$

The next statement gives a relation between $e_\nu(S_{n,n_0}, f)$ and $e_\pi(S_n, f)$.

PROPOSITION 2.15. *Let $f \in \mathbb{R}^D$ and $g = f - S(f)$. Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with transition matrix P and initial distribution ν. Let P be reversible with respect to π. Then*

$$e_\nu(S_{n,n_0}, f)^2 = e_\pi(S_n, f)^2 + \frac{1}{n^2}\sum_{j=1}^{n} L_{j+n_0}(g^2) + \frac{2}{n^2}\sum_{j=1}^{n-1}\sum_{k=j+1}^{n} L_{j+n_0}(gP^{k-j}g), \tag{2.9}$$

where

$$L_i(h) = \langle d_i, h\rangle = \left\langle (P^i - S)\left(\frac{\nu}{\pi} - 1\right), h \right\rangle, \quad h \in \mathbb{R}^D,\ i \in \mathbb{N}.$$

Proof. It is easy to see that

$$\mathbf{E}_{\nu,P}|S(f) - S_{n,n_0}(f)|^2 = \frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n} \mathbf{E}_{\nu,P}[g(X_{n_0+j})g(X_{n_0+i})]$$

$$\underset{(2.3)}{=} \frac{1}{n^2}\sum_{j=1}^{n}\sum_{x\in D} P^{n_0+j}(g^2)(x)\,\nu(x) + \frac{2}{n^2}\sum_{j=1}^{n-1}\sum_{k=j+1}^{n}\sum_{x\in D} P^{n_0+j}(gP^{k-j}g)(x)\,\nu(x).$$

Recall that reversibility with respect to π is equivalent to self-adjointness (s-a) of P. For every function $h \in \mathbb{R}^D$ and $i \in \mathbb{N}$ the following calculation holds:

$$\begin{aligned}\sum_{x\in D}(P^i h)(x)\,\nu(x) &= \left\langle P^i h, \frac{\nu}{\pi}\right\rangle = \langle P^i h, 1\rangle + \left\langle P^i h, \frac{\nu}{\pi} - 1\right\rangle \\ &\underset{\text{(s-a)}}{=} \langle P^i h, 1\rangle + \left\langle P^i\left(\frac{\nu}{\pi} - 1\right), h\right\rangle = \langle P^i h, 1\rangle + \left\langle (P^i - S)\left(\frac{\nu}{\pi} - 1\right), h\right\rangle \\ &= \sum_{x\in D}(P^i h)(x)\,\pi(x) + \langle d_i, h\rangle.\end{aligned}$$

Formula (2.9) is shown by using the previous calculation for $h = g^2$ and $h = gP^{k-j}g$. ∎

COROLLARY 2.16. *Under the same assumptions as in Proposition 2.15 we have, for $i \in \{1, \dots, |D| - 1\}$,*

$$e_\nu(S_{n,n_0}, u_i)^2 = \frac{1+\beta_i}{n(1-\beta_i)} - \frac{2\beta_i(1-\beta_i^n)}{n^2(1-\beta_i)^2} + \frac{1}{n^2}\sum_{j=1}^{n}\frac{1+\beta_i - 2\beta_i^{n-j+1}}{1-\beta_i} L_{j+n_0}(u_i^2),$$

where

$$L_k(u_i^2) = \sum_{l=1}^{|D|-1} \beta_l^k \left\langle \frac{\nu}{\pi}, u_l\right\rangle \langle u_l, u_i^2\rangle = \sum_{l=1}^{|D|-1} \beta_l^k \langle u_l, u_i^2\rangle \sum_{x\in D} u_l(x)\,\nu(x). \tag{2.10}$$

Proof. By substituting

$$e_\pi(S_n, u_i)^2 = \frac{1+\beta_i}{n(1-\beta_i)} - \frac{2\beta_i(1-\beta_i^n)}{n^2(1-\beta_i)^2}$$

and

$$\sum_{j=1}^{n-1}\sum_{k=j+1}^{n} L_{j+n_0}(u_i P^{k-j} u_i) = \sum_{j=1}^{n-1}\sum_{k=j+1}^{n} \beta_i^{k-j} L_{j+n_0}(u_i^2) = \sum_{j=1}^{n-1} \frac{L_{j+n_0}(u_i^2)(\beta_i - \beta_i^{n-j+1})}{1-\beta_i}$$

into (2.9) one obtains the error formula. The equality for $L_k(u_i^2)$ is an implication of (2.8). ∎

Equation (2.9) and the result of Corollary 2.16 are still exact error formulas. To get an upper bound for the error, we estimate the functional $L_k(\cdot)$. This estimate depends on the speed of convergence of νP^k to π.

LEMMA 2.17. *Let $h \in \mathbb{R}^D$, $k \in \mathbb{N}$ and recall that $\beta = \max\{\beta_1, |\beta_{|D|-1}|\}$. Then*

$$|L_k(h)| \le \beta^k \|\nu/\pi - 1\|_2 \|h\|_2 \le \beta^k \|\nu/\pi - 1\|_2 \sqrt{\|1/\pi\|_\infty}\, \|h\|_1. \tag{2.11}$$

Proof. After applying the Cauchy–Schwarz inequality (CS) to $L_k(h) = \langle d_k, h\rangle$ one obtains

$$|L_k(h)| \underset{(\mathrm{CS})}{\le} \|d_k\|_2 \|h\|_2 \le \|P^k - S\|_{\ell_2 \to \ell_2} \|\nu/\pi - 1\|_2 \|h\|_2.$$

By Lemma 2.8 the first inequality is proven and the rest is shown by using $\|h\|_2 \le \sqrt{\|1/\pi\|_\infty}\, \|h\|_1$. ∎

The last lemma ensures an exponential decay of $L_k(\cdot)$ for increasing $k \in \mathbb{N}$. This fact is used to show that there exists a constant $C_{\nu,\pi,\beta}$, which is independent of n and n_0, such that

$$|e_\nu(S_{n,n_0}, f)^2 - e_\pi(S_n, f)^2| \le C_{\nu,\pi,\beta}\, \|f\|_2^2\, \frac{\beta^{n_0}}{n^2}.$$

An immediate consequence of the inequality is an explicit error bound. The following two lemmas imply such an inequality and provide $C_{\nu,\pi,\beta}$ explicitly.

LEMMA 2.18. *Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with transition matrix P and initial distribution ν. Let P be reversible with respect to π. Let $f \in \mathbb{R}^D$ and*

$$U(\beta, n) = \sum_{j=1}^{n} \beta^j + 2\sum_{j=1}^{n-1}\sum_{k=j+1}^{n} \beta^k.$$

Then

$$|e_\nu(S_{n,n_0}, f)^2 - e_\pi(S_n, f)^2| \le U(\beta, n) \sqrt{\|1/\pi\|_\infty}\, \|\nu/\pi - 1\|_2 \|f\|_2^2\, \frac{\beta^{n_0}}{n^2}. \tag{2.12}$$

Proof. Let $g = f - S(f)$. The equation (2.9) implies

$$|e_\nu(S_{n,n_0}, f)^2 - e_\pi(S_n, f)^2| \le \frac{1}{n^2}\sum_{j=1}^{n} |L_{j+n_0}(g^2)| + \frac{2}{n^2}\sum_{j=1}^{n-1}\sum_{k=j+1}^{n} |L_{j+n_0}(g P^{k-j} g)|.$$

Then by (2.11) one gets

$$|L_{j+n_0}(g^2)| \le \beta^{j+n_0} \sqrt{\|1/\pi\|_\infty}\, \|\nu/\pi - 1\|_2 \|g\|_2^2,$$
$$|L_{j+n_0}(g P^{k-j} g)| \le \beta^{j+n_0} \sqrt{\|1/\pi\|_\infty}\, \|\nu/\pi - 1\|_2 \|g P^{k-j} g\|_1.$$

By the Cauchy–Schwarz inequality (CS) and $\|P^{k-j}\|_{\ell_2^0 \to \ell_2^0} = \beta^{k-j}$ it follows that

$$\|g P^{k-j} g\|_1 \underset{(\mathrm{CS})}{\le} \|g\|_2 \|P^{k-j} g\|_2 \le \|g\|_2^2 \|P^{k-j}\|_{\ell_2^0 \to \ell_2^0} = \beta^{k-j} \|g\|_2^2.$$

Let $\varepsilon_0 = \sqrt{\|1/\pi\|_\infty}\,\|\nu/\pi - 1\|_2 \beta^{n_0}$. Then

$$\begin{aligned}\sum_{j=1}^{n} |L_{j+n_0}(g^2)| + 2\sum_{j=1}^{n-1}\sum_{k=j+1}^{n} |L_{j+n_0}(gP^{k-j}g)| &\le \varepsilon_0\|g\|_2^2 \sum_{j=1}^{n}\beta^j + 2\varepsilon_0\|g\|_2^2 \sum_{j=1}^{n-1}\sum_{k=j+1}^{n}\beta^k \\ &= \varepsilon_0\|g\|_2^2\Big(\sum_{j=1}^{n}\beta^j + 2\sum_{j=1}^{n-1}\sum_{k=j+1}^{n}\beta^k\Big) \\ &\le U(\beta,n)\cdot\varepsilon_0\|f\|_2^2.\end{aligned}$$

The last inequality follows from $\|f - S(f)\|_2 \le \|f\|_2$. ∎

Note that one can also get a similar estimate as in (2.12) with respect to $\|f\|_4$ by using the first inequality of (2.11) instead of the second one. In the resulting estimate the factor $\sqrt{\|1/\pi\|_\infty}$ does not appear.

Let us consider $U(\beta, n)$. If $\beta < 1$, then the mapping $n \mapsto U(\beta, n)$ is bounded.

LEMMA 2.19. *Let $\beta < 1$. For all $n \in \mathbb{N}$ we have*

$$U(\beta,n) \le \frac{2}{(1-\beta)^2}.$$

Proof. By the infinite geometric series one obtains

$$U(\beta,n) \le \sum_{j=1}^{n}\beta^j + \frac{2\beta}{1-\beta}\sum_{j=1}^{n-1}\beta^j \le \frac{1+\beta}{1-\beta}\sum_{j=1}^{n}\beta^j \le \frac{2}{(1-\beta)^2}.$$ ∎

From Lemmas 2.18 and 2.19 it follows that

$$e_\nu(S_{n,n_0},f)^2 \le e_\pi(S_n,f)^2 + \frac{2\beta^{n_0}\sqrt{\|1/\pi\|_\infty}\,\|\nu/\pi - 1\|_2}{n^2(1-\beta)^2}\|f\|_2^2.$$

If the initial distribution ν is π then the error can be represented as in Proposition 2.11 and bounded as in Corollary 2.13.

The next theorem summarizes the main result of this section.

THEOREM 2.20. *Let $f \in \mathbb{R}^D$ and let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with transition matrix P and initial distribution ν. Let P be reversible with respect to π and let $\beta < 1$. Then*

$$e_\nu(S_{n,n_0},f)^2 \le \frac{2}{n(1-\beta_1)}\|f\|_2^2 + \frac{2\beta^{n_0}\sqrt{\|1/\pi\|_\infty}\|\nu/\pi - 1\|_2}{n^2(1-\beta)^2}\|f\|_2^2. \tag{2.13}$$

For $a_k = \langle f, u_k\rangle$ one has

$$\lim_{n\to\infty} n e_\nu(S_{n,n_0},f)^2 = \lim_{n\to\infty} n e_\pi(S_n,f)^2 = \sum_{k=1}^{|D|-1} a_k^2\,\frac{1+\beta_k}{1-\beta_k}. \tag{2.14}$$

Proof. By Lemma 2.18, Corollary 2.13 and Lemma 2.19 the estimate of (2.13) is proven. By Lemmas 2.18 and 2.19 the first equality of (2.14) holds. Then, by Proposition 2.11,

$$\lim_{n\to\infty} n e_\pi(S_n,f)^2 = \sum_{k=1}^{|D|-1} a_k^2\,\frac{1+\beta_k}{1-\beta_k}.$$ ∎

REMARK 2.21. The error bound (2.13) can be interpreted as follows: The burn-in n_0 is necessary to eliminate the influence of the initial distribution ν, while n must be large

to decrease $e_\pi(S_n, f)$. Unfortunately the dependence of the initial distribution on the estimate is disillusioning for an extension to general state spaces, because of the factor $\sqrt{\|1/\pi\|_\infty}$. One can avoid this factor if one considers error bounds with respect to $\|f\|_p$ with $p > 2$ (see Section 3.2).

Another consequence of Lemmas 2.18 and 2.19 is the following result concerning the asymptotic error for $\|f\|_2 \le 1$.

COROLLARY 2.22. *Under the same assumptions as in Theorem* 2.20*,*

$$\lim_{n\to\infty} n \sup_{\|f\|_2\le 1} e_\nu(S_{n,n_0}, f)^2 = \frac{1+\beta_1}{1-\beta_1}$$

and

$$\lim_{n_0\to\infty} \sup_{\|f\|_2\le 1} e_\nu(S_{n,n_0}, f)^2 = \frac{1+\beta_1}{n(1-\beta_1)} - \frac{2\beta_1(1-\beta_1^n)}{n^2(1-\beta_1)^2}.$$

Proof. Let us define

$$c_{n,n_0} = \frac{2\beta^{n_0}\sqrt{\|1/\pi\|_\infty}\,\|\nu/\pi - 1\|_2}{n^2(1-\beta)^2}.$$

One has $\lim_{n\to\infty} n \cdot c_{n,n_0} = 0$ and $\lim_{n_0\to\infty} c_{n,n_0} = 0$. For $\|f\|_2 \le 1$, by Lemmas 2.18 and 2.19,

$$|e_\nu(S_{n,n_0}, f)^2 - e_\pi(S_n, f)^2| \le c_{n,n_0}.$$

Hence

$$\sup_{\|f\|_2\le 1} e_\pi(S_n, f)^2 - c_{n,n_0} \le \sup_{\|f\|_2\le 1} e_\nu(S_{n,n_0}, f)^2 \le \sup_{\|f\|_2\le 1} e_\pi(S_n, f)^2 + c_{n,n_0}. \tag{2.15}$$

By Corollary 2.13 we have

$$\sup_{\|f\|_2\le 1} e_\pi(S_n, f)^2 = \frac{1+\beta_1}{n(1-\beta_1)} - \frac{2\beta_1(1-\beta_1^n)}{n^2(1-\beta_1)^2}.$$

Taking the limits in (2.15) yields the assertions. ∎

REMARK 2.23. The number

$$\tau = \frac{1+\beta_1}{1-\beta_1}$$

is called the *autocorrelation time* of P (see [Sok97, Mat99]). If one could sample from π then $\beta_1 = 0$ so that $\tau = 1$. Hence τ is the factor of computing time which quantifies the asymptotic difference of Markov chain Monte Carlo compared to Monte Carlo with an i.i.d. sample from the distribution π.

REMARK 2.24. Observe that one obtains from (2.15) a lower error bound for S_{n,n_0}. We have

$$\frac{1+\beta_1}{n(1-\beta_1)} - \frac{2}{n^2(1-\beta_1)^2} - c_{n,n_0} \le \sup_{\|f\|_2\le 1} e_\nu(S_{n,n_0}, f)^2 \le \frac{2}{n(1-\beta_1)} + c_{n,n_0}$$

with c_{n,n_0} defined as in the proof of Corollary 2.22. For a reasonable burn-in of the Markov chain the error can be effectively approximated by these estimates. We apply these estimates to illustrating examples (see Section 2.4). Now let us discuss which burn-in is reasonable.

2.3. Burn-in. Assume that computer resources for N steps of the Markov chain are available, i.e. $N = n + n_0$. The goal is to choose the burn-in n_0 and the number n such that the upper error bound is as small as possible. There is obviously a trade-off between the choice of n and n_0. In the next statement the error for an explicitly given burn-in is stated.

THEOREM 2.25. *Suppose that*

$$n_0 = \max\left\{\left\lceil \frac{\log(\sqrt{\|1/\pi\|_\infty}\,\|\nu/\pi - 1\|_2)}{\log(\beta^{-1})}\right\rceil, 0\right\}.$$

Then

$$\frac{1+\beta_1}{n(1-\beta_1)} - \frac{4}{n^2(1-\beta)^2} \le \sup_{\|f\|_2\le 1} e_\nu(S_{n,n_0}, f)^2 \le \frac{2}{n(1-\beta_1)} + \frac{2}{n^2(1-\beta)^2}.$$

Proof. The assertion follows from Theorem 2.20 and Remark 2.24. ∎

Note that $\log(\beta^{-1}) = (1-\beta) + \sum_{j=2}^\infty \frac{(1-\beta)^j}{j!}$ and $\log(\beta^{-1}) \ge 1-\beta$. One might use this observation to estimate the suggested burn-in. The choice of the burn-in of Theorem 2.25 is justified by the following.

Let us define

$$C = \sqrt{\|1/\pi\|_\infty}\,\|\nu/\pi - 1\|_2$$

and assume that $\beta_1 = \beta$. If the assumption does not hold we may estimate the error bound of Theorem 2.20 by using $(1-\beta_1)^{-1} \le (1-\beta)^{-1}$. For $\|f\|_2 \le 1$ we want to minimize the error estimate

$$\mathrm{est}(n, n_0) = \sqrt{\frac{2}{n(1-\beta)} + \frac{2C\beta^{n_0}}{n^2(1-\beta)^2}} \quad \text{under the constraint that} \quad N = n + n_0.$$

LEMMA 2.26. *For $\eta > 0$ let*

$$C > \left(\frac{\log(\beta^{-1})}{1-\beta}\right)^{1/\eta}, \tag{2.16}$$

$$N > (1+\eta)\,\frac{\log(C)}{\log(\beta^{-1})} + 2[\log(\beta^{-1}) - (1-\beta)]^{-1}. \tag{2.17}$$

Then there exists an

$$n_{\mathrm{opt}} \in \left[\frac{\log(C)}{\log(\beta^{-1})}, (1+\eta)\frac{\log(C)}{\log(\beta^{-1})}\right],$$

which minimizes the mapping $n_0 \mapsto \mathrm{est}(N - n_0, n_0)$.

If $\eta = 10^{-3}$, then (2.16) implies for $\beta = 0.99$ that $C > 152$ and for $C = 10^{30}$ that $\beta > 0.87$. Hence the assumptions are not restrictive, since β is usually close to 1, C is large ([2]) and the computational resources N should be sufficiently large.

Proof. Let

$$a = N - (1+\eta)\frac{\log(C)}{\log(\beta^{-1})} \quad \text{and} \quad b = N - \frac{\log(C)}{\log(\beta^{-1})}.$$

([2]) The constant C might depend exponentially on additional parameters: see the example "Random walk on the hypercube" in Section 2.4, or see Section 4.

Note that (2.17) gives that $b > a > 0$. It is enough to show that there exists an $m_{\mathrm{opt}} \in [a, b]$ which minimizes $n \mapsto \mathrm{est}^2(n)$ given by

$$\mathrm{est}^2(n) = (\mathrm{est}(n, N-n))^2 = \frac{2}{n(1-\beta)} + \frac{2C\beta^{N-n}}{n^2(1-\beta)^2}.$$

We have

$$\mathrm{est}^2(n)' = \frac{d}{dn}\,\mathrm{est}^2(n) = \frac{2}{n^2(1-\beta)}\left[\frac{C\beta^{N-n}}{(1-\beta)}\left(\log(\beta^{-1}) - \frac{2}{n}\right) - 1\right].$$

We will show for any $\tilde{a} \le a$ and $\tilde{b} \ge b$ that

$$\mathrm{est}^2(\tilde{a})' < 0 \quad \text{and} \quad \mathrm{est}^2(\tilde{b})' > 0,$$

consequently there exists an $m_{\mathrm{opt}} \in [a, b]$ which minimizes $\mathrm{est}^2(n)$. Let $\tilde{b} \ge b$. Then the inequality $\mathrm{est}^2(\tilde{b})' > 0$ follows by (2.17) and

$$\begin{aligned}
N > \frac{\log(C)}{\log(\beta^{-1})} & \frac{\left(\frac{2}{\log(C)} + \left(1 - \frac{1-\beta}{\log(\beta^{-1})}\right)\right)}{\left(1 - \frac{1-\beta}{\log(\beta^{-1})}\right)} \\
&\Leftrightarrow\ N > \frac{(1 - \frac{1-\beta}{\log(\beta^{-1})})\log(C) + 2}{\log(\beta^{-1}) - (1-\beta)} \\
&\Leftrightarrow\ N(\log(\beta^{-1}) - (1-\beta)) + \log(C)\left(\frac{1-\beta}{\log(\beta^{-1})} - 1\right) > 2 \\
&\Leftrightarrow\ \left(N - \frac{\log(C)}{\log(\beta^{-1})}\right)\log(\beta^{-1}) - \left(N - \frac{\log(C)}{\log(\beta^{-1})}\right)(1-\beta) > 2 \\
&\Leftrightarrow\ b\log(\beta^{-1}) - b(1-\beta) > 2 \\
&\Leftrightarrow\ \log(\beta^{-1}) - \frac{2}{b} > (1-\beta) \\
&\Leftrightarrow\ \frac{1}{1-\beta}\left(\log(\beta^{-1}) - \frac{2}{b}\right) - 1 = \frac{C\beta^{N-b}}{1-\beta}\left(\log(\beta^{-1}) - \frac{2}{b}\right) - 1 > 0 \\
&\Rightarrow\ \frac{C\beta^{N-\tilde{b}}}{1-\beta}\left(\log(\beta^{-1}) - \frac{2}{\tilde{b}}\right) - 1 > 0.
\end{aligned}$$

On the other hand for $\tilde{a} \le a$ we obtain $\mathrm{est}^2(\tilde{a})' < 0$. This is shown by the following calculation. By (2.16) one has

$$\begin{aligned}
C^\eta \underset{(2.16)}{>} & \frac{\log(\beta^{-1})}{(1-\beta)} \\
&\Leftrightarrow\ \log(\beta^{-1}) - (1-\beta)C^\eta < 0 \\
&\Rightarrow\ \log(\beta^{-1}) - (1-\beta)C^\eta < \frac{2}{a} \\
&\Leftrightarrow\ \log(\beta^{-1}) - \frac{2}{a} < (1-\beta)C^\eta \\
&\Leftrightarrow\ \frac{C^{-\eta}}{(1-\beta)}\left(\log(\beta^{-1}) - \frac{2}{a}\right) - 1 = \frac{C\beta^{N-a}}{(1-\beta)}\left(\log(\beta^{-1}) - \frac{2}{a}\right) - 1 < 0 \\
&\Rightarrow\ \frac{C\beta^{N-\tilde{a}}}{(1-\beta)}\left(\log(\beta^{-1}) - \frac{2}{\tilde{a}}\right) - 1 < 0.
\end{aligned}$$

Altogether this implies that there is an

$$n_{\mathrm{opt}} \in \left[\frac{\log(C)}{\log(\beta^{-1})}, (1+\eta)\frac{\log(C)}{\log(\beta^{-1})}\right]$$

which minimizes the mapping $n_0 \mapsto \mathrm{est}(N-n_0, n_0)$. ∎

If an error of at most $\varepsilon \in (0,1)$ is desired, then the suggested choice of the burn-in n_0 is independent of the precision ε: we choose

$$n_0 = \max\left\{\left\lceil \frac{\log(\sqrt{\|1/\pi\|_\infty}\,\|\nu/\pi - 1\|_2)}{\log(\beta^{-1})}\right\rceil, 0\right\}$$

and

$$n \geq \frac{1+\sqrt{1+4\varepsilon^2}}{(1-\beta)\varepsilon^2} \quad \text{to achieve} \quad e_\nu(S_{n,n_0}, f) \leq \varepsilon.$$

2.4. Examples. The goal is to compare the upper bounds of Theorems 2.20 and 2.25 with the exact error for a given function $f \in \mathbb{R}^D$. It is not known which f with $\|f\|_2 \leq 1$ maximizes $e_\nu(S_{n,n_0}, f)^2$. But by Corollary 2.22 one has

$$\lim_{n_0\to\infty} \sup_{\|f\|_2\leq 1} e_\nu(S_{n,n_0}, f)^2 = e_\pi(S_n, u_1)^2,$$

where u_1 is the eigenfunction corresponding to β_1. This motivates the study of the individual error for u_1, which gives the maximal error for integrands f with $\|f\|_2 \leq 1$ if n_0 goes to infinity. In this section illustrating examples are considered, where the eigenvalues and the eigenfunctions are available. The Markov chains are very well studied in the literature (see [Mei99, SC04, Str05, BD06, LPW09]).

Random walk on a circle. Let $T \geq 3$ be an odd natural number. Let $D = \mathbb{Z}_T$ be the underlying state space, where $\mathbb{Z}_T = \mathbb{Z} \bmod T$ denotes the cyclic group of order T. The $T \times T$ transition matrix of the random walk is

$$p(x,y) = \begin{cases} 1/2, & y = x \pm 1 \bmod T, \\ 0, & \text{otherwise.} \end{cases}$$

The transition matrix is reversible with respect to the uniform distribution given by $\pi(x) = 1/T$ for $x \in D$. Since T is an odd number, the transition matrix is aperiodic; for even T it would be periodic. The eigenvalues of the transition matrix are

$$\beta_0 = 1, \quad \beta_{2j-1} = \beta_{2j} = \cos(2\pi j/T), \quad j = 1, \ldots, (T-1)/2,$$

and the orthonormal eigenfunctions $\{u_0, u_1, \ldots, u_{T-1}\}$ are

$$u_0(x) = 1, \quad u_{2j-1}(x) = \sqrt{2}\cos(2\pi\, jx/T), \quad u_{2j}(x) = \sqrt{2}\sin(2\pi\, jx/T),$$

where $j = 1, \ldots, (T-1)/2$ and $x \in D$. Clearly $\beta = |\beta_{T-1}| = \cos(\pi/T)$, thus $\beta \neq \beta_1$.

Let us consider $f = u_1$. The initial distribution is chosen as $\nu = \delta_0$, so that the initial state is $0 \in D$. As $(u_1)^2 = u_0 + \frac{1}{\sqrt{2}}u_3$ we have

$$\langle u_i, (u_1)^2\rangle = \langle u_i, u_0\rangle + \frac{1}{\sqrt{2}}\langle u_i, u_3\rangle = \begin{cases} 1, & i = 0, \\ 1/\sqrt{2}, & i = 3, \\ 0, & \text{otherwise.} \end{cases}$$

Hence by (2.10) we obtain

$$L_k((u_1)^2) = \sum_{i=1}^{T-1} \beta_i^k \langle u_i, (u_1)^2 \rangle u_i(0) = \beta_3^k.$$

Additionally with $\beta_1 = \cos(2\pi/T)$,

$$e_\pi(S_n, u_1)^2 = \frac{1+\cos(2\pi/T)}{n(1-\cos(2\pi/T))} - \frac{2\cos(2\pi/T)(1-\cos^n(2\pi/T))}{n^2(1-\cos(2\pi/T))^2}.$$

The exact error is determined by Corollary 2.16 with $\beta_3 = \cos(4\pi/T)$ so that

$$e_\nu(S_{n,n_0}, u_1) = \Bigg(e_\pi(S_n, u_1)^2 + \frac{1}{n^2} \sum_{j=1}^{n} \left(\frac{1+\cos(2\pi/T) - 2\cos^{n-j+1}(2\pi/T)}{1-\cos(2\pi/T)} \right) \cos^{j+n_0}(4\pi/T) \Bigg)^{1/2}. \quad (2.18)$$

We apply Theorem 2.25 to get a lower error bound and (2.13) of Theorem 2.20 to get an upper error bound, since $\beta \neq \beta_1$. Hence the burn-in is chosen as suggested in Theorem 2.25, i.e.

$$n_0 = \left\lceil \frac{1}{2} \frac{\log(T^2 - T)}{\log(\cos^{-1}(\pi/T))} \right\rceil.$$

Then

$$e_\nu(S_{n,n_0}, u_1) \leq \left(\frac{2}{n(1-\cos(2\pi/T))} + \frac{2}{n^2(1-\cos(\pi/T))^2} \right)^{1/2} \quad (2.19)$$

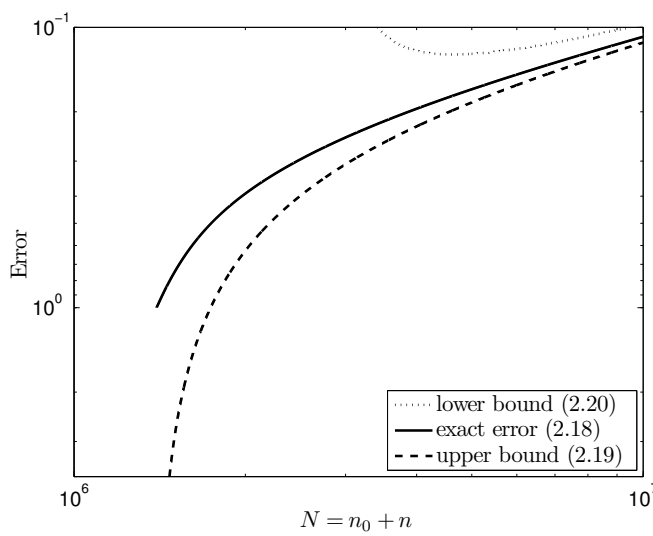

Fig. 1. Random walk on a circle: Exact error and error bounds, $T = 999$ and $n_0 = \lceil \frac{1}{2} \frac{\log(T^2-T)}{\log(\cos^{-1}(\pi/T))} \rceil = 1396699$.

and

$$\left(\frac{1+\cos(2\pi/T)}{n(1-\cos(2\pi/T))}-\frac{4}{n^2(1-\cos(\pi/T))^2}\right)^{1/2}\le e_\nu(S_{n,n_0},u_1). \tag{2.20}$$

We have an explicit exact error formula (2.18), a lower error bound (2.20) and an upper error bound (2.19).

In Figure 1 the different bounds of (2.19), (2.20) and the exact error of (2.18) are plotted for $T = 999$. The curves start at $N = n_0$, since the computational resources must be larger than the burn-in $n_0 = 1396699$. The lower error bound gives a non-trivial estimate if $N \ge n_0 + 1617911 = 3014610$, since for $n \ge \frac{4(1-\beta_1)}{(1+\beta_1)(1-\beta)^2} = 1617911$ one obtains a lower bound larger as zero.

Random walk on the hypercube. Let d be a natural number. Let $D = \{0,1\}^d$ be the state space and $|\tilde{x}| = \sum_{i=1}^d |\tilde{x}_i|$ for $\tilde{x} \in \{-1,0,1\}^d$. The $2^d \times 2^d$ transition matrix is given by

$$p(x,y) = \begin{cases} 1/2, & x = y, \\ 1/(2d), & |x-y| = 1, \\ 0, & \text{otherwise.} \end{cases}$$

The transition matrix is reversible with respect to $\pi(x) = 2^{-d}$ for $x \in D$. Furthermore, it is aperiodic and irreducible. We use a different notation for the index of the eigenvalues and orthonormal eigenfunctions, for $z \in \{0,1\}^d$ one has

$$\beta_z = 1 - |z|/d \quad \text{and} \quad u_z(x) = (-1)^{\sum_{i=1}^d z_i x_i}, \quad x \in D.$$

Set $[0] = (0,\dots,0)$ and set $[1] = (1,0,\dots,0)$ so that

$$\begin{aligned} \beta_{[0]} &= 1, & u_{[0]}(x) &= 1, \quad x \in D, \\ \beta_{[1]} &= 1 - 1/d, & u_{[1]}(x) &= (-1)^{x_1}, \quad x \in D. \end{aligned}$$

Obviously for all indices $z \in \{0,1\}^d$ we have $\beta_z \ge 0$ so that $\beta_{[1]} = \beta$.

Let us choose the initial state of the Markov chain deterministically in $(0,\dots,0) \in D$, i.e. $\nu = \delta_{[0]}$. By $(u_{[1]})^2 = u_{[0]}$ one has, for $z \in \{0,1\}^d$,

$$\langle u_z, (u_{[1]})^2 \rangle = \begin{cases} 1, & z = [0], \\ 0, & \text{otherwise.} \end{cases}$$

This implies

$$L_k((u_{[1]})^2) = 0, \quad k \in \mathbb{N}.$$

The error of S_n, if the initial state is chosen according to π, obeys

$$e_\pi(S_n, u_{[1]})^2 = \frac{2d-1}{n} - \frac{2(d^2-d)}{n^2}\left(1 - \left(1 - \frac{1}{d}\right)^n\right).$$

Then by Corollary 2.16,

$$e_\nu(S_{n,n_0}, u_{[1]}) = e_\pi(S_n, u_{[1]}). \tag{2.21}$$

The burn-in and the error bounds are determined by Theorem 2.25. One obtains

$$n_0 = \left\lceil \frac{1}{2} \frac{\log(2^{2d} - 2^d)}{\log(1-1/d)^{-1}} \right\rceil$$

such that

$$e_\nu(S_{n,n_0}, u_{[1]}) \le \sqrt{\frac{2d}{n} + \frac{2d^2}{n^2}}, \tag{2.22}$$

and

$$\sqrt{\frac{2d-1}{n} - \frac{4d^2}{n^2}} \le e_\nu(S_{n,n_0}, u_{[1]}). \tag{2.23}$$

In Figure 2 for $d = 50$ the exact error (2.21), the upper error bound (2.22) and the lower error bound (2.23) are plotted. It can be seen that after the burn-in the curves are close to each other. The error bounds are polynomial in d which is of the magnitude of $\log(|D|)$.

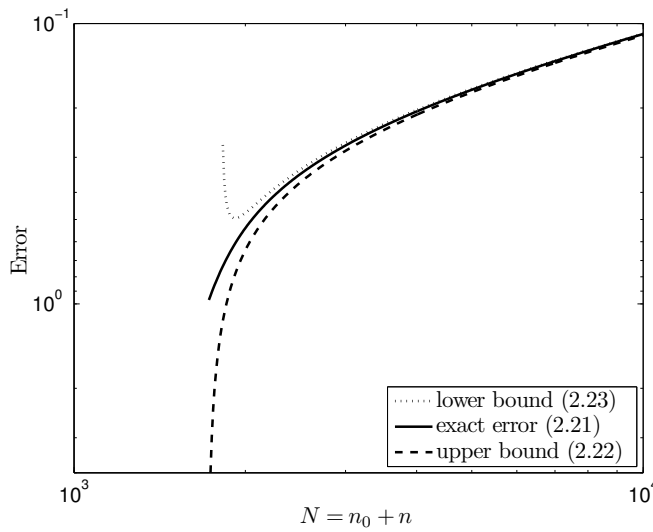

Fig. 2. Random walk on the hypercube: Exact error and error bounds, $d = 50$ and $n_0 = \lceil \frac{1}{2} \frac{\log(2^{2d} - 2^d)}{\log(1-1/d)^{-1}} \rceil = 1716$.

Random walk on the star. Let $T \ge 2$ be an even natural number. Let the state space be $D = \{0, 1, \dots, T\}$. The $(T+1) \times (T+1)$ transition matrix is given by

$$p(x,y) = \begin{cases} \theta, & x = y = 0, \\ \dfrac{1-\theta}{T}, & x = 0,\ y \in D \setminus \{0\}, \\ 1, & x \in D \setminus \{0\},\ y = 0, \\ 0, & \text{otherwise}, \end{cases}$$

with a parameter $\theta \in (0,1)$. The transition graph is star shaped since every state is connected solely to the center 0. The transition matrix is reversible with respect to π given for $x \in D$ by

$$\pi(x) = \begin{cases} \dfrac{1}{2-\theta}, & x = 0, \\ \dfrac{1-\theta}{T(2-\theta)}, & \text{otherwise}. \end{cases}$$

One obtains $\beta_0 = 1$, $\beta_T = \theta - 1$ and for $x \in D$ one has

$$u_0(x) = 1, \quad u_T(x) = \sqrt{1-\theta} \begin{cases} 1, & x = 0, \\ \dfrac{1}{\theta - 1}, & \text{otherwise.} \end{cases}$$

The eigenvalue $\beta_i = 0$ for $i \in \{1, \ldots, T-1\}$ is of multiplicity $T - 1$. Without loss of generality we may assume that for any $x \in D$ one has

$$u_1(x) = \begin{cases} 0, & x = 0, \\ \sqrt{\dfrac{2-\theta}{1-\theta}}, & x = 1, \ldots, T/2, \\ -\sqrt{\dfrac{2-\theta}{1-\theta}}, & x = T/2+1, \ldots, T. \end{cases}$$

The remaining eigenvectors $u_2, \ldots, u_{T-1}$ are arbitrarily chosen such that we get an orthonormal basis $\{u_0, u_1, \ldots, u_T\}$. One has an aperiodic and irreducible transition matrix where $\beta_1 = 0$ and $\beta = \max\{\beta_1, |\beta_T|\} = 1 - \theta$. We consider the error for $f = u_1$. The initial state is the center of the star, i.e. 0. Then $\nu = \delta_0$. From $(u_1)^2 = u_0 - \frac{1}{\sqrt{1-\theta}}\, u_T$ one gets

$$\langle u_i, (u_1)^2 \rangle = \begin{cases} 1, & i = 0, \\ -\dfrac{1}{\sqrt{1-\theta}}, & i = T, \\ 0, & \text{otherwise.} \end{cases}$$

By (2.10) this implies

$$L_k((u_1)^2) = \sum_{i=1}^{T} \beta_i^k \langle u_i, (u_1)^2 \rangle u_i(0) = -\beta_T^k = -(\theta - 1)^k.$$

The error where the Markov chain is initialized by the stationary distribution obeys

$$e_\pi(S_n, u_1)^2 = \frac{1}{n}.$$

Then by Corollary 2.16 it follows that

$$e_\nu(S_{n,n_0}, u_1) = \left(\frac{1}{n} - \frac{(\theta-1)^{n_0+1}((\theta-1)^n - 1)}{(\theta-2)n^2} \right)^{1/2}. \tag{2.24}$$

Recall that $\beta_1 \neq \beta$. However, we only use the error bounds of Theorem 2.25. The burn-in is chosen as

$$n_0 = \left\lceil \frac{\log((2-\theta)T)}{2\log(1-\theta)^{-1}} \right\rceil.$$

Then the upper bound is

$$e_\nu(S_{n,n_0}, u_1) \leq \sqrt{\frac{2}{\theta n} + \frac{2}{\theta^2 n^2}}, \tag{2.25}$$

and the lower bound is

$$\sqrt{\frac{1}{n} - \frac{4}{\theta^2 n^2}} \leq e_\nu(S_{n,n_0}, u_1). \tag{2.26}$$

In Figure 3 for $\theta = 0.1$ and $T = 10^5$ the exact error (2.24), the upper error bound (2.25) and the lower bound (2.26) are plotted. For $n \geq 4/\theta^2$ we get a non-trivial estimate by the

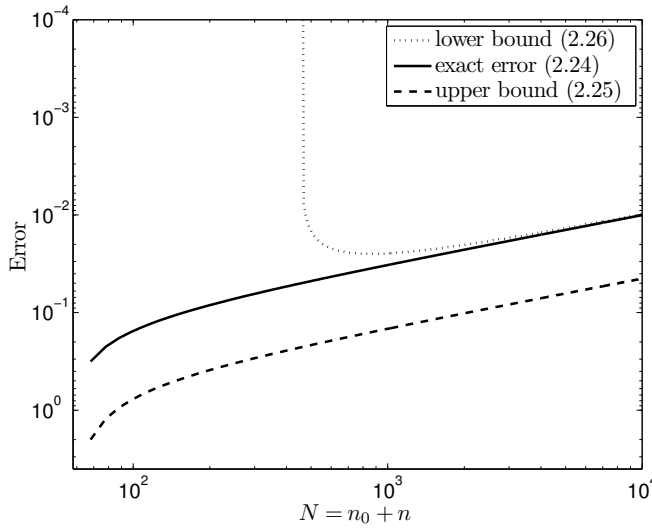

Fig. 3. Random walk on the star: Exact error and error bounds, $\theta = 0.1$, $T = 10^5$ and $n_0 = \lceil \frac{\log((2-\theta)T)}{2\log(1-\theta)^{-1}} \rceil = 58$.

lower bound. The upper error bound is shifted down since $\beta \neq \beta_1$. One could improve this by using (2.13) of Theorem 2.20 directly. In the present setting one loses asymptotically a factor of $\sqrt{2/\theta}$.

Let us summarize the important facts of this section. The error was considered for the eigenfunction u_1 corresponding to β_1. If n_0 goes to infinity, then u_1 is the function which maximizes the error for integrands f with $\|f\|_2 \leq 1$. The bound of Theorem 2.25 applied in this setting gives tight results if $\beta_1 = \beta$. Otherwise Theorem 2.20 achieves the right asymptotic coefficient if β_1 and β are known. For the examples considered, one knows the eigenvalues and the eigenfunctions explicitly. In applications it is usually difficult to estimate β_1 or β, but there are different auxiliary tools, e.g. the canonical path technique, conductance (see [JS89] and [DS91]), log-Sobolev inequalities and path coupling (see [LPW09]). However, if the eigenvalues β_1 and $\beta_{|D|-1}$ are available, then the error can be approximated by the lower and upper bound.

2.5. Notes and remarks. Let us comment on how the results fit into the published literature. An elementary and powerful technique to bound the error for S_{n,n_0} or S_n is based on Doeblin's theory (see [Str05, p. 27]). Let $A_k = (a_k(x,y))_{x,y\in D}$ be the kth Cesàro sum given by

$$A_k = \frac{1}{k}\sum_{j=0}^{k-1} P^j.$$

Assume that

$$\exists M \in \mathbb{N} \setminus \{1\},\, y_0 \in D \text{ and } \gamma > 0 \quad \text{such that} \quad \forall x \in D:\ a_M(x, y_0) \geq \gamma. \tag{2.27}$$

Then for any n_0 the error obeys

$$e_\nu(S_{n,n_0}, f)^2 \leq \frac{8(M-1)}{n\gamma} \|f\|_\infty^2.$$

Condition (2.27) states that there is a state y_0 where the expected value of visiting it, on average, until M from any other state is uniformly bounded from below with rate γ. If the transition matrix is irreducible then there exists an M such that $A_M > 0$ and (2.27) is satisfied (see for example [Beh00, Lemma 7.3, p. 50]). It is difficult to obtain γ and M. Let us state a toy example where one can compute γ and M explicitly. Let $D = \{0,1\}^d$. We consider a Markov chain which independently samples with respect to π with $\pi(x) = 2^{-d}$ for $x \in D$. This is Monte Carlo with an i.i.d. sample. Consequently, we get best possible parameters $\gamma = 2^{-d-1}$ and $M = 2$. The error estimate behaves exponentially badly in terms of d. In contrast, the estimate of Theorem 2.20 is independent of d. In general, even if one can get γ and M, these constants are often exponentially bad in terms of some other parameters. Usually γ is close to zero and M is huge. However, with this bound even the periodic case is covered and reversibility is not necessary. But on the other hand the optimal coefficient $\frac{1+\beta_1}{1-\beta_1}$ of the leading term of Corollary 2.13 is not reached and the burn-in cannot be used to tune the algorithm.

The approach to use the spectral representation of reversible transition matrices is not new. In [BD06] the result of Proposition 2.11 is presented. By the same arguments a slightly worse bound is shown in [Ald87, Proposition 4.1, p. 40]. It applies if $\beta_1 \geq 0$ and gives

$$e_\pi(S_n, f)^2 \leq \frac{2}{n(1-\beta_1)} \|f\|_2^2 + \frac{2\exp\{-n(1-\beta_1)\}}{n^2(1-\beta_1)^2} \|f\|_2^2. \tag{2.28}$$

Furthermore if the initial distribution ν is not stationary, a different algorithm is considered. Namely, the burn-in n_0^* is randomly chosen, independent of $(X_n)_{n\in\mathbb{N}}$, according to the Poisson distribution with parameter n_0, and

$$S_{n,n_0}^*(f) = \frac{1}{n} \sum_{j=1}^{n} f(X_{j+n_0^*}).$$

Then it is proven in [Ald87, Proposition 4.2, p. 41] that

$$e_\nu(S_{n,n_0}^*, f)^2 \leq e_\pi(S_n, f)^2 (1 + \|1/\pi\|_\infty \exp\{-n_0(1-\beta_1)\}).$$

This bound applies also for periodic Markov chains and after applying (2.28) it gives an estimate with respect to $\|\cdot\|_2$. The optimal coefficient $\frac{1+\beta_1}{1-\beta_1}$ of the leading term (see Corollary 2.22) is not reached, also if Corollary 2.13 instead of (2.28) is applied. The burn-in n_0^* is randomly chosen rather than deterministically, since then one can translate the discrete time Markov chain into a continuous time Markov chain and avoids discussions of negative eigenvalues. This technique is similar to the idea of considering a lazy Markov chain.

In [NP09] an explicit error bound is given which also holds for non-reversible Markov chains with an absolute ℓ_2-spectral gap, i.e. $\beta = \|P\|_{\ell_2^0 \to \ell_2^0} < 1$. In the proof of the error

bound the multiplicative reversibilization PP^* of P is used, where P^* is the adjoint operator of P acting on ℓ_2. It follows from [NP09, Corollary 4.2, p. 320] that

$$e_\nu(S_{n,n_0},f)^2 \le \frac{1+\beta}{n(1-\beta)}\|f\|_2^2 + \frac{2\beta}{(1-\beta)^2n^2}\|f\|_2^2 + \frac{2(1+\beta)\beta^{n_0}\|\nu/\pi-1\|_2}{(1-\beta)n^2}\|f\|_\infty\|f\|_2.$$

One obtains an error bound uniformly with respect to $\|f\|_2$ by using $\|f\|_\infty \le \|\frac{1}{\pi}\|_\infty^{1/2}\|f\|_2$. The spectral gap can be implied by aperiodicity and irreducibility of the Markov chain, see [LPW09, Lemma 12.1, p. 153]. But it is remarkable that the chain can be non-reversible. If $\beta = \beta_1$ then the error bound has the right coefficient of the leading term. Then it is essentially the same bound as in Theorem 2.20.

Also confidence estimates of S_{n,n_0} are of interest. The goal is to achieve for given precision $\varepsilon \in (0,1)$ and confidence parameter $\alpha \in (0,1)$ that

$$\Pr(|S_{n,n_0}(f) - S(f)| \ge \varepsilon) \le \alpha. \tag{2.29}$$

An estimate of the mean square error implies such confidence estimates.

LEMMA 2.27. *Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with transition matrix P and initial distribution ν and let $\varepsilon \in (0,1)$. Then*

$$\Pr(|S_{n,n_0}(f) - S(f)| \ge \varepsilon) \le \frac{e_\nu(S_{n,n_0},f)^2}{\varepsilon^2}.$$

Proof. The result is an application of the Markov inequality. ∎

Suppose that $\|f\|_2 \le 1$. If one applies Lemma 2.27 and the burn-in is chosen as in Theorem 2.25 then it follows for

$$n_0 \ge \frac{\log(\sqrt{\|1/\pi\|_\infty}\,\|\nu/\pi-1\|_2)}{\log(\beta^{-1})} \quad \text{and} \quad n \ge \frac{4\alpha^{-1}\varepsilon^{-2}}{1-\beta}$$

that (2.29) is true. Note that the burn-in is chosen independently of α. In [LPW09, Theorem 12.19, p. 165] a similar bound is deduced by coupling arguments. It implies a slightly worse result if the initial state is deterministically chosen. If

$$n_0 \ge \frac{\log(2\alpha^{-1}\|1/\pi\|_\infty)}{1-\beta} \quad \text{and} \quad n \ge \frac{4\alpha^{-1}\varepsilon^{-2}}{1-\beta}$$

then (2.29) is true. The main difference is the dependence of α on the choice of the burn-in. One can essentially boost this confidence level by using a median of independent runs of the Markov chain Monte Carlo method. This is explained in [NP09].

However, both results presented are far away from well known Chernoff bounds. These exponential inequalities for finite Markov chains are shown in [Gil98] for random walks on graphs. In [Lez98], this Chernoff bound was extended and refined for Markov chains on finite and general state spaces, furthermore for discrete and continuous time. For irreducible and reversible Markov chains on finite D and $\|f\|_\infty \le 1$ one deduces from [Lez98, Theorem 1.1, p. 850] that

$$\Pr(|S_{n,n_0}(f) - S(f)| \ge \varepsilon) \le 3\left\|\frac{\nu P^{n_0}}{\pi}\right\|_2 \exp\left\{-n(1-\beta_1)\frac{\varepsilon^2}{12}\right\}. \tag{2.30}$$

In other words, if

$$n_0 \geq \frac{\log(\|\nu/\pi - 1\|_2)}{\log(\beta^{-1})} \quad \text{and} \quad n \geq \frac{12\varepsilon^{-2}\log(6\alpha^{-1})}{1-\beta_1}$$

then (2.29) holds true. This is better than using Lemma 2.27. In [LP04] Hoeffding bounds for reversible Markov chains are presented.

Such exponential inequalities also imply an error bound of the mean square error by the following well known formula (see for example [Kal02, Lemma 2.4, p. 26]).

LEMMA 2.28. *Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with transition matrix P and initial distribution ν. Then*

$$e_\nu(S_{n,n_0}, f)^2 = \int_0^\infty \Pr(|S_{n,n_0}(f) - S(f)| \geq \sqrt{\varepsilon})\,\mathrm{d}\varepsilon.$$

By Lemma 2.28 and by (2.30) one obtains the error bound

$$\sup_{\|f\|_\infty \leq 1} e_\nu(S_{n,n_0}, f)^2 \leq \frac{36(1+\beta^{n_0}\|\nu/\pi - 1\|_2)}{n(1-\beta_1)}.$$

The asymptotic coefficient as described in Corollaries 2.13 and 2.22 is not reached. However, the error bound applies also for periodic Markov chains.

Let us provide a conclusion. Different explicit error bounds for finite state spaces are known. The results presented in Section 2.2 are not entirely new. In the literature one can find similar estimates where some of the assumptions like aperiodicity or reversibility are weakened. The justification and discussion of the burn-in in Section 2.3 and the lower bound of Theorem 2.25 seem to be new. In the following we will extend the results to general state spaces.

3. General state spaces

In the following we study the mean square error of Markov chain Monte Carlo methods on general state spaces. The state space can be countable or uncountable. In Section 3.1 we provide the basic definitions and properties of Markov chains on general state spaces. The estimates of the mean square error are shown in Section 3.2. We suggest and justify a recipe for choosing the burn-in in Section 3.3. Afterwards the error bound is applied to illustrating examples and finally we discuss how the results fit into the published literature.

3.1. Markov chains. In this section facts and definitions regarding Markov chains on general state spaces are stated. The paper [RR04] of Rosenthal and Roberts surveys various results about Markov chains on general state spaces. For further reading we refer to [Rev84, Num84, MT09].

Let $(D, \mathfrak{D})$ be a measurable space. In most examples D is contained in $\mathbb{R}^d$ and $\mathfrak{D}$ is $\mathcal{B}(D)$, the Borel σ-algebra of D. We now define transition kernels and Markov chains.

Definition 3.1 (Markov kernel, transition kernel). The function $K\colon D \times \mathfrak{D} \to [0,1]$ is called a *Markov kernel* or a *transition kernel* if

(i) for each $x \in D$ the mapping $A \in \mathfrak{D} \mapsto K(x,A)$ is a probability measure on $(D,\mathfrak{D})$,
(ii) for each $A \in \mathfrak{D}$ the mapping $x \in D \mapsto K(x,A)$ is a $\mathfrak{D}$-measurable real-valued function.

Definition 3.2 (Markov chain). A sequence of random variables $(X_n)_{n\in\mathbb{N}}$ on a probability space $(\Omega,\mathcal{F},\Pr)$ mapping into $(D,\mathfrak{D})$ is called a *Markov chain with transition kernel* K if for all $n \in \mathbb{N}$ and $A \in \mathfrak{D}$ one has

$$\Pr(X_{n+1} \in A \mid X_1,\dots,X_n) = \Pr(X_{n+1} \in A \mid X_n) = K(X_n,A) \quad \text{almost surely.}$$

The distribution

$$\nu(A) = \Pr(X_1 \in A), \quad A \in \mathfrak{D},$$

is called the *initial distribution*.

Suppose that we have a transition kernel K and a probability measure ν. For simplicity assume that $D \subset \mathbb{R}^d$ and $\mathfrak{D} = \mathcal{B}(D)$. For any transition kernel there exists a *random mapping representation* (see for example Kallenberg [Kal02, Lemma 2.22, p. 34]), a measurable function $\Phi\colon D \times [0,1] \to D$ which satisfies

$$\Pr(\Phi(x,Z) \in A) = K(x,A), \quad x \in D,\ A \in \mathfrak{D},$$

where the random variable $Z\colon (\Omega,\mathcal{F},\Pr) \to ([0,1],\mathcal{B}([0,1]))$ is uniformly distributed. Then a Markov chain can be constructed as follows. Let $(Z_n)_{n\in\mathbb{N}}$, with $Z_n : (\Omega,\mathcal{F},\Pr) \to ([0,1],\mathcal{B}([0,1]))$, be a sequence of i.i.d. random variables with uniform distribution, and assume that X_1 has distribution ν. Then one can see that $(X_n)_{n\in\mathbb{N}}$ defined by

$$X_n = \Phi(X_{n-1},Z_n), \quad n \geq 2,$$

is a Markov chain with transition kernel K and initial distribution ν.

The transition kernel K of a Markov chain describes the probability of getting from state $x \in D$ to $A \in \mathfrak{D}$ in one step, i.e. for all $k \in \mathbb{N}$ one has

$$K(x,A) = \Pr(X_{k+1} \in A \mid X_k = x).$$

The n-step transition kernel is inductively given by

$$K^n(x,A) = \int_D K^{n-1}(y,A)\,K(x,\mathrm{d}y) = \int_D K(y,A)\,K^{n-1}(x,\mathrm{d}y).$$

The first equality above is a definition, and for a proof of the second see [Rev84, Proposition 1.6, p. 11] or [MT09, Theorem 3.4.2, p. 61]. The function K^n is again a transition kernel. The n-step transition probability from state $x \in D$ to $A \in \mathfrak{D}$ is

$$\Pr(X_{k+n} \in A \mid X_k = x) = K^n(x,A).$$

This is seen by integrating over the conditional distribution of the previous step:

$$\begin{aligned}
\Pr(X_{k+1} \in A \mid X_k = x) &= K(x, A),\\
\Pr(X_{k+2} \in A \mid X_k = x) &= \int_D \Pr(X_{k+2} \in A \mid X_{k+1} = y, X_k = x) \Pr(X_{k+1} \in \mathrm{d}y \mid X_k = x)\\
&= \int_D \Pr(X_{k+2} \in A \mid X_{k+1} = y)\, K(x, \mathrm{d}y) = K^2(x, A),\\
&\vdots\\
\Pr(X_{k+n} \in A \mid X_k = x) &= \int_D \Pr(X_{k+n} \in A \mid X_{k+n-1} = y, X_k = x)\\
&\qquad\qquad \times \Pr(X_{k+n-1} \in \mathrm{d}y \mid X_k = x)\\
&= \int_D \Pr(X_{k+n} \in A \mid X_{k+n-1} = y)\, K^{n-1}(x, \mathrm{d}y) = K^n(x, A).
\end{aligned}$$

In the following let us assume that we have a Markov chain $(X_n)_{n\in\mathbb{N}}$ with transition kernel K and initial distribution ν. The expectation $\mathbf{E}_{\nu,K}$ is taken with respect to the joint distribution of $(X_n)_{n\in\mathbb{N}}$, say $W_{\nu,K}$, which is defined on $(D^{\mathbb{N}}, \sigma(\mathcal{A}))$ where

$$\begin{aligned}
D^{\mathbb{N}} &= \{\omega = (x_1, x_2, \dots) \mid x_i \in D \text{ for all } i \geq 1\},\\
\mathcal{A} &= \bigcup_{k\in\mathbb{N}} \{A_1 \times \cdots \times A_k \times D \times \cdots \mid A_i \in \mathfrak{D},\, i = 1, \dots, k\}
\end{aligned}$$

(see [MT09, Theorem 3.4.1, p. 60] or [Rev84, Theorem 2.8, p. 17]). For $k \in \mathbb{N}$ one has, for $A_1 \times \cdots \times A_k \subset D^k$,

$$\begin{aligned}
&W_{\nu,K}(A_1 \times \cdots \times A_k \times D \times \cdots) = \Pr(X_1 \in A_1, \dots, X_k \in A_k)\\
&\quad = \int_{A_1} \int_{A_2} \dots \int_{A_{k-1}} K(x_{k-1}, A_k) K(x_{k-2}, \mathrm{d}x_{k-1}) \dots K(x_1, \mathrm{d}x_2)\, \nu(\mathrm{d}x_1). \qquad (3.1)
\end{aligned}$$

Now we present properties of transition kernels. These properties have finite state space counterparts (see Section 2.1).

Denote by $\mathcal{M}(D)$ the set of *real-valued signed measures* ([1]) on $(D, \mathfrak{D})$. For any $\nu \in \mathcal{M}(D)$ let us define

$$\nu P^m(A) = \int_D K^m(x, A)\, \nu(\mathrm{d}x), \quad A \in \mathfrak{D},\, m \in \mathbb{N}.$$

Note that the mapping $\nu \mapsto \nu P^m$ defines a linear operator on $\mathcal{M}(D)$. If ν is a probability measure then νP^m is the distribution of X_{m+1}, where $(X_n)_{n\in\mathbb{N}}$ is a Markov chain with transition kernel K and initial distribution ν.

Definition 3.3 (stationarity). Let π be a probability measure on $(D, \mathfrak{D})$. Then π is called a *stationary distribution* of a transition kernel K if

$$\pi P(A) = \pi(A), \quad A \in \mathfrak{D}.$$

([1]) The set function $\mu\colon \mathfrak{D} \to \mathbb{R}$ is a real-valued signed measure if $\mu(\emptyset) = 0$ and for pairwise disjoint $A_1, A_2, \dots$ with $A_k \in \mathfrak{D}$ for $k \in \mathbb{N}$, one has $\mu(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} \mu(A_k)$.

Roughly speaking, this means: if we choose the initial state with respect to a stationary distribution π, then, after a single transition the same distribution arises, i.e.

$$\Pr(X_1 \in A) = \pi(A) = \pi P(A) = \Pr(X_2 \in A), \quad A \in \mathfrak{D}.$$

Definition 3.4 (reversibility). Let π be a probability measure on $(D, \mathfrak{D})$. A transition kernel K is called *reversible with respect to* π if

$$\int_B K(x, A)\, \pi(\mathrm{d}x) = \int_A K(x, B)\, \pi(\mathrm{d}x), \quad A, B \in \mathfrak{D}.$$

If a transition kernel K is reversible with respect to a distribution π, then π is a stationary distribution of K. If the initial distribution of a Markov chain with transition kernel K is π, then reversibility with respect to π is equivalent to

$$\Pr(X_1 \in A,\, X_2 \in B) = \Pr(X_1 \in B,\, X_2 \in A), \quad A, B \in \mathfrak{D}.$$

A Markov chain is called reversible with respect to π if the corresponding transition kernel is reversible with respect to π.

Definition 3.5 (lazy version). Let K be a transition kernel and let $\mathbf{1}_A(x)$ be the indicator function of $A \in \mathfrak{D}$ for $x \in D$. Then we call

$$\widetilde{K}(x, A) = \tfrac{1}{2}(\mathbf{1}_A(x) + K(x, A)), \quad x \in D,\ A \in \mathfrak{D},$$

the *lazy version* of K.

If π is a stationary distribution of K, then π is also a stationary distribution of $\widetilde{K}$. If K is reversible with respect to π, then $\widetilde{K}$ is also reversible with respect to π. For a Markov chain with transition kernel K and initial distribution ν we may define a *lazy Markov chain*, a Markov chain with transition kernel $\widetilde{K}$ and initial distribution ν.

Assume that π is a stationary distribution of a transition kernel K and let $f\colon D \to \mathbb{R}$ be an integrable function with respect to π. Let us define

$$P^m f(x) = \int_D f(y)\, K^m(x, \mathrm{d}y), \quad x \in D,\, m \in \mathbb{N}.$$

We call P the *Markov operator* or the *transition operator*. If a Markov chain $(X_n)_{n\in\mathbb{N}}$ with transition kernel K and initial distribution δ_x, the point mass at $x \in D$, is given, then $P^m f(x)$ is the expectation of $f(X_{m+1})$.

Let us state some well known properties of the operator P acting on functions and on signed measures.

Lemma 3.6. *Let π be a stationary distribution of a transition kernel K and let $f\colon D \to \mathbb{R}$ be an integrable function with respect to π. Then for $\nu \in \mathcal{M}(D)$,*

$$\int_D f(x)\, (\nu P^m)(\mathrm{d}x) = \int_D (P^m f)(x)\, \nu(\mathrm{d}x), \quad m \in \mathbb{N}, \tag{3.2}$$

whenever one of the integrals exists. In particular

$$S(f) = \int_D f(x)\, \pi(\mathrm{d}x) = \int_D (P^m f)(x)\, \pi(\mathrm{d}x), \quad m \in \mathbb{N}. \tag{3.3}$$

Proof. Equation (3.3) is an immediate consequence of (3.2) and stationarity. Hence one has to prove (3.2). The equality holds for indicator functions and for simple functions.

Then by the standard procedure of integration theory the equality can be extended to positive and afterwards to integrable functions. ∎

Note that if a Markov chain $(X_n)_{n\in\mathbb{N}}$ with transition kernel K and initial distribution ν is given, then (3.2) can be rewritten as

$$\mathbf{E}_{\nu,K}[f(X_{m+1})] = \mathbf{E}_{\nu,K}[\mathbf{E}_{\nu,K}[f(X_{m+1}) \mid X_1]].$$

The following result is well known (see for example [LS93, equation (1.2), p. 365]).

LEMMA 3.7. *Let the transition kernel K be reversible with respect to π and let $F\colon D \times D \to \mathbb{R}$. Then*

$$\int_D \int_D F(x,y)\, K^m(x,\mathrm{d}y)\, \pi(\mathrm{d}x) = \int_D \int_D F(y,x)\, K^m(x,\mathrm{d}y)\, \pi(\mathrm{d}x), \quad m \in \mathbb{N}, \tag{3.4}$$

whenever one of the integrals exists.

Proof. The reversibility of the transition kernel K implies reversibility of the m-step transition kernel K^m. Hence it is sufficient to show the assertion for $m = 1$. By using the reversibility one has

$$\int_D \int_D \mathbf{1}_{A\times B}(x,y)\, K(x,\mathrm{d}y)\, \pi(\mathrm{d}x) = \int_D \int_D \mathbf{1}_{A\times B}(y,x)\, K(x,\mathrm{d}y)\, \pi(\mathrm{d}x), \quad A, B \in \mathfrak{D}.$$

The equality of the integrals can be extended to arbitrary sets $C \in \mathfrak{D} \otimes \mathfrak{D}$, where $\mathfrak{D} \otimes \mathfrak{D}$ is the product σ-algebra. This is an application of Dynkin's Theorem. Then it is straightforward to consider the cases where F is a simple function, a positive function and finally an integrable one. ∎

For $p \in [1,\infty)$ let us define

$$L_p = L_p(D,\pi) = \left\{ f\colon D \to \mathbb{R} \;\middle|\; \|f\|_p^p = \int_D |f(x)|^p\, \pi(\mathrm{d}x) < \infty \right\}.$$

For $p = \infty$ the essential-supremum norm with respect to π is defined by

$$\|f\|_\infty = \operatorname*{ess\,sup}_{y\in D} |f(y)| = \inf_{N\in\mathfrak{D},\, \pi(N)=0} \sup_{y\in D\setminus N} |f(y)|,$$

and we set

$$L_\infty = L_\infty(D,\pi) = \{ f\colon D \to \mathbb{R} \mid \|f\|_\infty < \infty \}.$$

Sometimes it is convenient to consider bounded functions on D, not π-a.e. bounded ones, thus we define

$$L_B = L_B(D) = \left\{ f\colon D \to \mathbb{R} \;\middle|\; |f| = \sup_{x\in D} |f(x)| < \infty \right\}.$$

The next result is standard (see for example [BR95, Lemma 1, p. 334]).

LEMMA 3.8. *Let $p \in [1,\infty]$. For any transition kernel K with a stationary distribution π,*

$$\|Pf\|_p \le \|f\|_p \quad \textit{and} \quad \|P\|_{L_p \to L_p} = 1.$$

Proof. If $p < \infty$, then by the Jensen inequality (J) and (3.3) one obtains

$$\int_D |Pf(x)|^p \pi(\mathrm{d}x) \le \int_D \left(\int_D |f(y)| K(x, \mathrm{d}y) \right)^p \pi(\mathrm{d}x)$$
$$\underset{\text{(J)}}{\le} \int_D \int_D |f(y)|^p K(x, \mathrm{d}y)\, \pi(\mathrm{d}x) \underset{(3.3)}{=} \int_D |f(x)|^p \, \pi(\mathrm{d}x).$$

Since π is a stationary distribution of the transition kernel one has, for $N \in \mathfrak{D}$,

$$\pi(N) = 0 \;\Leftrightarrow\; K(\cdot, N) = 0 \ \pi\text{-a.e.}$$

Null sets with respect to π are the same as null sets with respect to $K(x, \cdot)$ for almost all $x \in D$. Hence

$$|Pf(x)| \le \int_D |f(y)| K(x, \mathrm{d}y) \le \|f\|_\infty \quad \pi\text{-a.e.}$$

and we have $\|Pf\|_p \le \|f\|_p$ for $p \in [1, \infty]$. Let $u(x) = 1$ for all $x \in D$. Then $Pu = u$ with $\|u\|_p = 1$ and we obtain $\|P\|_{L_p \to L_p} = 1$. ∎

The closed subspace

$$L_p^0 = \{f \in L_p \mid S(f) = 0\}$$

of L_p is important. Note that L_2 and L_2^0 are Hilbert spaces with inner product

$$\langle f, g \rangle = \int_D f(x) g(x) \, \pi(\mathrm{d}x).$$

Then

$$L_2 = L_2^0 \oplus (L_2^0)^\perp, \quad \text{where} \quad (L_2^0)^\perp = \{f \in L_2 \mid f \equiv c,\ c \in \mathbb{R}\}.$$

On the Hilbert spaces L_2 and L_2^0 there exists the *adjoint operator* P^* such that

$$\langle Pf, g \rangle = \langle f, P^* g \rangle.$$

Furthermore

$$\|P\|_{L_2^0 \to L_2^0} = \|P^*\|_{L_2^0 \to L_2^0} \quad \text{and} \quad \|P - S\|_{L_2 \to L_2} = \|P^* - S\|_{L_2 \to L_2}.$$

The following facts about adjoint operators are helpful. Let $T : L_p \to L_p$, with $p \in [1, \infty)$, be a bounded linear operator. Then the adjoint operator $T^* : L_q \to L_q$, with $q \in (1, \infty]$, is defined as follows. Suppose that p and q are such that $p^{-1} + q^{-1} = 1$. It is well known that L_q is isometrically isomorphic to the dual space $(L_p)'$, where the isomorphism is given by

$$A \colon L_q \to (L_p)', \quad A(g)(f) = \langle f, g \rangle, \quad f \in L_p.$$

Then there exists the dual operator $T^\times \colon (L_p)' \to (L_p)'$ and the adjoint operator acting on L_q can be defined as $T^* = A^{-1} T^\times A$. This is illustrated by the diagram below.

$$\begin{array}{ccc} L_q & \xrightarrow{T^*} & L_q \\ \downarrow{\scriptstyle A} & & \uparrow{\scriptstyle A^{-1}} \\ (L_p)' & \xrightarrow{T^\times} & (L_p)' \end{array}$$

Furthermore, for all $f \in L_p$ and for all $g \in L_q$ one has

$$\begin{aligned}\langle f, T^*g\rangle &= \langle f, A^{-1}T^\times Ag\rangle = A(A^{-1}T^\times Ag)(f)\\ &= (T^\times A)(g)(f) \underset{\text{(dual operator)}}{=} A(g)(Tf) = \langle Tf, g\rangle.\end{aligned}$$

Then

$$\begin{aligned}\|T\|_{L_p\to L_p} &= \|T^\times\|_{(L_p)'\to(L_p)'} = \sup_{\|Ag\|_{(L_p)'}\le 1} \|T^\times Ag\|_{(L_p)'}\\ &= \sup_{\|g\|_q\le 1} \|A^{-1}T^\times Ag\|_q = \|T^*\|_{L_q\to L_q}.\end{aligned}$$

If $T = P - S$, then it follows that

$$\|P - S\|_{L_p\to L_p} = \|P^* - S\|_{L_q\to L_q}.$$

Let $\nu \in \mathcal{M}(D)$. If there exists a density of ν with respect to π then we denote it by $\frac{d\nu}{d\pi}$ and for $q \in [1,\infty]$ let

$$\|\nu\|_q = \begin{cases} \left\|\dfrac{d\nu}{d\pi}\right\|_q, & \nu \ll \pi,\\ \infty, & \text{otherwise.}\end{cases}$$

Set

$$\mathcal{M}_q = \mathcal{M}_q(D,\pi) = \{\nu \in \mathcal{M}(D) \mid \|\nu\|_q < \infty\}.$$

The function space L_q is isometrically isomorphic to the space $\mathcal{M}_q$ of signed measures, in symbols $L_q \cong \mathcal{M}_q$. The space $\mathcal{M}_2$ of signed measures is a Hilbert space and the inner product is the L_2 inner product of the densities:

$$\langle \nu, \mu\rangle = \int_D \frac{d\nu}{d\pi}(x)\,\frac{d\mu}{d\pi}(x)\,\pi(\mathrm{d}x) = \left\langle \frac{d\nu}{d\pi}, \frac{d\mu}{d\pi}\right\rangle, \quad \nu,\mu \in \mathcal{M}_2.$$

Furthermore set

$$\mathcal{M}_q^0 = \{\nu \in \mathcal{M}_q \mid \nu(D) = 0\}.$$

Then

$$\mathcal{M}_2 = \mathcal{M}_2^0 \oplus (\mathcal{M}_2^0)^\perp, \quad \text{where} \quad (\mathcal{M}_2^0)^\perp = \{\nu \in \mathcal{M}_2 \mid \nu = c\pi,\, c \in \mathbb{R}\}.$$

Clearly, $\mathcal{M}_2^0$ is also a Hilbert space. We have $L_2^0 \cong \mathcal{M}_2^0$ and $(L_2^0)^\perp \cong (\mathcal{M}_2^0)^\perp$. Let us recall that the transition kernel applies to signed measures $\nu \in \mathcal{M}_q$ as

$$\nu P(A) = \int_D K(x,A)\,\nu(\mathrm{d}x), \quad A \in \mathfrak{D}.$$

LEMMA 3.9. *Let K be a transition kernel and let π be a stationary distribution of K.*

(i) *Let $q \in (1,\infty]$ and $\nu \in \mathcal{M}_q$. Then*

$$\frac{d(\nu P)}{d\pi}(x) = P^*\left(\frac{d\nu}{d\pi}\right)(x) \quad \pi\text{-a.e.}$$

and

$$\|P\|_{L_2^0\to L_2^0} = \|P\|_{\mathcal{M}_2^0\to\mathcal{M}_2^0}.$$

(ii) *Reversibility with respect to π is equivalent to P being self-adjoint acting on L_2 and $\mathcal{M}_2$, i.e.*

$$\langle Pf, g\rangle = \langle f, Pg\rangle \quad \textit{and} \quad \langle \nu P, \mu\rangle = \langle \nu, \mu P\rangle.$$

Proof. (i) For all $f \in L_p$ with p chosen such that $p^{-1} + q^{-1} = 1$ one has

$$\left\langle f, \frac{d(\nu P)}{d\pi}\right\rangle = \int_D f(x)\,(\nu P)(\mathrm{d}x) \underset{(3.2)}{=} \int_D (Pf)(x)\,\nu(\mathrm{d}x) = \left\langle Pf, \frac{d\nu}{d\pi}\right\rangle = \left\langle f, P^*\left(\frac{d\nu}{d\pi}\right)\right\rangle.$$

Hence π-a.e.,

$$\frac{d(\nu P)}{d\pi}(x) = P^*\left(\frac{d\nu}{d\pi}\right)(x).$$

By using the previous equation one obtains

$$\begin{aligned}\|P\|_{\mathcal{M}_2^0\to\mathcal{M}_2^0} &= \sup_{\|\frac{d\mu}{d\pi}\|_2=1,\,\mu(D)=0} \left\|\frac{d(\mu P)}{d\pi}\right\|_2 = \sup_{\|\frac{d\mu}{d\pi}\|_2=1,\,S(\frac{d\mu}{d\pi})=0} \left\|P^*\left(\frac{d\mu}{d\pi}\right)\right\|_2 \\ &= \|P^*\|_{L_2^0\to L_2^0} = \|P\|_{L_2^0\to L_2^0}.\end{aligned}$$

(ii) It is clear that self-adjointness implies reversibility. The other direction follows by

$$\langle Pf, g\rangle = \int_D\int_D f(y)g(x)K(x,\mathrm{d}y)\,\pi(\mathrm{d}x) \underset{(3.4)}{=} \int_D\int_D f(x)g(y)\,K(x,\mathrm{d}y)\,\pi(\mathrm{d}x) = \langle f, Pg\rangle.$$

The result with respect to $\mathcal{M}_2$ is shown by using (i) and the self-adjointness of P on L_2. ∎

In the following we introduce several convergence properties of a Markov chain $(X_n)_{n\in\mathbb{N}}$ with transition kernel K and initial distribution ν. We assume that π is a stationary distribution of K. The goal is to quantify the speed of convergence of νP^m to π for increasing $m \in \mathbb{N}$. For further details let us refer to [RR97a], [RR04] or [Che05].

Definition 3.10 (L_2-spectral gap). Let P be the Markov operator with corresponding transition kernel K. Then there exists an (absolute) *L_2-spectral gap* if

$$\beta = \|P\|_{L_2^0\to L_2^0} < 1,$$

and the value of the L_2-spectral gap is $1-\beta$.

Let us briefly explain what this means for a reversible transition kernel. If the transition kernel K is reversible with respect to π, then let $\mathbf{spec}(P|L_2)$ be the spectrum of the self-adjoint operator P acting on L_2 and $\mathbf{spec}(P|L_2^0)$ be the spectrum of P acting on L_2^0. Since $\|P\|_{L_2\to L_2} \le 1$ the spectrum $\mathbf{spec}(P|L_2)$ is contained in $[-1,1]$. Let us define

$$\lambda = \inf\{\alpha \mid \alpha \in \mathbf{spec}(P|L_2^0)\} \quad \text{and} \quad \Lambda = \sup\{\alpha \mid \alpha \in \mathbf{spec}(P|L_2^0)\}.$$

Since P is self-adjoint, it is well known that

$$\lambda = \inf_{\|g\|_2=1,\, g\in L_2^0} \langle Pg, g\rangle \quad \text{and} \quad \Lambda = \sup_{\|g\|_2=1,\, g\in L_2^0} \langle Pg, g\rangle.$$

Then we have

$$\mathbf{spec}(P|L_2^0) \subset [\lambda, \Lambda] \quad \text{and} \quad \beta = \|P\|_{L_2^0\to L_2^0} = \max\{\Lambda, |\lambda|\}.$$

The existence of an L_2-spectral gap implies that $-1 < \lambda \le \Lambda < 1$, so there is a gap between $1 \in \mathbf{spec}\,(P|L_2)$ and β, the second largest absolute value of $\mathbf{spec}(P|L_2)$.

DEFINITION 3.11 (L_2-geometric ergodicity). A transition kernel K with stationary distribution π is called *L_2-geometrically ergodic* if for all probability measures $\nu \in \mathcal{M}_2$ there exist $\alpha \in [0,1)$ and $C_\nu < \infty$ such that

$$\|\nu P^n - \pi\|_2 \leq C_\nu \, \alpha^n, \quad n \in \mathbb{N}.$$

An L_2-spectral gap implies L_2-geometric ergodicity.

PROPOSITION 3.12. *Let K be a transition kernel with stationary distribution π. Assume that the Markov operator P has an L_2-spectral gap, i.e. $1-\beta > 0$. Then the transition kernel K is L_2-geometrically ergodic.*

Proof. If $\nu \in \mathcal{M}_2$ and $\nu(D) = 1$, then one obtains $(\nu - \pi)(D) = 0$ and the proof is completed by

$$\|\nu P^n - \pi\|_2 = \|(\nu - \pi)P^n\|_2 \leq \|P\|^n_{\mathcal{M}_2^0 \to \mathcal{M}_2^0} \|\nu - \pi\|_2 = \beta^n \|\nu - \pi\|_2. \quad \blacksquare$$

If the transition kernel is reversible with respect to π, then L_2-geometric ergodicity and the existence of an L_2-spectral gap are equivalent:

PROPOSITION 3.13 ([RR97a, Theorem 2.1, p. 17]). *Let the transition kernel K be reversible with respect to π. Then the following statements are equivalent:*

(i) *The transition kernel is L_2-geometrically ergodic.*
(ii) *The Markov operator P has an L_2-spectral gap.*

For further details and more equivalents of L_2-geometric ergodicity, see [RR97a, RT01]. The next definition is similar to L_p-exponential convergence in [Che05].

DEFINITION 3.14 (L_p-exponential convergence). Let $p \in [1,\infty]$, let $\alpha \in [0,1)$ and $M < \infty$. Then the transition kernel K with stationary distribution π is called *L_p-exponentially convergent with (α, M)* if

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

The transition kernel is called *L_p-exponentially convergent* if there exist an $M < \infty$ and an $\alpha \in [0,1)$ such that K is L_p-exponentially convergent with (α, M).

The Markov chain is called L_2-geometrically ergodic or L_p-exponentially convergent if the corresponding transition kernel K is L_2-geometrically ergodic or L_p-exponentially convergent.

Let p and q be such that $p^{-1} + q^{-1} = 1$. Then L_p-exponential convergence implies convergence of νP^n to the stationary distribution π in $\mathcal{M}_q$ for increasing $n \in \mathbb{N}$.

COROLLARY 3.15. *Let $p \in [1,\infty)$ and $\nu \in \mathcal{M}_q$ with $p^{-1} + q^{-1} = 1$. Let the transition kernel K with stationary distribution π be L_p-exponentially convergent with (α, M). Then*

$$\|\nu P^n - \pi\|_q \leq M\|\nu - \pi\|_q \alpha^n, \quad n \in \mathbb{N}.$$

Proof. Indeed,

$$\begin{aligned}\|\nu P^n - \pi\|_q &= \|(\nu-\pi)P^n\|_q = \left\|\frac{d((\nu-\pi)P^n)}{d\pi}\right\|_q = \left\|(P^n)^*\left(\frac{d\nu}{d\pi}-1\right)\right\|_q \\ &= \left\|((P^n)^*-S)\left(\frac{d\nu}{d\pi}-1\right)\right\|_q \le \|(P^n-S)^*\|_{L_q\to L_q}\left\|\frac{d\nu}{d\pi}-1\right\|_q \\ &\le \|P^n-S\|_{L_p\to L_p}\left\|\frac{d\nu}{d\pi}-1\right\|_q \le M\|\nu-\pi\|_q\alpha^n. \ \blacksquare\end{aligned}$$

In the following we consider relations between the existence of an L_2-spectral gap and L_p-exponential convergence. First, let us add some helpful inequalities.

LEMMA 3.16. *Let π be a stationary distribution of a transition kernel K. Then*

$$\|P^n\|_{L_2^0\to L_2^0} = \|P^n - S\|_{L_2\to L_2} \le \beta^n, \quad n\in\mathbb{N}. \tag{3.5}$$

If $p\in[1,\infty]$ then

$$\|P^n\|_{L_p^0\to L_p^0} \le \|P^n - S\|_{L_p\to L_p} \le 2\|P^n\|_{L_p^0\to L_p^0}, \quad n\in\mathbb{N}. \tag{3.6}$$

Proof. Note that if P is a *normal operator*, i.e. $PP^* = P^*P$, then $\|P^n\|_{L_2^0\to L_2^0} = \beta^n$, otherwise one has $\|P^n\|_{L_2^0\to L_2^0} \le \|P\|^n_{L_2^0\to L_2^0} = \beta^n$. From

$$\begin{aligned}\|P^n - S\|_{L_2\to L_2} &= \sup_{\|f\|_2\le 1}\|(P^n-S)f\|_2 = \sup_{\|f\|_2\le 1}\|P^n(f-S(f))\|_2 \\ &\le \sup_{\|g\|_2\le 1,\, S(g)=0}\|P^n g\|_2 = \|P^n\|_{L_2^0\to L_2^0}\end{aligned}$$

and

$$\begin{aligned}\|P^n\|_{L_p^0\to L_p^0} &= \sup_{\|g\|_p\le 1,\, S(g)=0}\|P^n g\|_p = \sup_{\|g\|_p\le 1,\, S(g)=0}\|P^n g - S(g)\|_p \\ &\le \sup_{\|f\|_p\le 1}\|(P^n-S)f\|_p = \|P^n - S\|_{L_p\to L_p}\end{aligned}$$

claim (3.5) and the first part of (3.6) follow. Furthermore one obtains

$$\begin{aligned}\|P^n - S\|_{L_p\to L_p} &= \sup_{\|f\|_p\le 1}\|P^n f - Sf\|_p = 2\sup_{\|f\|_p\le 1}\left\|P^n\left(\frac{1}{2}(f-Sf)\right)\right\|_p \\ &\le 2\sup_{\|g\|_p\le 1,\, S(g)=0}\|P^n g\|_p = 2\|P^n\|_{L_p^0\to L_p^0},\end{aligned}$$

which finishes the proof. ∎

In a general setting it follows that an L_2-spectral gap implies L_p-exponential convergence for all $p\in(1,\infty)$.

PROPOSITION 3.17. *Let $p\in(1,\infty)$. Let π be a stationary distribution of a transition kernel K and $n\in\mathbb{N}$. The existence of an L_2-spectral gap, $1-\beta>0$, implies L_p-exponential convergence. We have*

$$\|P^n - S\|_{L_p\to L_p} \le \begin{cases} 2^{2/p}\,\beta^{2n(p-1)/p}, & p\in(1,2), \\ 2^{2(p-1)/p}\,\beta^{2n/p}, & p\in[2,\infty). \end{cases} \tag{3.7}$$

Proof. Let $p \in (1,2)$. Lemma 3.16 gives

$$\|P^n - S\|_{L_2 \to L_2} \le \beta^n \quad \text{and} \quad \|P^n - S\|_{L_1 \to L_1} \le 2.$$

We apply Proposition A.4 (Riesz–Thorin Interpolation Theorem), where $T = P^n - S$ and $q_1 = 2$, $q_2 = 1$ such that $\theta = (2-p)/p$. The case where $p \in (2,\infty)$ follows by the same interpolation argument, since by Lemma 3.16 one has

$$\|P^n - S\|_{L_2 \to L_2} \le \beta^n \quad \text{and} \quad \|P^n - S\|_{L_\infty \to L_\infty} \le 2. \ \blacksquare$$

From Proposition 3.17 and actually already from (3.5) it follows that an L_2-spectral gap implies L_2-exponential convergence. With the additional assumption of normality of P one can prove the converse.

PROPOSITION 3.18. *Let π be a stationary distribution of the transition kernel K. Let the Markov operator P be normal, i.e. $PP^* = P^*P$. Then the following statements are equivalent:*

(i) *There exists an L_2-spectral gap, i.e. $1 - \beta > 0$.*
(ii) *There exist $\alpha \in [0,1)$ and $M < \infty$ such that the transition kernel K is L_2-exponentially convergent with (α, M).*

In particular (ii) *implies*

$$\beta = \|P - S\|_{L_2 \to L_2} \le \alpha,$$

so that

$$\beta = \min\{\alpha \mid \exists M < \infty \text{ with } \|P^n - S\|_{L_2 \to L_2} \le M\alpha^n,\ n \in \mathbb{N}\}.$$

Proof. By (3.5), (i) implies (ii) with $(\alpha, M) = (\beta, 1)$. Now we show that (ii) implies (i). One has

$$\|P\|^2_{L_2^0 \to L_2^0} = \|PP^*\|_{L_2^0 \to L_2^0},$$

where PP^* is self-adjoint and $(P^*)^n = (P^n)^*$ for all $n \in \mathbb{N}$. Then

$$\begin{aligned}\|P^n - S\|^2_{L_2 \to L_2} &= \|P^n\|^2_{L_2^0 \to L_2^0} = \|P^n (P^n)^*\|_{L_2^0 \to L_2^0} \\ &= \|P^n (P^*)^n\|_{L_2^0 \to L_2^0} \underset{\text{(normality)}}{=} \|(PP^*)^n\|_{L_2^0 \to L_2^0},\end{aligned}$$

so that

$$\|P^n - S\|_{L_2 \to L_2} \le M\alpha^n \ \Leftrightarrow\ \|(PP^*)^n\|_{L_2^0 \to L_2^0} \le M^2 \alpha^{2n}. \tag{3.8}$$

By the spectral radius formula and the self-adjointness (s-a) of PP^* one obtains

$$\begin{aligned}\|P\|^2_{L_2^0 \to L_2^0} &= \|PP^*\|_{L_2^0 \to L_2^0} \underset{\text{(s-a)}}{=} r[PP^*] \\ &= \lim_{n\to\infty} (\|(PP^*)^n\|_{L_2^0 \to L_2^0})^{1/n} \underset{(3.8)}{\le} \alpha^2 \lim_{n\to\infty} (M^2)^{1/n} \le \alpha^2.\end{aligned}$$

Hence the proof is complete. ∎

By an interpolation argument we deduce that L_∞-exponential convergence or L_1-exponential convergence imply an L_2-spectral gap if the Markov operator is normal.

PROPOSITION 3.19. *Let π be a stationary distribution of the transition kernel K. Let K be L_1-exponentially convergent or L_∞-exponentially convergent with (α, M). Suppose that*

the Markov operator P is normal, i.e. $PP^* = P^*P$. Then there exists an L_2-spectral gap, in particular

$$\beta = \|P - S\|_{L_2 \to L_2} \leq \sqrt{\alpha}. \tag{3.9}$$

Proof. We show that L_1-exponential convergence with (α, M) implies $\beta \leq \sqrt{\alpha}$. For L_∞-exponentially convergent Markov chains the claim follows by the same arguments, where the roles of L_∞ and L_1 are interchanged.

By the assumptions of the proposition and Lemma 3.16 one has

$$\|P^n - S\|_{L_1 \to L_1} \leq \alpha^n M \quad \text{and} \quad \|P^n - S\|_{L_\infty \to L_\infty} \leq 2.$$

By Proposition A.4 (Riesz–Thorin Interpolation Theorem), where $T = P^n - S$ and $q_1 = 1$, $q_2 = \infty$, $\theta = 1/2$ one obtains L_2-exponential convergence with $(\sqrt{\alpha}, 2^{3/2} M^{1/2})$. Then Proposition 3.18 implies $\beta \leq \sqrt{\alpha}$, completing the proof. ∎

Another way to measure the convergence of νP^n to π for increasing $n \in \mathbb{N}$ is by using the total variation distance, defined as follows.

Definition 3.20 (total variation distance). The *total variation distance* between two probability measures $\nu, \mu \in \mathcal{M}(D)$ is defined by

$$\|\nu - \mu\|_{\mathrm{tv}} = \sup_{A \in \mathfrak{D}} |\nu(A) - \mu(A)|.$$

The total variation distance can be considered as an L_1-norm.

Lemma 3.21 ([RR04, Proposition 3, p. 28]). *Let $\nu, \mu \in \mathcal{M}(D)$ be probability measures. Then*

$$\|\nu - \mu\|_{\mathrm{tv}} = \frac{1}{2} \sup_{|f| \leq 1} \left| \int_D f(x) \, (\nu(\mathrm{d}x) - \mu(\mathrm{d}x)) \right|, \tag{3.10}$$

where $|f| = \sup_{x \in D} |f(x)|$. If $\nu, \mu \in \mathcal{M}_1$, then $\|\nu - \mu\|_{\mathrm{tv}} = \frac{1}{2} \|\nu - \mu\|_1$.

Now we can define uniform ergodicity of a transition kernel K.

Definition 3.22 (uniform ergodicity, π-a.e. uniform ergodicity). Let $M < \infty$ and $\alpha \in [0, 1)$. Then the transition kernel K with stationary distribution π is called *uniformly ergodic with* (α, M) if for all $x \in D$,

$$\|K^n(x, \cdot) - \pi\|_{\mathrm{tv}} \leq M \alpha^n, \quad n \in \mathbb{N}. \tag{3.11}$$

If the inequality of (3.11) holds π-a.e., rather than for all $x \in D$, then the transition kernel K is called *π-a.e. uniformly ergodic with* (α, M). A Markov chain with transition kernel K is called *uniformly ergodic* or *π-a.e. uniformly ergodic* if there exist $M < \infty$ and $\alpha \in [0, 1)$ such that K is uniformly ergodic or π-a.e uniformly ergodic with (α, M).

Obviously, if the transition kernel is uniformly ergodic then it is also π-a.e. uniformly ergodic. Note that in other references, e.g. [Che05], uniform ergodicity is called strong ergodicity.

Uniform ergodicity is closely related to L_∞-exponential convergence. An important relation is presented in the following proposition. Recall that $L_B = L_B(D)$ denotes the class of bounded functions on D.

PROPOSITION 3.23. *Let $\alpha \in [0,1)$ and $M < \infty$. Let π be a stationary distribution of the transition kernel K. Then the following statements are equivalent:*

(i′) *The transition kernel K is uniformly ergodic with (α, M).*
(ii′) *The transition operator P satisfies*

$$\|P^n - S\|_{L_B \to L_B} \leq 2M\,\alpha^n, \quad n \in \mathbb{N}.$$

Furthermore (i′) *and* (ii′) *imply the following equivalent statements:*

(i) *K is π-a.e. uniformly ergodic with (α, M).*
(ii) *K is L_∞-exponentially convergent with $(\alpha, 2M)$.*

Proof. Lemma 3.21 yields the equivalence of (i′) and (ii′). To prove the equivalence of (i) and (ii), let us first show that π-a.e.

$$\sup_{\|f\|_\infty \leq 1} |P^n f(x) - S(f)| = \sup_{|f| \leq 1} |P^n f(x) - S(f)|.$$

Note that

$$\pi(N) = 0 \;\Leftrightarrow\; K^n(\cdot, N) = 0 \quad \pi\text{-a.e.}$$

for all $N \in \mathfrak{D}$ and $n \in \mathbb{N}$, since π is the stationary distribution. Suppose that $f \in L_\infty$. Obviously, if $N \in \mathfrak{D}$ and $\pi(N) = 0$ then π-a.e.

$$|P^n f(x) - S(f)| = |P^n(\mathbf{1}_{N^c} f)(x) - S(\mathbf{1}_{N^c} f)|.$$

Let $\|f\|_\infty \leq 1$, i.e. $\pi(\{x \in D : f(x) > 1\}) = 0$. Define

$$g(x) = \begin{cases} f(x), & f(x) \leq 1, \\ 1, & f(x) > 1, \end{cases}$$

so that $f(x) = g(x)$ π-a.e. and $|g| \leq 1$. Thus, π-a.e.

$$|P^n f(x) - S(f)| = |P^n g(x) - S(g)| \leq \sup_{|g| \leq 1} |P^n g(x) - S(g)|,$$

so that π-a.e.

$$\sup_{\|f\|_\infty \leq 1} |P^n f(x) - S(f)| \leq \sup_{|g| \leq 1} |P^n g(x) - S(g)|.$$

The inequality in the other direction is clearly also correct, i.e. π-a.e.

$$\sup_{\|f\|_\infty \leq 1} |P^n f(x) - S(f)| = \sup_{|g| \leq 1} |P^n g(x) - S(g)|.$$

By applying the essential supremum on both sides of the previous equation and (3.10) one obtains

$$\|P^n - S\|_{L_\infty \to L_\infty} = 2 \operatorname*{ess\,sup}_{x \in D} \|K^n(x, \cdot) - \pi\|_{\mathrm{tv}}.$$

This completes the proof. ∎

It is known that there are transition kernels where the Markov operators have an L_2-spectral gap and the transition kernels are not uniformly ergodic (see [MT96]). Furthermore, uniform ergodicity implies an L_2-spectral gap (see [RR97a]). In this sense uniform ergodicity is a stronger property than the existence of an L_2-spectral gap.

PROPOSITION 3.24. *Let $\alpha \in [0,1)$ and $M < \infty$. Let the transition kernel K be reversible with respect to π. Then the following statements are equivalent:*

(i) *K is L_1-exponentially convergent with $(\alpha, 2M)$.*
(ii) *K is L_∞-exponentially convergent with $(\alpha, 2M)$.*
(iii) *K is π-a.e. uniformly ergodic with (α, M).*

Each of these conditions implies that the Markov operator has an L_2-spectral gap. We have

$$\beta = \|P\|_{L_2^0 \to L_2^0} \le \alpha.$$

Proof. First we prove the equivalence of (i) and (ii). By reversibility one can see for $f \in L_1$ and $h \in L_\infty$ that

$$\langle (P^n - S)f, h\rangle \underset{(3.4)}{=} \langle f, (P^n - S)h\rangle.$$

The adjoint operator of $P^n - S$ acting on L_1 is $P^n - S$ acting on L_∞. Since

$$\|P^n - S\|_{L_1 \to L_1} = \|P^n - S\|_{L_\infty \to L_\infty},$$

the equivalence is obvious.

By Proposition 3.23, (ii) is equivalent to (iii).

The last implication follows by an interpolation argument. Proposition A.4 (Riesz–Thorin Theorem) with $q_1 = \infty$, $q_2 = 1$ and $\theta = 1/2$ is applied. Then

$$\|P^n\|_{L_2^0 \to L_2^0} \underset{(3.5)}{=} \|P^n - S\|_{L_2 \to L_2} \le 4M\alpha^n, \quad n \in \mathbb{N}. \tag{3.12}$$

Because of the self-adjointness (s-a) of P one can apply the spectral radius formula to obtain

$$\beta = \|P\|_{L_2^0 \to L_2^0} \underset{(\text{s-a})}{=} r[P] = \lim_{n\to\infty} \|P^n\|^{1/n}_{L_2^0 \to L_2^0} \underset{(3.12)}{\le} \alpha \lim_{n\to\infty} (4M)^{1/n} = \alpha. \ \blacksquare$$

In Figure 1 we present a survey of the discussed relations between various convergence and ergodicity notions.

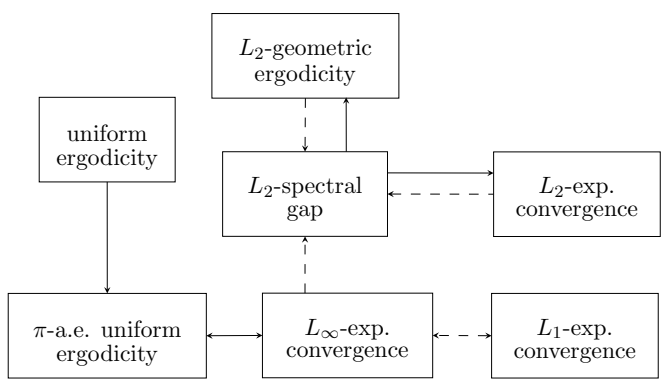

Fig. 1. Ergodicity notions and their relations. Solid lines represent implications without any assumption of reversibility. Dashed lines represent implications under the assumption of reversibility.

3.2. Error bounds. In this section we prove error bounds on general state spaces. We assume that we have a Markov chain $(X_n)_{n\in\mathbb{N}}$ with transition kernel K and initial distribution ν, where π is a stationary distribution, and compute

$$S_{n,n_0}(f) = \frac{1}{n}\sum_{j=1}^{n} f(X_{j+n_0})$$

as approximation to $S(f) = \int_D f(x)\,\pi(\mathrm{d}x)$. The error is measured in the mean square sense, i.e.

$$e_\nu(S_{n,n_0}, f) = (\mathbf{E}_{\nu,K}|S_{n,n_0}(f) - S(f)|^2)^{1/2}.$$

Now let us present a helpful result.

LEMMA 3.25. *Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with transition kernel K and initial distribution ν. Then for $i, j \in \mathbb{N}$ with $j \le i$,*

$$\mathbf{E}_{\nu,K}[f(X_i)f(X_j)] = \int_D P^j(fP^{i-j}f)(x)\,\nu(\mathrm{d}x). \tag{3.13}$$

Moreover, if π is a stationary distribution and $\nu = \pi$ then

$$\mathbf{E}_{\pi,K}[f(X_i)f(X_j)] = \langle f, P^{i-j}f\rangle. \tag{3.14}$$

Proof. The calculation

$$\begin{aligned}
\mathbf{E}_{\nu,K}[f(X_i)f(X_j)] &= \underbrace{\int_D \dots \int_D}_{i\text{ times}} f(x_i)f(x_j)\,K(x_{i-1},\mathrm{d}x_i)\dots K(x_1,\mathrm{d}x_2)\,\nu(\mathrm{d}x_1)\\
&= \underbrace{\int_D \dots \int_D}_{j\text{ times}} f(x_j)P^{i-j}f(x_j)\,K(x_{j-1},\mathrm{d}x_j)\dots K(x_1,\mathrm{d}x_2)\,\nu(\mathrm{d}x_1)\\
&= \int_D P^j(fP^{i-j}f)(x)\,\nu(\mathrm{d}x)
\end{aligned}$$

proves (3.13), and by (3.3) one can see (3.14). ∎

First we assume that the initial distribution of the Markov chain is stationary. Hence it is not necessary to do any burn-in, i.e. $n_0 = 0$. The resulting method is denoted by S_n instead of $S_{n,0}$. Afterwards we turn to the general method S_{n,n_0} where the initial distribution might differ from a stationary one.

In the next statement we assume that the transition kernel is reversible with respect to π. Then we can apply the spectral theorem for self-adjoint bounded linear operators (see Theorem A.2).

PROPOSITION 3.26. *Let $f \in L_2$ and $g = f - S(f)$. Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with transition kernel K and initial distribution π, let K be reversible with respect to π and let*

$$\lambda = \inf\{\alpha \mid \alpha \in \mathbf{spec}(P|L_2^0)\}, \qquad \Lambda = \sup\{\alpha \mid \alpha \in \mathbf{spec}(P|L_2^0)\}.$$

Suppose that $\Lambda < 1$. Then

$$e_\pi(S_n, f)^2 = \frac{1}{n^2}\int_\lambda^\Lambda W(n,\alpha)\,\mathrm{d}\langle E_{\{\alpha\}}g, g\rangle = \frac{1}{n^2}\langle W(n,P)g, g\rangle, \tag{3.15}$$

where E denotes the spectral measure ([2]) *which corresponds to $P\colon L_2^0 \to L_2^0$ and recall that*

$$W(n,\alpha) = \frac{n(1-\alpha^2) - 2\alpha(1-\alpha^n)}{(1-\alpha)^2}, \quad \alpha \in [-1,1).$$

Proof. Since $f \in L_2$ we have $g \in L_2^0$. The error obeys

$$\begin{aligned} e_\pi(S_n,f)^2 &= \mathbf{E}_{\pi,K}\left|\frac{1}{n}\sum_{j=1}^n g(X_j)\right|^2 = \frac{1}{n^2}\,\mathbf{E}_{\pi,K}\left|\sum_{j=1}^n g(X_j)\right|^2 \\ &= \frac{1}{n^2}\sum_{j=1}^n \mathbf{E}_{\pi,K}[g(X_j)^2] + \frac{2}{n^2}\sum_{j=1}^{n-1}\sum_{i=j+1}^{n} \mathbf{E}_{\pi,K}[g(X_j)g(X_i)]. \end{aligned}$$

For $i,j \in \mathbb{N}$ with $j \le i$ we obtain

$$\mathbf{E}_{\pi,K}[g(X_i)g(X_j)] \underset{(3.14)}{=} \langle g, P^{i-j}g\rangle = \int_\lambda^\Lambda \alpha^{i-j}\,\mathrm{d}\langle E_{\{\alpha\}}g,g\rangle,$$

where the last equality is an application of Theorem A.2. Altogether this gives

$$\begin{aligned} e_\pi(S_n,f)^2 &= \frac{1}{n^2}\int_\lambda^\Lambda \Big[n + 2\sum_{j=1}^{n-1}\sum_{i=j+1}^{n}\alpha^{i-j}\Big]\mathrm{d}\langle E_{\{\alpha\}}g,g\rangle \\ &= \frac{1}{n^2}\int_\lambda^\Lambda \left[n + 2\frac{(n-1)\alpha - n\alpha^2 + \alpha^{n+1}}{(1-\alpha)^2}\right]\mathrm{d}\langle E_{\{\alpha\}}g,g\rangle \\ &= \frac{1}{n^2}\int_\lambda^\Lambda W(n,\alpha)\,\mathrm{d}\langle E_{\{\alpha\}}g,g\rangle = \frac{1}{n^2}\langle W(n,P)g,g\rangle. \ \blacksquare \end{aligned}$$

By the spectral theorem we have a representation of the error depending on the Markov operator P. In this setting one can show a relation between the operator norm of $W(n,P)\colon L_2^0 \to L_2^0$ and the maximal error of S_n for integrands f which satisfy $\|f\|_2 \le 1$:

COROLLARY 3.27. *Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with transition kernel K and initial distribution π, let K be reversible with respect to π and suppose that $\Lambda < 1$. Then*

$$\sup_{\|f\|_2\le 1} e_\pi(S_n,f)^2 = \frac{1}{n^2}\|W(n,P)\|_{L_2^0\to L_2^0} = \frac{1+\Lambda}{n(1-\Lambda)} - \frac{2\Lambda(1-\Lambda^n)}{n^2(1-\Lambda)^2} \le \frac{2}{n(1-\Lambda)}.$$

Proof. The last inequality follows by Lemma 2.12. The mapping $\alpha \mapsto W(n,\alpha)$ of Proposition 3.26 is increasing (see also Lemma 2.12). For $g = f - S(f)$ we have

$$\begin{aligned} e_\pi(S_n,f)^2 &= \frac{1}{n^2}\int_\lambda^\Lambda W(n,\alpha)\,\mathrm{d}\langle E_{\{\alpha\}}g,g\rangle \le \frac{1}{n^2}W(n,\Lambda)\int_\lambda^\Lambda \mathrm{d}\langle E_{\{\alpha\}}g,g\rangle \\ &= \frac{1}{n^2}W(n,\Lambda)\langle g,g\rangle = \left(\frac{1+\Lambda}{n(1-\Lambda)} - \frac{2\Lambda(1-\Lambda^n)}{n^2(1-\Lambda)^2}\right)\|g\|_2^2. \end{aligned}$$

([2]) The definition of a spectral measure and the spectral theorem for linear, bounded self-adjoint operators are stated in Section A.1.

The assertion is proven by

$$W(n,\Lambda) = \max_{\alpha\in\mathbf{spec}(P|L_2^0)} |W(n,\alpha)| = \|W(n,P)\|_{L_2^0\to L_2^0} = \sup_{\|g\|_2\le 1,\, g\in L_2^0} \langle W(n,P)g, g\rangle$$
$$= \sup_{\|g\|_2\le 1,\, g\in L_2^0} n^2 \cdot e_\pi(S_n,g)^2 \le n^2 \sup_{\|f\|_2\le 1} e_\pi(S_n,f)^2. \ \blacksquare$$

If the transition kernel K is reversible with respect to π and the Markov operator has an L_2-spectral gap, then

$$\beta = \|P\|_{L_2^0\to L_2^0} = \max\{\Lambda, |\lambda|\} < 1.$$

Note that Proposition 3.26 holds already if $\Lambda < 1$. Hence an L_2-spectral gap is not necessary. If the transition kernel K is not reversible but one has an L_2-spectral gap, then the following error bound can be shown.

PROPOSITION 3.28. *Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with transition kernel K and initial distribution π. Let π be a stationary distribution of K. Let $f \in L_2$ and assume that there exists an L_2-spectral gap $1-\beta > 0$. Then*

$$e_\pi(S_n,f)^2 \le \frac{2}{n(1-\beta)}\|f\|_2^2. \tag{3.16}$$

Proof. Let $g = f - S(f)$. The error obeys

$$e_\pi(S_n,f)^2 = \frac{1}{n^2}\sum_{j=1}^n \mathbf{E}_{\pi,K}[g(X_j)^2] + \frac{2}{n^2}\sum_{j=1}^{n-1}\sum_{i=j+1}^n \mathbf{E}_{\pi,K}[g(X_j)g(X_i)].$$

For $i,j\in\mathbb{N}$ with $j\le i$, by the Cauchy–Schwarz inequality (CS) we have

$$\mathbf{E}_{\pi,K}[g(X_i)g(X_j)] = \langle g, P^{i-j}g\rangle \underset{(\mathrm{CS})}{\le} \|P^{i-j}\|_{L_2^0\to L_2^0}\|g\|_2^2.$$

Then with $W(n,\beta)$ from Proposition 3.26 one has

$$e_\pi(S_n,f)^2 \le \frac{W(n,\beta)}{n^2}\|g\|_2^2 \underset{(2.6)}{\le} \frac{2}{n(1-\beta)}\|f\|_2^2. \ \blacksquare$$

The estimates of the error under the assumption that the initial distribution is stationary seem to be restrictive. If we could sample π directly we would approximate $S(f)$ by Monte Carlo with an i.i.d. sample. However, even if this is possible it might happen that the direct sampling procedure is computationally expensive, so it is reasonable to generate only the initial state by sampling from π and afterwards run a Markov chain with stationary distribution π.

The error of a Markov chain Monte Carlo method with stationary initial distribution is related to the error with a not necessarily stationary initial distribution.

PROPOSITION 3.29. *Let $r\in[1,2]$, let $f\in L_{2r}$ and let $\nu\in\mathcal{M}_{r/(r-1)}$ be a probability measure. Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with transition kernel K and initial distribution ν and let π be a stationary distribution of K. Then*

$$e_\nu(S_{n,n_0},f)^2 = e_\pi(S_n,f)^2 + \frac{1}{n^2}\sum_{j=1}^n L_{j+n_0}(g^2) + \frac{2}{n^2}\sum_{j=1}^{n-1}\sum_{k=j+1}^n L_{j+n_0}(gP^{k-j}g), \tag{3.17}$$

where $g = f - S(f)$ *and*

$$L_i(h) = \left\langle (P^i - S)h, \frac{d\nu}{d\pi} - 1 \right\rangle, \quad h \in L_r,\ i \in \mathbb{N}.$$

Proof. The proof is adapted from [Rud09, Lemma 6, p. 17]. One has

$$\begin{aligned}\mathbf{E}_{\nu,K}|S(f) - S_{n,n_0}(f)|^2 &= \frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n}\mathbf{E}_{\nu,K}[g(X_{n_0+j})g(X_{n_0+i})] \\ &= \frac{1}{n^2}\sum_{j=1}^{n}\int_D P^{n_0+j}(g^2)(x)\,\nu(\mathrm{d}x) + \frac{2}{n^2}\sum_{j=1}^{n-1}\sum_{k=j+1}^{n}\int_D P^{n_0+j}(gP^{k-j}g)(x)\,\nu(\mathrm{d}x).\end{aligned}$$

For $h \in L_r$ and $\nu \in \mathcal{M}_{r/(r-1)}$ we see for all $i \in \mathbb{N}$ that $\frac{d\nu}{d\pi} \cdot P^i h$ is integrable with respect to π. Then

$$\begin{aligned}\int_D (P^i h)(x)\,\nu(\mathrm{d}x) &= \left\langle P^i h, \frac{d\nu}{d\pi} \right\rangle = \langle P^i h, 1\rangle + \left\langle P^i h, \frac{d\nu}{d\pi} - 1 \right\rangle \\ &= \langle P^i h, 1\rangle + \left\langle P^i h, \frac{d\nu}{d\pi} - 1 \right\rangle - \underbrace{\left\langle h, S\left(\frac{d\nu}{d\pi} - 1\right)\right\rangle}_{=0} \\ &= \langle P^i h, 1\rangle + \left\langle (P^i - S)h, \frac{d\nu}{d\pi} - 1 \right\rangle \\ &= \int_D (P^i h)(x)\,\pi(\mathrm{d}x) + \left\langle (P^i - S)h, \frac{d\nu}{d\pi} - 1 \right\rangle.\end{aligned}$$

Formula (3.17) is shown by using this calculation for $h = g^2$ and $h = gP^{k-j}g$. ∎

Equation (3.17) is still an exact error formula. The next lemma provides an estimate of the functional $L_k(\cdot)$ for $k \in \mathbb{N}$.

LEMMA 3.30. *Let* $r \in [1,2]$, $\nu \in \mathcal{M}_{r/(r-1)}$ *and* $h \in L_r$. *Recall that* $\beta = \|P\|_{L_2^0 \to L_2^0}$.

(i) *If* $r \in (1,2]$, *then*

$$|L_k(h)| \le 2^{2/r}\beta^{2k\frac{r-1}{r}}\left\|\frac{d\nu}{d\pi} - 1\right\|_{\frac{r}{r-1}}\|h\|_r, \quad k \in \mathbb{N}. \tag{3.18}$$

(ii) *If* $r = 1$ *and the transition kernel is* L_1*-exponentially convergent with* (α, M), *then*

$$|L_k(h)| \le M\alpha^k\left\|\frac{d\nu}{d\pi} - 1\right\|_{\infty}\|h\|_1, \quad k \in \mathbb{N}. \tag{3.19}$$

Proof. After applying Hölder's inequality (HI) with conjugate parameter r and $s = \frac{r}{r-1}$ to $L_k(h) = \langle (P^k - S)h, d\nu/d\pi - 1\rangle$ one has

$$|L_k(h)| \underset{(\mathrm{HI})}{\le} \|(P^k - S)h\|_r\left\|\frac{d\nu}{d\pi} - 1\right\|_s \le \|P^k - S\|_{L_r \to L_r}\left\|\frac{d\nu}{d\pi} - 1\right\|_s\|h\|_r.$$

By equation (3.7) the claim of (i) is proven and by the L_1-exponential convergence the inequality of (ii) holds. ∎

Note that if $r = 2$ then $|L_k(h)| \le \beta^k\|d\nu/d\pi - 1\|_2\|h\|_2$ (see (3.5)). This is better than (3.18) by a factor of two, but not essentially different.

In Lemma 3.30 we have seen that under suitable assumptions one can ensure exponential decay of $L_k(\cdot)$ for increasing $k \in \mathbb{N}$. This fact is used to show for reversible Markov chains which are L_1-exponentially ergodic with (α, M) that there exists a constant $C_{\nu,\alpha,M}$, independent of n and n_0, such that

$$|e_\nu(S_{n,n_0}, f)^2 - e_\pi(S_n, f)^2| \le C_{\nu,\alpha,M} \|f\|_2^2 \frac{\alpha^{n_0}}{n^2}.$$

An immediate consequence of this inequality is an explicit error bound. The following lemma and remark imply such an inequality and provide $C_{\nu,\alpha,M}$ explicitly.

LEMMA 3.31. *Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with transition kernel K and initial distribution ν, where $\nu \in \mathcal{M}_\infty$. Let K be reversible with respect to π and L_1-exponentially convergent with (α, M). Let $f \in L_2$ and*

$$U(\alpha, n) = \sum_{j=1}^{n} \alpha^j + 2 \sum_{j=1}^{n-1} \sum_{k=j+1}^{n} \alpha^k.$$

Then

$$|e_\nu(S_{n,n_0}, f)^2 - e_\pi(S_n, f)^2| \le \frac{U(\alpha, n)}{n^2} M \left\| \frac{d\nu}{d\pi} - 1 \right\|_\infty \alpha^{n_0} \|f\|_2^2. \tag{3.20}$$

Proof. Let $g = f - S(f)$. Equation (3.17) implies

$$|e_\nu(S_{n,n_0}, f)^2 - e_\pi(S_n, f)^2| \le \frac{1}{n^2} \sum_{j=1}^{n} |L_{j+n_0}(g^2)| + \frac{2}{n^2} \sum_{j=1}^{n-1} \sum_{k=j+1}^{n} |L_{j+n_0}(gP^{k-j}g)|.$$

By (3.19) of Lemma 3.30 one obtains

$$|L_{j+n_0}(g^2)| \le M\alpha^{j+n_0} \left\| \frac{d\nu}{d\pi} - 1 \right\|_\infty \|g\|_2^2,$$
$$|L_{j+n_0}(gP^{k-j}g)| \le M\alpha^{j+n_0} \left\| \frac{d\nu}{d\pi} - 1 \right\|_\infty \|gP^{k-j}g\|_1.$$

By the reversibility and L_1-exponential convergence of K we see from Proposition 3.24 that $\beta = \|P\|_{L_2^0 \to L_2^0} \le \alpha$. Then by the Cauchy–Schwarz inequality (CS),

$$\|gP^{k-j}g\|_1 \underset{(\mathrm{CS})}{\le} \|g\|_2 \|P^{k-j}g\|_2 \le \|g\|_2^2 \|P^{k-j}\|_{L_2^0 \to L_2^0} \le \alpha^{k-j} \|g\|_2^2.$$

Let $\varepsilon_0 = \alpha^{n_0} M \left\| d\nu/d\pi - 1 \right\|_\infty$. Then

$$\begin{aligned}
&\sum_{j=1}^{n} |L_{j+n_0}(g^2)| + 2 \sum_{j=1}^{n-1} \sum_{k=j+1}^{n} |L_{j+n_0}(gP^{k-j}g)| \\
&\qquad \le \varepsilon_0 \|g\|_2^2 \sum_{j=1}^{n} \alpha^j + 2\varepsilon_0 \|g\|_2^2 \sum_{j=1}^{n-1} \sum_{k=j+1}^{n} \alpha^k \\
&\qquad = \varepsilon_0 \|g\|_2^2 \Big(\sum_{j=1}^{n} \alpha^j + 2 \sum_{j=1}^{n-1} \sum_{k=j+1}^{n} \alpha^k \Big) \\
&\qquad = \varepsilon_0 \cdot U(\alpha, n) \cdot \|g\|_2^2 \le \varepsilon_0 \cdot U(\alpha, n) \cdot \|f\|_2^2. \quad \blacksquare
\end{aligned}$$

REMARK 3.32. The function $U(\alpha, n)$ has already been studied in Lemma 2.19. Let us repeat the result. For all $n \in \mathbb{N}$ we have

$$U(\alpha, n) \leq \frac{2}{(1-\alpha)^2}.$$

Then from Lemma 3.31 it follows that

$$e_\nu(S_{n,n_0}, f)^2 \leq e_\pi(S_n, f)^2 + \frac{2M\left\|\frac{d\nu}{d\pi} - 1\right\|_\infty \alpha^{n_0}}{n^2(1-\alpha)^2}\|f\|_2^2.$$

If the initial distribution ν is π then one has the error formula of Proposition 3.26.

REMARK 3.33. Note that in Lemma 3.31 reversibility of K was essentially used to apply Proposition 3.24. If the Markov operator is normal, i.e. $PP^* = P^*P$, then by Proposition 3.19, $\beta = \|P\|_{L_2^0 \to L_2^0} \leq \sqrt{\alpha}$. By this observation we get a very similar estimate to that in Lemma 3.31 for normal Markov operators which are not necessarily reversible. The only difference with (3.20) is that α has to be replaced by $\sqrt{\alpha}$. Then

$$U(\sqrt{\alpha}, n) \leq \frac{2}{(1-\sqrt{\alpha})^2} \leq \frac{8}{(1-\alpha)^2}.$$

The last inequality follows from $1 - \alpha^r \geq r(1-\alpha)$ for $r \in [0,1]$, which is a consequence of the *Bernoulli inequality with a real exponent* ([3]) .

The next theorem summarizes the main result for a Markov chain with a reversible and L_1-exponentially convergent transition kernel.

THEOREM 3.34. *Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with transition kernel K and initial distribution ν. Let K be reversible with respect to π and L_1-exponentially convergent with (α, M). Let $f \in L_2$ and assume that $\nu \in \mathcal{M}_\infty$. Then*

$$e_\nu(S_{n,n_0}, f)^2 \leq \frac{2}{n(1-\Lambda)}\|f\|_2^2 + \frac{2M\left\|\frac{d\nu}{d\pi} - 1\right\|_\infty \alpha^{n_0}}{n^2(1-\alpha)^2}\|f\|_2^2 \tag{3.21}$$

and for $g = f - S(f)$ we have

$$\lim_{n\to\infty} n e_\nu(S_{n,n_0}, f)^2 = \lim_{n\to\infty} n e_\pi(S_n, f)^2 = \langle (I+P)(I-P)^{-1} g, g\rangle. \tag{3.22}$$

Proof. By Lemmas 3.31 and 2.19 the first equality of (3.22) holds true. By the reversibility of the transition kernel Proposition 3.26 applies, so that

$$\lim_{n\to\infty} n e_\pi(S_n, f)^2 = \lim_{n\to\infty} \frac{1}{n}\langle W(n,P)g, g\rangle = \langle (I+P)(I-P)^{-1} g, g\rangle.$$

The rest follows via Lemma 3.31, Corollary 3.27 and Lemma 2.19. ∎

REMARK 3.35. Under the assumptions of Theorem 3.34, Proposition 3.24 shows that π-a.e. uniform ergodicity with $(\alpha, \widetilde{M})$ is equivalent to L_1-exponential convergence with $(\alpha, 2\widetilde{M})$. Hence one can restate Theorem 3.34 for uniformly ergodic Markov chains and obtains the same result with $M = 2\widetilde{M}$. This is the general state space counterpart to Theorem 2.20, where $\widetilde{M}$ is of the magnitude of $\|1/\pi\|_\infty$ and $\beta = \alpha$.

([3]) The Bernoulli inequality with real exponent $r \in [0,1]$ states for any real number $x > -1$ that $(1+x)^r \leq 1 + rx$.

Furthermore note that if the Markov operator is normal and not necessarily reversible, then one can get a similar error bound by using Remark 3.33.

REMARK 3.36. The error bound of (3.21) might be interpreted as follows: The burn-in n_0 is reasonable to eliminate the influence of the initial distribution, while n has to decrease $e_\pi(S_n, f)$. For large n the error behaves exactly as the error where one started with the stationary distribution. Hence the bias of the initial distribution disappears after sufficiently many steps. If the initial distribution falls together with the stationary one, then the bias of the initial part vanishes completely.

Another consequence of Lemmas 3.31 and 2.19 is the following result concerning the asymptotic error for $\|f\|_2 \leq 1$.

COROLLARY 3.37. *Under the same assumptions as in Theorem* 3.34,

$$\lim_{n\to\infty} n \sup_{\|f\|_2\leq 1} e_\nu(S_{n,n_0}, f)^2 = \frac{1+\Lambda}{1-\Lambda}$$

and

$$\lim_{n_0\to\infty} \sup_{\|f\|_2\leq 1} e_\nu(S_{n,n_0}, f)^2 = \frac{1+\Lambda}{n(1-\Lambda)} - \frac{2\Lambda(1-\Lambda^n)}{n^2(1-\Lambda)^2}.$$

Proof. Let us define

$$c_{n,n_0} = \frac{2\alpha^{n_0} M \|\frac{d\nu}{d\pi} - 1\|_\infty}{n^2(1-\alpha)^2}.$$

One has $\lim_{n\to\infty} nc_{n,n_0} = 0$ and $\lim_{n_0\to\infty} c_{n,n_0} = 0$. For $\|f\|_2 \leq 1$ Lemmas 3.31 and 2.19 yield

$$|e_\nu(S_{n,n_0}, f)^2 - e_\pi(S_n, f)^2| \leq c_{n,n_0}.$$

Hence

$$\sup_{\|f\|_2\leq 1} e_\pi(S_n, f)^2 - c_{n,n_0} \leq \sup_{\|f\|_2\leq 1} e_\nu(S_{n,n_0}, f)^2 \leq \sup_{\|f\|_2\leq 1} e_\pi(S_n, f)^2 + c_{n,n_0}. \tag{3.23}$$

Recall that $\Lambda = \sup\{\alpha \mid \alpha \in \mathbf{spec}(P|L_2^0)\}$. Then by Corollary 3.27 we have

$$\sup_{\|f\|_2\leq 1} e_\pi(S_n, f)^2 = \frac{1+\Lambda}{n(1-\Lambda)} - \frac{2\Lambda(1-\Lambda^n)}{n^2(1-\Lambda)^2}.$$

Taking the limits in (3.23) yields the assertions. ∎

In many examples it is known that the transition kernel is L_1-exponentially convergent or π-a.e. uniformly ergodic, but it is difficult to obtain reasonable values of (α, M) explicitly. Then at least the asymptotic result can be used. This is similar to results of [Sok97, Bré99, Mat99].

REMARK 3.38. Observe that we have a lower and an upper bound of the error of S_{n,n_0}. Exactly as in Remark 2.24 one deduces by (3.23) that

$$\frac{1+\Lambda}{n(1-\Lambda)} - \frac{2}{n^2(1-\Lambda)^2} - c_{n,n_0} \leq \sup_{\|f\|_2\leq 1} e_\nu(S_{n,n_0}, f)^2 \leq \frac{2}{n(1-\Lambda)} + c_{n,n_0}.$$

We showed an error bound of S_{n,n_0} with respect to $\|\cdot\|_2$ for Markov chains which are reversible and L_1-exponentially convergent. The condition of L_1-exponential convergence is rather restrictive. This motivates the study of Markov chains which satisfy a weaker

convergence property, namely we assume that there is an L_2-spectral gap, i.e. $1-\beta>0$. This is enough to obtain error bounds for integrands $f\in L_p$ with $p\in(2,\infty]$. The following lemmas lead to show that there exists a constant $C_{\nu,\beta,p}$, independent of n_0 and n, such that

$$|e_\nu(S_{n,n_0},f)^2-e_\pi(S_n,f)^2|\le C_{\nu,\beta,p}\|f\|_p^2\frac{\beta^{n_0}}{n^2}.$$

Note that it is not assumed that the Markov chain is reversible with respect to π.

LEMMA 3.39. *Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with transition kernel K and initial distribution ν. Let π be a stationary distribution of K. Let $f\in L_p$, let $\nu\in\mathcal{M}_{\max\{2,p/(p-2)\}}$ with $p\in(2,\infty]$ and*

$$V(\beta,n,p)=4\begin{cases}2^{4/p}\displaystyle\sum_{j=1}^{n}\beta^{2j\frac{p-2}{p}}+2^{\frac{3p+2}{p}}\sum_{j=1}^{n-1}\beta^{2j\frac{p-3}{p}}\sum_{k=j+1}^{n}\beta^{2k/p}, & p\in(2,4),\\ 2\displaystyle\sum_{j=1}^{n}\beta^{j}+2^{\frac{3p+2}{p}}\sum_{j=1}^{n-1}\beta^{2j/p}\sum_{k=j+1}^{n}\beta^{k\frac{p-2}{p}}, & p\in[4,\infty].\end{cases}$$

(i) *If $p\in(2,4)$, then*

$$|e_\nu(S_{n,n_0},f)^2-e_\pi(S_n,f)^2|\le\frac{V(\beta,n,p)}{n^2}\,\beta^{2n_0\frac{p-2}{p}}\left\|\frac{d\nu}{d\pi}-1\right\|_{\frac{p}{p-2}}\|f\|_p^2.$$

(ii) *If $p\in[4,\infty]$, then*

$$|e_\nu(S_{n,n_0},f)^2-e_\pi(S_n,f)^2|\le\frac{V(\beta,n,p)}{n^2}\,\beta^{n_0}\left\|\frac{d\nu}{d\pi}-1\right\|_{2}\|f\|_p^2.$$

Proof. First, let $g=f-S(f)$ and observe that for $p\ge1$ one obtains

$$\|g\|_p\le\|f\|_p+|S(f)|\le\|f\|_p+\|f\|_1\le2\|f\|_p. \tag{3.24}$$

Equation (3.17) implies

$$\begin{aligned}|e_\nu(S_{n,n_0},f)^2-e_\pi(S_n,f)^2|&\\ \le\frac{1}{n^2}\sum_{j=1}^{n}|L_{j+n_0}(g^2)|+\frac{2}{n^2}\sum_{j=1}^{n-1}\sum_{k=j+1}^{n}|L_{j+n_0}(gP^{k-j}g)|.&\end{aligned} \tag{3.25}$$

Let $p\in(2,4)$. Then it follows by (3.18) with $r=\frac{p}{2}$ and $r/(r-1)=\frac{p}{p-2}$ that

$$|L_{j+n_0}(g^2)|\le2^{4/p}\beta^{2j\frac{p-2}{p}}\,\beta^{2n_0\frac{p-2}{p}}\left\|\frac{d\nu}{d\pi}-1\right\|_{\frac{p}{p-2}}\|g\|_p^2,$$

$$|L_{j+n_0}(gP^{k-j}g)|\le2^{4/p}\beta^{2j\frac{p-2}{p}}\,\beta^{2n_0\frac{p-2}{p}}\left\|\frac{d\nu}{d\pi}-1\right\|_{\frac{p}{p-2}}\|gP^{k-j}g\|_{p/2}.$$

By applying the Cauchy–Schwarz inequality (CS) and (3.7) one obtains

$$\|gP^{k-j}g\|_{p/2}\underset{(\mathrm{CS})}{\le}\|g\|_p\|P^{k-j}g\|_p\le\|g\|_p^2\|P^{k-j}\|_{L_p^0\to L_p^0}\underset{(3.7)}{\le}2^{2\frac{p-1}{p}}\beta^{2\frac{k-j}{p}}\|g\|_p^2.$$

Let $\varepsilon_0(p) = \beta^{2n_0\frac{p-2}{p}}\|d\nu/d\pi - 1\|_{\frac{p}{p-2}}$. Then

$$\begin{aligned}
&\sum_{j=1}^{n}|L_{j+n_0}(g^2)| + 2\sum_{j=1}^{n-1}\sum_{k=j+1}^{n}|L_{j+n_0}(gP^{k-j}g)| \\
&\qquad\le 2^{4/p}\varepsilon_0(p)\|g\|_p^2\sum_{j=1}^{n}\beta^{2j\frac{p-2}{p}} + 2^{\frac{3p+2}{p}}\varepsilon_0(p)\|g\|_p^2\sum_{j=1}^{n-1}\beta^{2j\frac{p-3}{p}}\sum_{k=j+1}^{n}\beta^{2k/p} \\
&\qquad= \varepsilon_0(p)\|g\|_p^2\Big(2^{4/p}\sum_{j=1}^{n}\beta^{2j\frac{p-2}{p}} + 2^{\frac{3p+2}{p}}\sum_{j=1}^{n-1}\beta^{2j\frac{p-3}{p}}\sum_{k=j+1}^{n}\beta^{2k/p}\Big) \\
&\qquad\underset{(3.24)}{\le} V(\beta,n,p)\cdot\varepsilon_0(p)\|f\|_p^2.
\end{aligned}$$

Thus, claim (i) is proved.

Let us turn to (ii), i.e. $p\in[4,\infty]$. Equation (3.18) with $r=2$ is used to get

$$\begin{aligned}
|L_{j+n_0}(g^2)| &\le 2\beta^{j+n_0}\left\|\frac{d\nu}{d\pi}-1\right\|_2\|g\|_4^2, \\
|L_{j+n_0}(gP^{k-j}g)| &\le 2\beta^{j+n_0}\left\|\frac{d\nu}{d\pi}-1\right\|_2\|gP^{k-j}g\|_2.
\end{aligned}$$

By Hölder's inequality (HI) with conjugate parameters $\frac{p}{2}$ and $\frac{p}{p-2}$ one obtains

$$\begin{aligned}
\|gP^{k-j}g\|_2 &\underset{(\mathrm{HI})}{\le} \|g\|_p\|P^{k-j}g\|_{\frac{2p}{p-2}} \le \|P^{k-j}\|_{L^0_{2p/(p-2)}\to L^0_{2p/(p-2)}}\|g\|_p\|g\|_{\frac{2p}{p-2}} \\
&\le \|P^{k-j}\|_{L^0_{2p/(p-2)}\to L^0_{2p/(p-2)}}\|g\|_p^2 \underset{(3.7)}{\le} 2^{\frac{p+2}{p}}\beta^{(k-j)\frac{p-2}{p}}\|g\|_p^2.
\end{aligned}$$

Note that in the third inequality above it was essential that $p\in[4,\infty]$ for using $\|g\|_{\frac{2p}{p-2}} \le \|g\|_p$. Thus, for $\varepsilon_0 = \beta^{n_0}\|\frac{d\nu}{d\pi}-1\|_2$ one has

$$\begin{aligned}
&\sum_{j=1}^{n}|L_{j+n_0}(g^2)| + 2\sum_{j=1}^{n-1}\sum_{k=j+1}^{n}|L_{j+n_0}(gP^{k-j}g)| \\
&\qquad\le \varepsilon_0\|g\|_4^2 2\sum_{j=1}^{n}\beta^j + \varepsilon_0\|g\|_p^2 2^{2+\frac{p+2}{p}}\sum_{j=1}^{n-1}\beta^{2j/p}\sum_{k=j+1}^{n}\beta^{k\frac{p-2}{p}} \\
&\qquad\le \varepsilon_0\|g\|_p^2\Big(2\sum_{j=1}^{n}\beta^j + 2^{\frac{3p+2}{p}}\sum_{j=1}^{n-1}\beta^{2j/p}\sum_{k=j+1}^{n}\beta^{k\frac{p-2}{p}}\Big) \underset{(3.24)}{\le} V(\beta,n,p)\cdot\varepsilon_0\|f\|_p^2.
\end{aligned}$$

Finally, substituting this in (3.25) completes the proof. ∎

Let us consider $V(\beta,n,p)$. If $p\in(2,\infty]$ and $1-\beta>0$, then we show that the mapping $n\mapsto V(\beta,n,p)$ is bounded.

Lemma 3.40. *Let $p\in(2,\infty]$ and $1-\beta>0$. For all $n\in\mathbb{N}$,*

$$V(\beta,n,p) \le \frac{64p}{(p-2)(1-\beta)^2}. \tag{3.26}$$

Proof. The inequalities indicated by $(\star)$ below follow from $1-\beta^r \geq r(1-\beta)$ for $r \in [0,1]$. First, let $p \in (2,4)$. By the geometric series one can estimate

$$\begin{aligned}
\frac{V(\beta,n,p)}{4} &= 2^{4/p}\sum_{j=1}^{n}\beta^{2j\frac{p-2}{p}} + 2^{\frac{3p+2}{p}}\sum_{j=1}^{n-1}\beta^{2j\frac{p-3}{p}}\sum_{k=j+1}^{n}\beta^{2k/p} \\
&= 2^{4/p}\sum_{j=1}^{n}\beta^{2j\frac{p-2}{p}} + 2^{\frac{3p+2}{p}}\beta^{2/p}\sum_{j=1}^{n-1}\beta^{2j\frac{p-2}{p}}\sum_{k=0}^{n-j-1}\beta^{2k/p} \\
&\leq 2^{4/p}\sum_{j=1}^{n}\beta^{2j\frac{p-2}{p}} + \frac{2^{\frac{3p+2}{p}}\beta^{2/p}}{1-\beta^{2/p}}\sum_{j=1}^{n-1}\beta^{2j\frac{p-2}{p}} \\
&\leq \frac{2^{4/p}+\beta^{2/p}2^{4/p}(2^{3-2/p}-1)}{1-\beta^{2/p}}\sum_{j=1}^{n}\beta^{2j\frac{p-2}{p}} \leq \frac{2^{3+2/p}}{1-\beta^{2/p}}\sum_{j=1}^{n}\beta^{2j\frac{p-2}{p}} \\
&\underset{p\in(2,4)}{\leq} \frac{16}{(1-\beta^{2/p})(1-\beta^{2(p-2)/p})} \underset{(\star)}{\leq} \frac{4p^2}{(p-2)(1-\beta)^2} \underset{p\in(2,4)}{\leq} \frac{16p}{(p-2)(1-\beta)^2}.
\end{aligned}$$

For $p \in [4,\infty]$, again by the geometric series, we can estimate

$$\begin{aligned}
\frac{V(\beta,n,p)}{4} &= 2\sum_{j=1}^{n}\beta^{j} + 2^{\frac{3p+2}{p}}\sum_{j=1}^{n-1}\beta^{2j/p}\sum_{k=j+1}^{n}\beta^{k\frac{p-2}{p}} \\
&= 2\sum_{j=1}^{n}\beta^{j} + 2^{\frac{3p+2}{p}}\beta^{\frac{p-2}{p}}\sum_{j=1}^{n-1}\beta^{j}\sum_{k=0}^{n-j-1}\beta^{k\frac{p-2}{p}} \\
&\leq \left(2+\frac{2^{\frac{3p+2}{p}}\beta^{\frac{p-2}{p}}}{1-\beta^{\frac{p-2}{p}}}\right)\sum_{j=1}^{n}\beta^{j} \leq \frac{2+\beta^{\frac{p-2}{p}}(2^{\frac{3p+2}{p}}-2)}{1-\beta^{\frac{p-2}{p}}}\sum_{j=1}^{n}\beta^{j} \\
&\underset{p\in[4,\infty]}{\leq} \frac{8\sqrt{2}}{1-\beta^{\frac{p-2}{p}}}\sum_{j=1}^{n}\beta^{j} \leq \frac{8\sqrt{2}}{(1-\beta)(1-\beta^{\frac{p-2}{p}})} \underset{(\star)}{\leq} \frac{8\sqrt{2}\,p}{(p-2)(1-\beta)^2}. \qquad \blacksquare
\end{aligned}$$

The main error bound of S_{n,n_0} for Markov chains with an L_2-spectral gap is presented in the next theorem.

THEOREM 3.41. *Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with transition kernel K and initial distribution ν. Let π be a stationary distribution of K. For $p \in (2,\infty]$ let $f \in L_p$ and $\nu \in \mathcal{M}_{\max\{2,p/(p-2)\}}$. Suppose that the Markov operator has an L_2-spectral gap, i.e. $1-\beta > 0$. Then*

$$e_\nu(S_{n,n_0},f)^2 \leq e_\pi(S_n,f)^2 + \frac{64p\|f\|_p^2}{n^2(p-2)(1-\beta)^2}\begin{cases}\beta^{2n_0\frac{p-2}{p}}\left\|\frac{d\nu}{d\pi}-1\right\|_{\frac{p}{p-2}}, & p\in(2,4),\\ \beta^{n_0}\left\|\frac{d\nu}{d\pi}-1\right\|_2, & p\in[4,\infty],\end{cases}$$

where

$$e_\pi(S_n,f)^2 \leq \begin{cases}\dfrac{2}{n(1-\Lambda)}\|f\|_p & \text{if } K \text{ is reversible with respect to } \pi,\\ \dfrac{2}{n(1-\beta)}\|f\|_p & \text{otherwise.}\end{cases}$$

Furthermore

$$\lim_{n\to\infty} ne_\nu(S_{n,n_0},f)^2 = \lim_{n\to\infty} ne_\pi(S_n,f)^2 \tag{3.27}$$

and if K is reversible with respect to π then (3.27) *is equal to*

$$\langle (I+P)(I-P)^{-1}g,g\rangle, \quad \textit{where} \quad g=f-S(f).$$

Proof. By Lemmas 3.39 and 3.40 the equality in (3.27) is true. If the transition kernel is reversible, then by Proposition 3.26 the asymptotic result holds since

$$\lim_{n\to\infty} ne_\pi(S_n,f)^2 = \lim_{n\to\infty} \frac{1}{n}\langle W(n,P)g,g\rangle = \langle (I+P)(I-P)^{-1}g,g\rangle.$$

By Lemmas 3.39 and 3.40 one obtains the estimate of $e_\nu(S_{n,n_0},f)^2$. The estimate of $e_\pi(S_n,f)^2$ follows by Proposition 3.28 and for a reversible transition kernel by Corollary 3.27. ∎

REMARK 3.42. A large burn-in n_0 guarantees that the influence of the initial distribution disappears and a large n makes $e_\pi(S_n,f)$ small. The condition of L_1-exponential convergence could be replaced by the existence of an L_2-spectral gap by paying the price of considering error bounds in terms of L_p-norms of the integrand for $p\in(2,\infty]$. If p converges to 2, then the bound goes to infinity. However, for $p>2$ one has an explicit error bound. If the initial and stationary distribution is the same, then the influence of the initial part vanishes for all $p\in(2,\infty]$.

REMARK 3.43. Let

$$c_{n,n_0}(p) = \frac{64p}{n^2(p-2)(1-\beta)^2} \begin{cases} \beta^{2n_0\frac{p-2}{p}} \left\| \dfrac{d\nu}{d\pi} - 1 \right\|_{\frac{p}{p-2}}, & p\in(2,4), \\ \beta^{n_0} \left\| \dfrac{d\nu}{d\pi} - 1 \right\|_2, & p\in[4,\infty]. \end{cases}$$

For $\|f\|_p \le 1$ we have, by Lemmas 3.39 and 3.40,

$$|e_\nu(S_{n,n_0},f)^2 - e_\pi(S_n,f)^2| \le c_{n,n_0}(p).$$

Observe that this implies a lower error bound for S_{n,n_0}. We do not use it because of the lack of a general lower bound of $\sup_{\|f\|_p\le 1} e_\pi(S_n,f)^2$ for $p\in(2,\infty]$.

REMARK 3.44. Let K be a transition kernel which is reversible with respect to π. We use the notation $\beta_K=\beta$ and $\Lambda_K=\Lambda$ to indicate the transition kernel. The lazy version of K is given by $\widetilde{K}$. Then

$$\beta_{\widetilde{K}} = \Lambda_{\widetilde{K}} = \tfrac{1}{2}(1+\Lambda_K).$$

If one has an estimate of Λ_K, then one also has an estimate of $\beta_{\widetilde{K}}$ and one can apply Theorem 3.41. There are some techniques, e.g. canonical paths (see [Yue00]) and the conductance concept (see [LS88, LS93] and [JS89, DS91]), which are helpful to estimate Λ_K. However, in general it is a challenging task.

3.3. Burn-in. Assume that computational resources for $N=n+n_0$ steps of the Markov chain are available. The burn-in n_0 and the sample size n should be chosen such that the error bound is as small as possible. One encounters the same trade-off as for finite state spaces. In the next statement the error bound for an explicit burn-in is stated.

THEOREM 3.45.

(i) *Suppose that we have a Markov chain which is reversible with respect to π and L_1-exponentially convergent with (α, M). Let*

$$n_0 = \max\left\{\left\lceil \frac{\log(M\left\|\frac{d\nu}{d\pi}-1\right\|_\infty)}{\log(\alpha^{-1})}\right\rceil, 0\right\}.$$

Then

$$\sup_{\|f\|_2\le 1} e_\nu(S_{n,n_0},f)^2 \le \frac{2}{n(1-\beta)} + \frac{2}{n^2(1-\alpha)^2} \le \frac{2}{n(1-\alpha)} + \frac{2}{n^2(1-\alpha)^2}.$$

(ii) *Suppose that we have a Markov chain with Markov operator P which has an L_2-spectral gap $1-\beta > 0$. For $p \in (2,\infty]$ let $n_0(p)$ be the smallest natural number (including zero) which is greater than or equal to*

$$\frac{1}{\log(\beta^{-1})}\begin{cases} \frac{p}{2(p-2)}\log\left(\frac{32p}{p-2}\left\|\frac{d\nu}{d\pi}-1\right\|_{\frac{p}{p-2}}\right), & p\in(2,4),\\ \log\left(64\left\|\frac{d\nu}{d\pi}-1\right\|_2\right), & p\in[4,\infty]. \end{cases}$$

Then

$$\sup_{\|f\|_p\le 1} e_\nu(S_{n,n_0(p)},f)^2 \le \frac{2}{n(1-\beta)} + \frac{2}{n^2(1-\beta)^2}.$$

Proof. Assertion (i) follows from Theorem 3.34 and Proposition 3.24. Claim (ii) is an application of Theorem 3.41. ∎

Note that $\log(\beta^{-1}) = (1-\beta) + \sum_{j=2}^\infty (1-\beta)^j/j!$ and $\log(\beta^{-1}) \ge 1-\beta$. This can be used to estimate the suggested burn-in. Now we justify the choice of the burn-in.

For simplicity we assume that $\alpha = \beta$. Let us define

$$C(p) = \begin{cases} M\left\|\frac{d\nu}{d\pi}-1\right\|_\infty, & p=2,\\ \frac{32p}{p-2}\left\|\frac{d\nu}{d\pi}-1\right\|_{\frac{p}{p-2}}, & p\in(2,4),\\ 64\left\|\frac{d\nu}{d\pi}-1\right\|_2, & p\in[4,\infty]. \end{cases}$$

We consider numerical experiments under the following conditions. Suppose that

- the computational resources are either $N = 10^5$ or $N = 10^6$.
- $\beta = 0.9$ or $\beta = 0.99$ or $\beta = 0.999$.
- $C = C(p) = 10^{30}$, independent of p.

Then the suggested burn-in in Theorem 3.45 for $p = 2$ and $p \in [4,\infty]$ has the form

$$n_0^{\{2\}\cup[4,\infty)} = \left\lceil \frac{\log(C)}{\log(\beta^{-1})}\right\rceil,$$

whereas for $p \in (2,4)$ it still depends on p,

$$n_0^{(2,4)} = \left\lceil \frac{p}{2(p-2)}\,\frac{\log(C)}{\log(\beta^{-1})}\right\rceil.$$

The error for $\|f\|_p \le 1$ where $p \in \{2\} \cup [4, \infty)$ is bounded by

$$\mathrm{est}_{\{2\}\cup[4,\infty)}(n, n_0) = \sqrt{\frac{2}{n(1-\beta)} + \frac{2C\beta^{n_0}}{n^2(1-\beta)^2}},$$

whereas for $p \in (2, 4)$ we have the upper estimate

$$\mathrm{est}_{(2,4)}(n, n_0) = \sqrt{\frac{2}{n(1-\beta)} + \frac{2C\beta^{2n_0\frac{p-2}{p}}}{n^2(1-\beta)^2}}.$$

With the restriction $N = n + n_0$ one can numerically compute a burn-in, which approximates the minimal upper error bound. This is a 1-dimensional minimization problem with different parameters. Let us denote the numerically computed values of the burn-in by $n_{\mathrm{opt}}^{\{2\}\cup[4,\infty)}$ for $p \in \{2\} \cup [4, \infty)$ and $n_{\mathrm{opt}}^{(2,4)}$ for $p \in (2, 4)$ respectively.

Table 1. For $C = 10^{30}$ and $p = 2.1$. The numerically computed value $n_{\mathrm{opt}}^{\mathrm{Int}}$ which approximately minimizes the mapping $n_0 \mapsto \mathrm{est}_{\mathrm{Int}}(N - n_0, n_0)$, with either $\mathrm{Int} = \{2\} \cup [4, \infty)$ or $\mathrm{Int} = (2, 4)$.

N	β	$n_{\mathrm{opt}}^{\{2\}\cup[4,\infty)}$ (by Maple)	$n_0^{\{2\}\cup[4,\infty)} = \lceil \frac{\log(C)}{\log(\beta^{-1})} \rceil$ (suggested above)	$n_{\mathrm{opt}}^{(2,4)}$ (by Maple)	$n_0^{(2,4)} = \lceil \frac{p}{2(p-2)} \frac{\log(C)}{\log(\beta^{-1})} \rceil$ (suggested above, $p = 2.1$)
10^5	0.9	656	656	6655	6885
10^6	0.9	656	656	6655	6885
10^5	0.99	6873	6874	69642	72169
10^6	0.99	6874	6874	69715	72169
10^5	0.999	68977	69043	79011	724952
10^6	0.999	69041	69043	699520	724952

Table 1 gives a collection of $n_{\mathrm{opt}}^{\{2\}\cup[4,\infty)}$ and $n_{\mathrm{opt}}^{(2,4)}$ where $p = 2.1$. The suggested n_0 of Theorem 3.45 is close to the numerically computed values of the burn-in, which approximately minimize the error bound. For $N = 10^5$ and $\beta = 0.999$ the difference between $n_{\mathrm{opt}}^{(2,4)}$ and $n_0^{(2,4)}$ is large. In this situation Theorem 3.41 gives an error smaller than 1 for no choice of n and n_0 with $N = 10^5$. The available resources $N = n + n_0$ are too small for the suggested burn-in to be reached. If the computational resources are large enough, then the computed values $n_{\mathrm{opt}}^{\{2\}\cup[4,\infty)}$ and $n_{\mathrm{opt}}^{(2,4)}$ are of the same magnitude as the suggested $n_0^{\{2\}\cup[4,\infty)}$ and $n_0^{(2,4)}$.

If an error of at most $\varepsilon \in (0, 1)$ is desired, then the suggested choice $n_0^{\{2\}\cup[4,\infty)}$ or $n_0^{(2,4)}$, depending on p, of the burn-in is independent of the precision ε. We choose n_0 as suggested in Theorem 3.45 and

$$n \ge \frac{1 + \sqrt{1 + 4\varepsilon^2}}{(1-\beta)\varepsilon^2} \quad \text{to achieve} \quad e_\nu(S_{n,n_0}, f) \le \varepsilon.$$

Let the Markov chain be reversible with respect to π and let $\Lambda = \beta$. For different fixed values n_0 a plot of

$$\mathrm{est}_{\{2\}\cup[4,\infty)}(N - n_0, n_0) \quad \text{and} \quad \sup_{\|f\|_2 \le 1} e_\pi(S_N, f) = \sqrt{\frac{1+\Lambda}{N(1-\Lambda)} - \frac{2\Lambda(1-\Lambda^N)}{N^2(1-\Lambda)^2}}$$

is presented in Figure 2. Roughly speaking, one can see that if the burn-in is chosen too small a vertical shifting takes place and if the burn-in is chosen too large a horizontal shifting appears. Summarizing one can say that if β, C and p are given, then one should

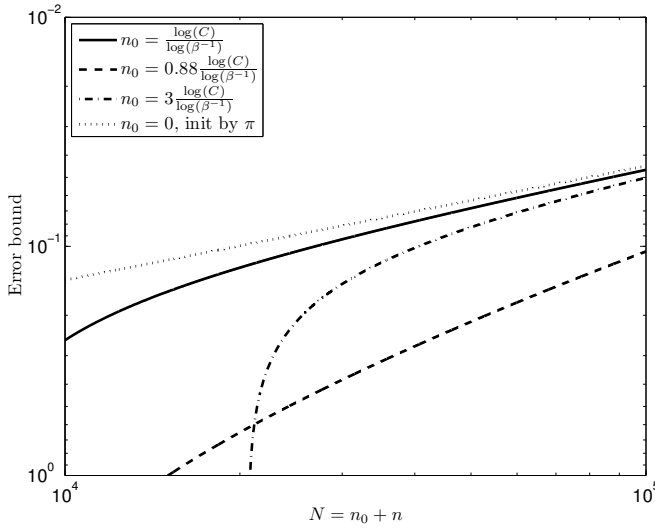

Fig. 2. For $\beta = \Lambda = 0.99$ and $C = 10^{30}$ the mapping $N \mapsto \mathrm{est}_{\{2\}\cup[4,\infty)}(N - n_0, n_0)$ is plotted for different values of n_0. The dotted curve is the graph of the mapping $N \mapsto \sup_{\|f\|_2 \le 1} e_\pi(S_N, f)$.

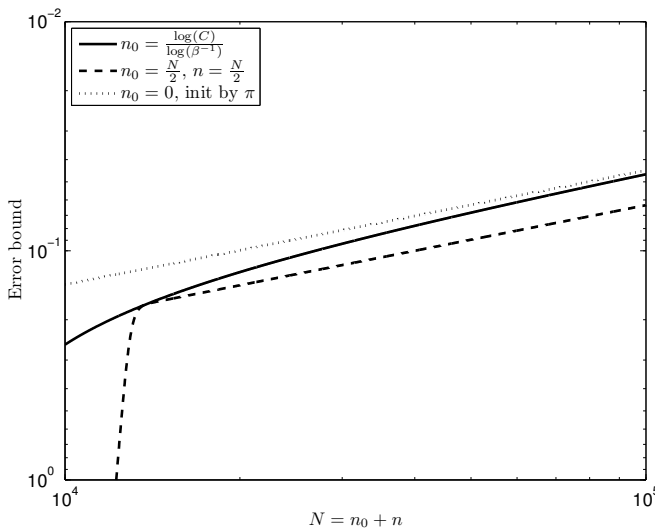

Fig. 3. For $\beta = \Lambda = 0.99$ and $C = 10^{30}$ the mapping $N \mapsto \mathrm{est}_{\{2\}\cup[4,\infty)}(N - n_0, n_0)$ is plotted for different values of n_0. The dotted curve is the graph of the mapping $N \mapsto \sup_{\|f\|_2 \le 1} e_\pi(S_N, f)$.

choose the burn-in as suggested above. If there is an estimate of $\log(C)/\log(\beta^{-1})$, then one should ensure that it is not smaller than the real quotient. As seen in Figure 2, if it

is slightly smaller there is already a strong influence. By choosing the burn-in too large the influence is less heavy.

If nothing is known about β or C another strategy is to choose $n = n_0 = N/2$ for even N. This has the advantage that no information about β or C is needed. In Figure 3 we plotted

$$\mathrm{est}_{\{2\}\cup[4,\infty)}(N/2, N/2),\ \mathrm{est}_{\{2\}\cup[4,\infty)}(N - n_0^{\{2\}\cup[4,\infty)}, n_0^{\{2\}\cup[4,\infty)}) \text{ and } \sup_{\|f\|_2\le 1} e_\pi(S_N, f)$$

where $N \in [10^4, 10^5]$. Asymptotically the price of a factor of $\sqrt{2}$ is paid, i.e. asymptotically the error is $\sqrt{2}$ times worse than $\sup_{\|f\|_2\le 1} e_\pi(S_N, f)$ (see Figure 3). This strategy works well and reaches the same rate of convergence as in Theorem 3.45.

3.4. Examples. For the examples in Section 2.4 one can provide all eigenfunctions and eigenvalues. Usually it is a challenging task to obtain the necessary information on the spectral structure of the Markov operator, in particular on general state spaces. This section contains examples to illustrate the error bounds. The literature provides some tools which can be applied to estimate the quantities of interest, e.g. Λ, β. These tools are briefly introduced. For further details we refer to the literature. Note that the initial distributions of the Markov chains of the following examples are to demonstrate the error bounds and not to minimize the burn-in.

Bounded state spaces. Suppose that the state space D is a measurable subset of $\mathbb{R}^d$. The σ-algebra $\mathfrak{D}$ is given by $\mathcal{B}(D)$. We say a transition kernel K has a *transition density* with respect to a positive measure μ if there is a function $k\colon D \times D \to [0,\infty]$ such that

$$K(x, A) = \int_A k(x, y)\, \mu(\mathrm{d}y), \quad x \in D,\ A \in \mathcal{B}(D).$$

We write k^n for the transition density of K^n.

Let D be a bounded set and let $\varrho\colon D \to [0,\infty]$ be integrable with respect to the Lebesgue measure, with $\int_D \varrho(x)\,\mathrm{d}x > 0$. Then

$$\pi_\varrho(A) = \frac{\int_A \varrho(x)\,\mathrm{d}x}{\int_D \varrho(x)\,\mathrm{d}x}, \quad A \in \mathcal{B}(D),$$

is a probability measure on $(D, \mathcal{B}(D))$. We say ϱ is an *unnormalized density* with respect to the Lebesgue measure if $\int_D \varrho(x)\,\mathrm{d}x \neq 1$. Let K be a transition kernel with transition density k with respect to the Lebesgue measure and assume that π_ϱ is a stationary distribution of K. Furthermore, let $s \in [0,1]$ and define

$$K_s(x, A) = (1 - s)K(x, A) + s\mathbf{1}_A(x), \quad x \in D,\ A \in \mathcal{B}(D).$$

The transition kernel K_s is called the *s-modified transition kernel of* K. If $s = 1/2$ then the lazy version of K is given and if $s = 0$ then one has K. For all $s \in [0,1]$, π_ϱ is a stationary distribution of K_s. The goal is to approximate

$$S(f) = \int_D f(x)\, \pi_\varrho(\mathrm{d}x).$$

One finds for $n \in \mathbb{N}$ that

$$K_s^n(x,A) = \sum_{i=1}^n s^{n-i}(1-s)^i \binom{n}{i} K^i(x,A) + s^n \mathbf{1}_A(x), \quad x \in D,\ A \in \mathcal{B}(D). \tag{3.28}$$

The case $s = 0$ is reasonable if we define $0^0 = 1$. The following lemma gives a condition which implies L_1-exponential convergence of the s-modified transition kernel. For simplicity let us assume that $\int_D \varrho(x)\,\mathrm{d}x = 1$.

LEMMA 3.46. *If there exist $\alpha \in [0,1)$ and $M < \infty$ such that*

$$2s^n + \int_D \operatorname*{ess\,sup}_{y\in D} \left| \sum_{i=1}^n s^{n-i}(1-s)^i \binom{n}{i} \frac{k^i(x,y)}{\varrho(y)} - (1-s^n) \right| \varrho(x)\,\mathrm{d}x \le \alpha^n M, \quad n \in \mathbb{N},$$

then the transition kernel K_s is L_1-exponentially convergent with (α, M).

Proof. The Markov operator of K_s is denoted by P_s. Then

$$\begin{aligned}
\|(P_s^n - S)f\|_1 &= \int_D \left| \int_D f(y) \left(\sum_{i=1}^n s^{n-i}(1-s)^i \binom{n}{i} k^i(x,y)\,\mathrm{d}y \right) + s^n f(x) - S(f) \right| \varrho(x)\,\mathrm{d}x \\
&\le \int_D \int_D |f(y)| \left| \sum_{i=1}^n s^{n-i}(1-s)^i \binom{n}{i} \frac{k^i(x,y)}{\varrho(y)} - (1-s^n) \right| \varrho(y)\,\mathrm{d}y\, \varrho(x)\,\mathrm{d}x \\
&\quad + s^n \int_D |f(x) - S(f)| \varrho(x)\,\mathrm{d}x \\
&\le \|f\|_1 \int_D \operatorname*{ess\,sup}_{y\in D} \left| \sum_{i=1}^n s^{n-i}(1-s)^i \binom{n}{i} \frac{k^i(x,y)}{\varrho(y)} - (1-s^n) \right| \varrho(x)\,\mathrm{d}x \\
&\quad + s^n \|f - S(f)\|_1 \\
&\le \|f\|_1 \left(2s^n + \int_D \operatorname*{ess\,sup}_{y\in D} \left| \sum_{i=1}^n s^{n-i}(1-s)^i \binom{n}{i} \frac{k^i(x,y)}{\varrho(y)} - (1-s^n) \right| \varrho(x)\,\mathrm{d}x \right),
\end{aligned}$$

which proves the assertion. ∎

For $n = 1$ and $s = 0$ one has a criterion for L_1-exponential convergence with $(\alpha, 1)$ for the transition kernel K.

COROLLARY 3.47. *If there exists an $\alpha \in [0,1)$ such that*

$$\int_D \operatorname*{ess\,sup}_{y\in D} \left| \frac{k(x,y)}{\varrho(y)} - 1 \right| \varrho(x)\,\mathrm{d}x \le \alpha,$$

then the transition kernel K is L_1-exponentially convergent with $(\alpha, 1)$.

EXAMPLE 1. Let us present an easy example borrowed from [Ros95, p. 402]. Let $D = [0,1]$ and $\mathfrak{D} = \mathcal{B}([0,1])$. The transition kernel is defined by

$$K(x,A) = \int_A \frac{1+x+y}{x+3/2}\,\mathrm{d}y, \quad x \in [0,1],\ A \in \mathcal{B}([0,1]).$$

The stationary distribution is given by

$$\pi(A) = \frac{1}{2} \int_A (x+3/2)\,\mathrm{d}x, \quad A \in \mathcal{B}([0,1]).$$

The transition kernel K is reversible with respect to π. These properties can be checked straightforwardly. We have

$$\int_0^1 \operatorname*{ess\,sup}_{y\in[0,1]} \left|\frac{k(x,y)}{\varrho(y)} - 1\right| \varrho(x)\,\mathrm{d}x = \int_0^1 \operatorname*{ess\,sup}_{y\in[0,1]} \frac{|x+y-2xy-1/2|}{2(y+3/2)(x+3/2)}\,\varrho(x)\,\mathrm{d}x$$
$$= \int_0^1 \operatorname*{ess\,sup}_{y\in[0,1]} \frac{|x+y-2xy-1/2|}{4(y+3/2)}\,\mathrm{d}x = \frac{1}{6}\int_0^1 |x-1/2|\,\mathrm{d}x = \frac{1}{24}.$$

Hence Corollary 3.47 gives that the transition kernel is L_1-exponentially convergent with $(1/24, 1)$. Because of the reversibility one can apply Proposition 3.24 to find that the transition kernel is π-a.e. uniformly ergodic with $(1/24, 1/2)$. Furthermore there exists an L_2-spectral gap: one has $\beta \le \alpha = 1/24$.

Let $\delta \in (0, 2/3)$ and let the initial distribution ν be given by

$$\nu(A) = \frac{1}{\delta}\int_A \mathbf{1}_{[0,\delta]}(x)\,\mathrm{d}x, \quad A \in \mathcal{B}([0,1]).$$

Hence the initial state is uniformly distributed in $[0, \delta]$. Then

$$\left\|\frac{d\nu}{d\pi_\varrho} - 1\right\|_\infty = \operatorname*{ess\,sup}_{x\in[0,1]} \left|\frac{4\cdot\mathbf{1}_{[0,\delta]}(x)}{\delta(2x+3)} - 1\right| = \frac{4}{3\delta} - 1.$$

Theorem 3.45 (i) suggests the choice

$$n_0 = \left\lceil \frac{\log(\frac{4}{3\delta} - 1)}{\log(24)} \right\rceil$$

such that

$$\sup_{\|f\|_2\le 1} e_\nu(S_{n,n_0}, f)^2 \le \frac{48}{23n} + \frac{1152}{529n^2} < \frac{5}{n}.$$

Example 2. It is taken from [Ros03, p. 172]. Let $D = [-1, 1]$ and $\mathfrak{D} = \mathcal{B}([-1, 1])$. The transition kernel is defined by

$$K(x, A) = \int_A (\mathbf{1}_{[-1,0]}(x)\mathbf{1}_{(0,1]}(y) + \mathbf{1}_{(0,1]}(x)\mathbf{1}_{[-1,0]}(y))\,\mathrm{d}y, \quad x \in [-1, 1],\ A \in \mathcal{B}([-1, 1]).$$

For $x \in [-1, 0]$ the next state is uniformly distributed in $(0, 1]$ and for $x \in (0, 1]$ the next state is uniformly distributed in $[-1, 0]$. The transition kernel is reversible with respect to the uniform distribution on D, thus π_ϱ is given by $\varrho(x) = 1/2$ for $x \in D$. For $n \in \mathbb{N}$ we have

$$K^n(x, A) = \begin{cases} K(x, A), & n \text{ odd}, \\ K^2(x, A), & n \text{ even}, \end{cases}$$

where

$$K^2(x, A) = \int_A (\mathbf{1}_{[-1,0]}(x)\mathbf{1}_{[-1,0]}(y) + \mathbf{1}_{(0,1]}(x)\mathbf{1}_{(0,1]}(y))\,\mathrm{d}y, \quad x \in [-1, 1],\ A \in \mathcal{B}([-1, 1]).$$

The spectrum of P is completely known: $\mathbf{spec}(P|L_2) = \{1, 0, -1\}$ with

$$\mathrm{Eig}(P, 1) = \{f \in L_2 \mid f \equiv c,\ c \in \mathbb{R}\} = (L_2^0)^\perp,$$
$$\mathrm{Eig}(P, 0) = \left\{f \in L_2 \;\middle|\; \int_{-1}^0 f(x)\mathrm{d}x = \int_0^1 f(x)\mathrm{d}x = 0\right\},$$
$$\mathrm{Eig}(P, -1) = \{f \in L_2 \mid f(x) = c\,(\mathbf{1}_{[-1,0]}(x) - \mathbf{1}_{(0,1]}(x)),\ c \in \mathbb{R}\},$$

where $\mathrm{Eig}(P,\lambda)$ denotes the eigenspace of the eigenvalue λ. Clearly $\mathbf{spec}(P|L_2^0) = \{0,-1\}$. To apply the error bounds one has to pass over to $\widetilde{K}$, the lazy version of K. Let $\widetilde{P}$ be the transition operator which corresponds to $\widetilde{K}$. We write $\beta = \beta_{\widetilde{K}}$ and $\Lambda = \Lambda_{\widetilde{K}}$ to indicate the transition kernel $\widetilde{K}$. We have $\mathbf{spec}(\widetilde{P}|L_2) = \{1, 1/2, 0\}$ and $\mathbf{spec}(\widetilde{P}|L_2^0) = \{1/2, 0\}$. The operator $\widetilde{P}$ has an L_2-spectral gap, one obtains

$$\beta_{\widetilde{K}} = \Lambda_{\widetilde{K}} = \|\widetilde{P}\|_{L_2^0 \to L_2^0} = 1/2.$$

Note that $\widetilde{K} = K_{1/2}$. By the special structure of K^n one finds for $x, y \in D$ that

$$\frac{1}{2^n}\sum_{i=1}^{n}\binom{n}{i}\frac{k^i(x,y)}{\varrho(y)} = \frac{1}{2^{n-1}}\begin{cases}\displaystyle\sum_{i=0}^{(n-1)/2}\binom{n}{2i+1}k(x,y) + \sum_{i=1}^{(n-1)/2}\binom{n}{2i}k^2(x,y), & n \text{ odd},\\ \displaystyle\sum_{i=0}^{n/2-1}\binom{n}{2i+1}k(x,y) + \sum_{i=1}^{n/2-1}\binom{n}{2i}k^2(x,y), & n \text{ even},\end{cases}$$

$$= (k(x,y) + k^2(x,y)) - \frac{k^2(x,y)}{2^{n-1}} = 1 - \frac{k^2(x,y)}{2^{n-1}}.$$

It follows that

$$\int_{-1}^{1} \operatorname*{ess\,sup}_{y\in[-1,1]} \left|\frac{1}{2^n}\sum_{i=1}^{n}\binom{n}{i}\frac{k^i(x,y)}{\varrho(y)} - 1 + \frac{1}{2^n}\right| \varrho(x)\,\mathrm{d}x$$

$$= \int_{-1}^{1} \operatorname*{ess\,sup}_{y\in[-1,1]} \left|\frac{1}{2^n} - \frac{k^2(x,y)}{2^{n-1}}\right| \frac{1}{2}\,\mathrm{d}x = \frac{1}{2^n}.$$

By Lemma 3.46 with $s = 1/2$ we see that the kernel $\widetilde{K}$ is L_1-exponentially convergent with $(1/2, 3)$, i.e.

$$\|\widetilde{P}^n - S\|_{L_1\to L_1} \le \frac{3}{2^n}, \quad n \in \mathbb{N}.$$

The parameter $\alpha = 1/2$ of L_1-exponential convergence is optimal, since $\beta_{\widetilde{K}} = 1/2$ and in general for reversible, L_1-exponentially convergent transition kernel with (α, M) one has $\beta \le \alpha$.

Let $\delta \in (0,1)$. Assume that the initial distribution is given by

$$\nu(A) = \frac{1}{\delta}\int_A \mathbf{1}_{[0,\delta]}(x)\,\mathrm{d}x, \quad A \in \mathcal{B}([-1,1]),$$

i.e. the initial state is chosen with respect to the uniform distribution in $[0,\delta]$. Then

$$\left\|\frac{d\nu}{d\pi_\varrho} - 1\right\|_\infty = \operatorname*{ess\,sup}_{x\in[-1,1]}\left|\frac{2\cdot\mathbf{1}_{[0,\delta]}(x)}{\delta} - 1\right| = \frac{2}{\delta} - 1.$$

Theorem 3.45(i) suggests the choice

$$n_0 = \left\lceil \frac{\log(3(2/\delta - 1))}{\log(2)} \right\rceil$$

such that for S_{n,n_0}, which uses a Markov chain with transition kernel $\widetilde{K}$ and initial distribution ν, one has

$$\sup_{\|f\|_2\le 1} e_\nu(S_{n,n_0}, f) \le \sqrt{\frac{4}{n} + \frac{8}{n^2}}. \tag{3.29}$$

By Remark 3.38, by the L_1-exponential convergence of $\widetilde{K}$ with $(1/2, 3)$ and $\Lambda_{\widetilde{K}} = \beta_{\widetilde{K}}$ one obtains the lower bound

$$\sqrt{\frac{3}{n} - \frac{16}{n^2}} \leq \sup_{\|f\|_2 \leq 1} e_\nu(S_{n,n_0}, f). \tag{3.30}$$

By Corollary 3.37, for all $u \in \mathrm{Eig}(\widetilde{P}, \frac{1}{2}) = \mathrm{Eig}(P, 0)$ with $\|u\|_2 = 1$,

$$e_\pi(S_n, u)^2 = \sup_{\|f\|_2 \leq 1} e_\pi(S_n, f)^2 = \lim_{n_0 \to \infty} \sup_{\|f\|_2 \leq 1} e_\nu(S_{n,n_0}, f)^2.$$

This motivates the comparison of the lower error bound, the upper error bound and the exact error for a specific $u \in \mathrm{Eig}(\widetilde{P}, 1/2)$. Namely, let

$$u(x) = \begin{cases} -1, & x \in [-1, -1/2] \cup [0, 1/2), \\ 1, & x \in (-1/2, 0] \cup (1/2, 1]. \end{cases}$$

Since $u^2 = 1$ we get

$$L_j(u^2) = 0 \quad \text{and} \quad L_j(uP^k u) = \frac{1}{2^k} L_j(u^2) = 0, \quad \text{for } j, k \in \mathbb{N}.$$

Hence by Proposition 3.29 one has

$$e_\nu(S_{n,n_0}, u) = e_\pi(S_n, u) = \sqrt{\frac{3}{n} - \frac{4(1 - 2^{-n})}{n^2}}. \tag{3.31}$$

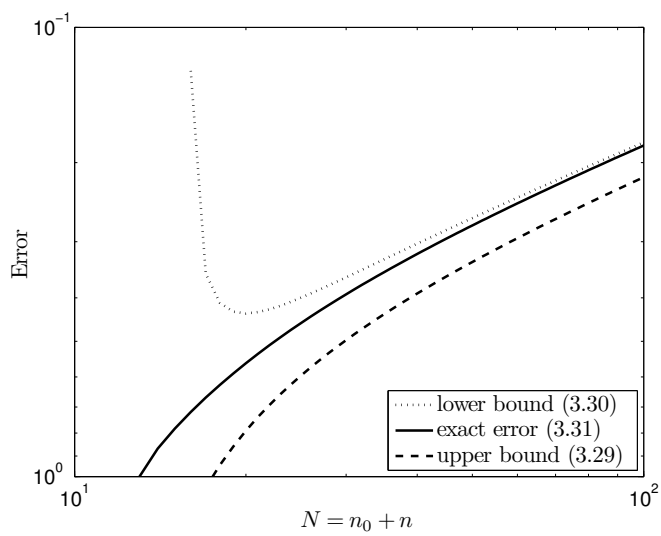

Fig. 4. Example 2: Exact error and error bounds, $\delta = 10^{-3}$ and $n_0 = \lceil \frac{\log(3(2/\delta - 1))}{\log(2)} \rceil = 13$.

In Figure 4 for $\delta = 10^{-3}$ the exact error (3.31), the upper error bound (3.29) and the lower bound (3.30) are plotted. The lower bound leads to a non-trivial estimate if $N \geq n_0 + 6 = 19$. The curve of the upper error estimate is shifted down, because the coefficient of the leading term is worse than the coefficient of the leading term of the exact error $e_\nu(S_{n,n_0}, u)$.

Lemma 3.46 provides a tool which can be used to show L_1-exponential convergence for several examples. Unfortunately it is rather difficult to apply for more sophisticated applications. Next let us present the Metropolis–Hastings algorithm.

Metropolis–Hastings algorithm. The Metropolis–Hastings algorithm, suggested in [MR$^+$53] and extended in [Has70], is widely used in applications. The following introduction is based on Mengersen and Tweedie [MT96]. Suppose that the state space D is contained in $\mathbb{R}^d$ and equipped with $\mathcal{B}(D)$. Let π_ϱ be a probability measure on $(D, \mathcal{B}(D))$ given by a possibly unnormalized density ϱ with respect to the Lebesgue measure:

$$\pi_\varrho(A) = \frac{\int_A \varrho(x)\,\mathrm{d}x}{\int_D \varrho(x)\,\mathrm{d}x}, \quad A \in \mathcal{B}(D).$$

Let $q\colon D \times D \to [0,\infty]$ be such that $q(x,\cdot)$ is integrable with respect to the Lebesgue measure for all $x \in D$ and assume that

$$Q(x,A) = \int_A q(x,y)\,\mathrm{d}y + \mathbf{1}_A(x)\biggl(1 - \int_D q(x,y)\,\mathrm{d}y\biggr), \quad x \in D,\ A \in \mathcal{B}(D),$$

is a transition kernel. It might happen that for some $x \in D$ one has $Q(x,\{x\}) > 0$. If $Q(x,\{x\}) = 0$ for all $x \in D$ then q is a transition density of Q. The question is how to modify Q to get a transition kernel with stationary distribution π_ϱ. For $x, y \in D$ let

$$\theta(x,y) = \begin{cases} \min\biggl\{\dfrac{\varrho(y)q(y,x)}{\varrho(x)q(x,y)}, 1\biggr\}, & \varrho(x)q(x,y) > 0, \\ 1, & \varrho(x)q(x,y) = 0, \end{cases}$$

be the *acceptance probability*. Then the *Metropolis–Hastings transition kernel* K_ϱ is defined by

$$\begin{aligned} K_\varrho(x,A) &= \int_A \theta(x,y)\,Q(x,\mathrm{d}y) + \mathbf{1}_A(x)\biggl(\int_D (1-\theta(x,y))\,Q(x,\mathrm{d}y)\biggr) \\ &= \int_A \theta(x,y)q(x,y)\,\mathrm{d}y + \mathbf{1}_A(x)\biggl(\int_D (1-\theta(x,y))q(x,y)\,\mathrm{d}y + Q(x,\{x\})\biggr), \end{aligned}$$

where $x \in D$ and $A \in \mathcal{B}(D)$. In this setting Q is called the *proposal transition kernel* of K_ϱ. If $q(x,y) = q(y,x)$ for all $x, y \in D$, then we call K_ϱ the *Metropolis transition kernel*. By the construction one can see that the transition kernel K_ϱ is reversible with respect to π_ϱ, thus one has the desired stationary distribution.

LEMMA 3.48. *The Metropolis–Hastings transition kernel K_ϱ is reversible with respect to π_ϱ.*

Proof. It is enough to show that

$$\int_A K_\varrho(x,B)\,\pi_\varrho(\mathrm{d}x) = \int_B K_\varrho(x,A)\,\pi_\varrho(\mathrm{d}x)$$

for disjoint $A, B \in \mathcal{B}(D)$. Then the assertion follows by the symmetry $\theta(x,y)q(x,y)\varrho(x) = \theta(y,x)q(y,x)\varrho(y)$ and Fubini's Theorem. ∎

The Metropolis–Hastings algorithm, which simulates a transition of the Metropolis-Hastings transition kernel, goes as follows: Let $x \in D$ be the current state. Choose a proposal state y with respect to $Q(x,\cdot)$. Toss a coin whose "head" probability is $\theta(x,y)$.

If "head" then accept the proposal state, i.e. return y. Otherwise reject the proposal, i.e. return x. Schematically, a single step of the Metropolis–Hastings algorithm is presented in the Procedure Metropolis-Step(x, Q, ϱ).

Procedure Metropolis-Step(x,Q,ϱ)

input: current state x, proposal kernel Q, unnormalized density ϱ.
output: next state y.

Choose y with respect to $Q(x,\cdot)$;

Compute

$$\theta(x,y) = \begin{cases} \min\{\frac{\varrho(y)q(y,x)}{\varrho(x)q(x,y)}, 1\}, & \varrho(x)q(x,y) > 0, \\ 1, & \varrho(x)q(x,y) = 0; \end{cases}$$

if rand() $\geq \theta(x,y)$ **then**
  $y := x$;
end
Return y.

If $\tilde{q}(y) = q(x,y)$ for all $x, y \in D$ then the proposal transition kernel samples independently of x. In this situation one can apply the following result.

THEOREM 3.49 (see [MT96, Theorem 2.1, p. 105]). *Let $\tilde{q}\colon D \to [0,\infty]$ be a function with $\int_D \tilde{q}(x)\,\mathrm{d}x = 1$. Let the proposal transition kernel of the Metropolis–Hastings transition kernel K_ϱ be $Q(x,A) = \int_A \tilde{q}(y)\,\mathrm{d}y$ for $x \in D$ and $A \in \mathcal{B}(D)$. If there exists a $\gamma > 0$ such that*

$$\frac{\tilde{q}(y)}{\varrho(y)} \geq \gamma, \quad y \in D,$$

then K_ϱ is uniformly ergodic, and

$$\|K_\varrho^n(x,\cdot) - \pi\|_{\mathrm{tv}} \leq (1-\gamma)^n, \qquad x \in D,\ n \in \mathbb{N}.$$

REMARK 3.50. The proof is based on the well known equivalence that a transition kernel K is uniformly ergodic iff the whole state space D is a small set. A set $R \in \mathcal{B}(D)$ is called *small* if there exists a $\gamma > 0$, an $m \in \mathbb{N}$ and a probability measure ψ such that

$$K^m(x,A) \geq \gamma\psi(A), \quad x \in R,\ A \in \mathcal{B}(D).$$

The result of Theorem 3.49 will be demonstrated for a toy example, stated in [MT96, p. 107].

Let $D = \mathbb{R}$ and $\mathfrak{D} = \mathcal{B}(\mathbb{R})$. Note that the state space is unbounded. The desired distribution is given by the density

$$\varrho(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right), \quad y \in \mathbb{R},$$

i.e. π_ϱ is an $N(0,1)$ distribution. By $N(\mu,\xi^2)$ we denote the normal distribution with mean μ and variance ξ^2. Furthermore, assume that the proposal transition kernel samples

independently from $N(0,\xi^2)$ so that

$$\tilde{q}(y) = \frac{1}{\sqrt{2\pi}\xi} \exp\left(-\frac{y^2}{2\xi^2}\right), \quad y \in \mathbb{R}.$$

Let $\xi^2 > 1$. Then

$$\frac{\tilde{q}(x)}{\varrho(x)} \geq \xi^{-1},$$

which implies that

$$\|K_\varrho^n(x,\cdot) - \pi_\varrho\|_{\mathrm{tv}} \leq (1-\xi^{-1})^n, \quad x \in D,\ n \in \mathbb{N}.$$

By the reversibility with respect to π_ϱ of the Metropolis–Hastings transition kernel, an immediate consequence is that uniform ergodicity implies L_1-exponential convergence, since π-a.e. uniform ergodicity is equivalent to L_1-exponential convergence. Hence we have a transition kernel which is L_1-exponentially convergent with $(1-\xi^{-1},1)$. This implies that the Markov operator P which corresponds to the transition kernel K_ϱ has an L_2-spectral gap: we have $1-\beta \geq \xi^{-1}$.

Let $\delta \in (0,1)$ and $x_0 \in [0,\infty)$. The initial state is chosen uniformly distributed in $[x_0-\delta, x_0+\delta]$. Then

$$\frac{d\nu}{d\pi}(x) = \sqrt{\frac{\pi}{2}} \cdot \frac{\mathbf{1}_{[x_0-\delta,x_0+\delta]}(x)}{\delta} \exp\left(\frac{x^2}{2}\right), \quad x \in D.$$

We obtain

$$\left\| \frac{d\nu}{d\pi} - 1 \right\|_\infty = \sqrt{\frac{\pi}{2}} \cdot \frac{\exp((x_0+\delta)^2/2)}{\delta} - 1 \leq \sqrt{\frac{\pi}{2}} \cdot \frac{\exp((x_0+\delta)^2/2)}{\delta}.$$

The method S_{n,n_0} uses a Markov chain with transition kernel K_ϱ and initial distribution ν. The burn-in is almost chosen as suggested in Theorem 3.45(i). We use $\log(1-\xi^{-1}) \geq \xi^{-1}$ to estimate the burn-in, so we set

$$n_0 = \lceil \xi(\log(\delta^{-1}) + (x_0+\delta)^2/2 + 0.23) \rceil.$$

Then

$$\sup_{\|f\|_2 \leq 1} e_\nu(S_{n,n_0}, f)^2 \leq \frac{2\xi}{n} + \frac{2\xi^2}{n^2}.$$

Contracting normals. The next example is described in [Bax05]; see also [RR97b, RT99]. Let $D = \mathbb{R}$, $\mathfrak{D} = \mathcal{B}(\mathbb{R})$ and $\theta \in (-1,1)$. Note that the state space is unbounded. The transition kernel is given by

$$K(x,A) = \frac{1}{\sqrt{2\pi(1-\theta^2)}} \int_A \exp\left(-\frac{(\theta x - y)^2}{2(1-\theta^2)}\right) \mathrm{d}y, \quad x \in \mathbb{R},\ A \in \mathcal{B}(\mathbb{R}),$$

so that $K(x,\cdot)$ is an $N(\theta x, 1-\theta^2)$ distribution. By some elementary calculation one can see that a stationary distribution is

$$\pi(A) = \frac{1}{\sqrt{2\pi}} \int_A \exp\left(-\frac{y^2}{2}\right) \mathrm{d}y, \quad A \in \mathcal{B}(\mathbb{R}),$$

i.e. π is an $N(0,1)$ distribution. The transition kernel K is reversible with respect to π. Suppose that $\theta \in (0,1)$. Then the Markov operator is *positive semi-definite*, i.e. $\langle Pf, f\rangle \geq 0$ for all $f \in L_2$. The next result is an application of [Bax05, Theorem 1.3,

p. 702] where the Markov operator is self-adjoint and positive semi-definite. The same example is considered in [Bax05, p. 728] and [ŁN11, p. 33].

LEMMA 3.51 (see [Bax05, Theorem 1.3, pp. 702 and 728]). *Let $\theta \in (0,1)$, $c \in (1,\infty)$ and set*

$$\lambda = \theta^2 + \frac{2(1-\theta^2)}{1+c^2},$$
$$K = 2 + \theta^2(c^2-1),$$
$$B = 2\left[\Phi\left(\frac{(1+\theta)c}{\sqrt{1-\theta^2}}\right) - \Phi\left(\frac{\theta c}{\sqrt{1-\theta^2}}\right)\right], \quad \textit{where} \quad \Phi(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} \exp\left(-\frac{y^2}{2}\right) \mathrm{d}y,$$
$$\alpha = 1 + \frac{\log\left(\frac{K-B}{1-B}\right)}{\log(\lambda^{-1})},$$
$$\hat{\beta} = \max\{\lambda, (1-B)^{1/\alpha}\} < 1.$$

Then

$$\beta = \|P\|_{L_2^0 \to L_2^0} \le \hat{\beta}.$$

Let us illustrate the last lemma. For any fixed θ one can numerically minimize the upper estimate $\hat{\beta}$ of β, depending on c. For example let $\theta = 0.5$. Then one gets $\hat{\beta} = 0.8946$ for $c = 1.6041$.

There exists an L_2-spectral gap, so we can apply Theorem 3.45 for $p \in (2,\infty]$. Let $\delta \in (0,1)$ and $x_0 \in [0,\infty)$. The initial state is chosen uniformly distributed on $[x_0-\delta, x_0+\delta]$. The density of the initial distribution with respect to π is given by

$$\frac{d\nu}{d\pi}(x) = \sqrt{\frac{\pi}{2}} \cdot \frac{\mathbf{1}_{[x_0-\delta,x_0+\delta]}(x)}{\delta} \exp\left(\frac{x^2}{2}\right).$$

Then for all $q \in [1,\infty]$ it follows that

$$\left\|\frac{d\nu}{d\pi} - 1\right\|_q \le \left\|\frac{d\nu}{d\pi} - 1\right\|_\infty = \sqrt{\frac{\pi}{2}} \cdot \frac{\exp((x_0+\delta)^2/2)}{\delta} - 1 \le \sqrt{\frac{\pi}{2}} \cdot \frac{\exp((x_0+\delta)^2/2)}{\delta}.$$

The burn-in is chosen as suggested in Theorem 3.45, where we use the previously stated estimate of $\|d\nu/d\pi - 1\|_q$. Suppose that the burn-in $n_0(p)$ is the smallest natural number (including zero) which is greater than or equal to

$$\frac{1}{\log(\hat{\beta}^{-1})} \begin{cases} \dfrac{p}{2(p-2)}\left[\log\left(\dfrac{16p}{p-2}\right) + \log(\sqrt{2\pi}\,\delta^{-1}) + (x_0+\delta)^2/2\right], & p \in (2,4), \\ \log(\delta^{-1}) + (x_0+\delta)^2/2 + 4.39, & p \in [4,\infty]. \end{cases}$$

Then

$$\sup_{\|f\|_p \le 1} e_\nu(S_{n,n_0}, f)^2 \le \frac{2}{n(1-\hat{\beta})} + \frac{2}{n^2(1-\hat{\beta})^2}.$$

In Table 2 one can see how much resources N are sufficient to obtain an error less than $\varepsilon = 0.01$.

3.5. Notes and remarks. In the last decades explicit error bounds and confidence estimates of Markov chain Monte Carlo methods on general state spaces attracted more and more attention. Let us present how our results fit into the published literature.

Table 2. Contracting Normals: The initial distribution ν is chosen with $x_0 = 0$ and $\delta = 0.1$. The burn-in of Theorem 3.45 is computed for $p = 2.1$ and n is such that one obtains an error less than $\varepsilon = 0.01$. The estimate $\hat{\beta}$ of β is computed by a minimizing procedure of Maple for $c \geq 1.01$.

θ	c	$\hat{\beta}$	n_0 (for $p = 2.1$)	n (for precision $\varepsilon = 0.01$)	N
0.91	1.12845	0.999664	$2.82241 \cdot 10^5$	$5.94614 \cdot 10^7$	$5.97437 \cdot 10^7$
0.92	1.11691	0.999816	$5.16275 \cdot 10^5$	$1.08759 \cdot 10^8$	$1.09275 \cdot 10^8$
0.93	1.10499	0.999912	$1.08257 \cdot 10^6$	$2.28043 \cdot 10^8$	$2.29126 \cdot 10^8$
0.94	1.09260	0.999966	$2.76738 \cdot 10^6$	$5.82923 \cdot 10^8$	$5.85690 \cdot 10^8$
0.95	1.07964	0.999990	$9.60536 \cdot 10^6$	$2.02337 \cdot 10^9$	$2.03297 \cdot 10^9$
0.96	1.06599	0.999998	$5.58578 \cdot 10^7$	$1.17624 \cdot 10^{10}$	$1.18183 \cdot 10^{10}$

In the seminal work of Lovász and Simonovits [LS93] an estimate of $e_\pi(S_n, f)^2$ is shown. The paper deals with the computation of the volume of a convex body by a randomized algorithm based on Markov chains. Let us explain the result of [LS93, Theorem 1.9, p. 375] in detail. Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with transition kernel K and initial distribution ν and let K be reversible with respect to a probability measure π. Then let us define the *conductance* as

$$\varphi(K, \pi) = \inf_{0<\pi(A)\leq 1/2} \frac{\int_A K(x, A^c)\, \pi(\mathrm{d}x)}{\pi(A)}.$$

Assume that the Markov operator is positive semi-definite, i.e. $\langle Pf, f\rangle \geq 0$ for all $f \in L_2$. Then

$$e_\pi(S_n, f)^2 \leq \frac{4}{\varphi(K, \pi)^2 \cdot n} \|f\|_2^2. \tag{3.32}$$

This result is slightly worse than Proposition 3.26. In Proposition 3.26 one has an exact error formula for $e_\pi(S_n, f)^2$. Mainly the spectral structure of the Markov operator is used. In Corollary 3.27 this exact error formula is further estimated and one obtains

$$e_\pi(S_n, f)^2 \leq \frac{2}{(1-\Lambda)n} \|f\|_2^2, \quad \text{where} \quad \Lambda = \sup\{\alpha \mid \alpha \in \mathbf{spec}(P|L_2^0)\}. \tag{3.33}$$

The *Cheeger inequality* $1 - \Lambda \geq \varphi(K, \pi)^2/2$ (see Section A.3) provides a relation between Λ and $\varphi(K, \pi)$, so that (3.33) implies (3.32). Note that in Proposition 3.26 and Corollary 3.27 it is not assumed that the Markov operator is positive semi-definite, so the assumptions are slightly less restrictive. But if one has a transition kernel K which determines a not necessarily positive semi-definite transition operator, then one can pass over to the lazy version of K to obtain positive semi-definiteness. However, the estimate of (3.32) covers the important facts and it seems that the refinement of Proposition 3.26 is well known.

The paper of Mathé [Mat99] contains results concerning the asymptotic integration error for uniformly ergodic Markov chains which are reversible with respect to π. For example it is shown that for any initial distribution $\nu \in \mathcal{M}_\infty$ one has

$$\lim_{n\to\infty} n \sup_{\|f\|_2\leq 1} e_\nu(S_{n,n_0}, f)^2 = \frac{1+\Lambda}{1-\Lambda}$$

and for $f \in L_2$ it is proven that

$$\lim_{n\to\infty} n e_\nu(S_{n,n_0}, f)^2 = \langle (I-P)^{-1}(I+P)g, g\rangle, \quad \text{where} \quad g = f - S(f).$$

The same result is part of Corollary 3.37 and for individual f part of Theorem 3.34. In [Mat04] the asymptotic integration error is studied for Markov chains not necessarily reversible and not necessarily uniformly ergodic. It is assumed that the transition kernel is V-uniformly ergodic (see (3.36)). For further details we refer to [Mat04].

In [Rud09, Theorem 8, p. 19] an explicit upper error bound of $e_\nu(S_{n,n_0}, f)^2$ for general state spaces is provided. The result is based on [LS93, Theorem 1.9, p. 375] and the assumptions are the same. Namely, the transition kernel K is reversible with respect to π and the transition operator P is positive semi-definite. After a burn-in

$$n_0 \geq \frac{\log(\|\frac{d\nu}{d\pi}\|_\infty)}{\varphi(K,\pi)^2} \quad \text{the error obeys} \quad e_\nu(S_{n,n_0}, f)^2 \leq \frac{100}{\varphi(K,\pi)^2 \cdot n}\|f\|_\infty^2. \tag{3.34}$$

The proof is based on Proposition 3.29 which provides the crucial relation between $e_\nu(S_{n,n_0}, f)^2$ and $e_\pi(S_{n,n_0}, f)^2$. By Theorems 3.41 and 3.45 one obtains a refined error estimate and a refined recipe for the choice of n_0. Note that positive semi-definiteness and reversibility are not needed in Theorem 3.41. It is enough that there exists an L_2-spectral gap, i.e. $1 - \beta > 0$.

Independently of [Rud09, Theorem 8, p. 19] in the work of Belloni and Chernozhukov [BC09, Theorem 3, p. 2031] a similar error bound for S_{n,n_0} is proven. It is also based on [LS93, Theorem 1.9, p. 375], so that again the transition kernel is assumed to be reversible with respect to π and the Markov operator must be positive semi-definite. Then it is shown that

$$e_\nu(S_{n,n_0}, f)^2 \leq e_\pi(S_n, f)^2 + 8\|f\|_\infty^2 \|\nu P^{n_0} - \pi\|_{\mathrm{tv}}.$$

Let the initial distribution ν be R-warm, i.e. $\sup_{A\in\mathfrak{D},\,\pi(A)>0} \nu(A)/\pi(A) \leq R$. Then by [LS93, Corollary 1.5, p. 372] one obtains

$$\|\nu P^{n_0} - \pi\|_{\mathrm{tv}} \leq \sqrt{R}\left(1 - \frac{\varphi(K,\pi)^2}{2}\right)^{n_0}.$$

Hence by [LS93, Theorem 1.9, p. 375] one has

$$e_\nu(S_{n,n_0}, f)^2 \leq \frac{4}{\varphi(K,\pi)^2 \cdot n}\|f\|_2^2 + 8\sqrt{R}\left(1 - \frac{\varphi(K,\pi)^2}{2}\right)^{n_0}\|f\|_\infty^2. \tag{3.35}$$

The explicit error bound for S_{n,n_0}, when the initial distribution is not the stationary one, is the same as in [Rud09, Theorem 8, p. 19]. Note that the burn-in depends on the desired precision. We can choose $R = \|d\nu/d\pi\|_\infty$ and if one uses $\|f\|_2 \leq \|f\|_\infty$, then the upper bound of (3.35) can be further estimated and one obtains an estimate with respect to $\|\cdot\|_\infty$.

Another result is due to Łatuszyński and Niemiro [ŁN11]. The integration error for V-uniformly ergodic Markov chains is estimated, where $V\colon D \to [1,\infty)$ is a drift function. The weighted class of functions

$$L_V = L_V(D) = \left\{ f\colon D \to \mathbb{R} \;\middle|\; |f|_V = \sup_{x\in D} \frac{|f(x)|}{V(x)} < \infty \right\}$$

is studied. Let $\alpha \in [0,1)$ and $M < \infty$. A transition kernel K is called *V-uniformly ergodic with (α, M)* if

$$\|P^n - S\|_{L_V \to L_V} \le M\alpha^n, \quad n \in \mathbb{N}. \tag{3.36}$$

One can replace the drift function V by $V^{1/r}$ for all $r \ge 1$. Then there exist an $\alpha(r) \in [0,1)$ and an $M(r) < \infty$ such that

$$\|P^n - S\|_{L_{V^{1/r}} \to L_{V^{1/r}}} \le M(r)\alpha(r)^n, \quad n \in \mathbb{N}$$

(see for example [Mat04] or [Bax05]). Now let us state a less general version of the main result of [ŁN11, Theorem 3.1, p. 28]. For $r = 2$ and $g = f - S(f)$ one has

$$e_\nu(S_{n,n_0}, f)^2 \le \frac{|g^2|_V}{n}\left(1 + \frac{2M(2)\alpha(2)}{1-\alpha(2)}\right)\left(\|V\|_1 + \frac{M^2\alpha^{n_0}\|\nu - \pi\|_V}{n(1-\alpha)}\right), \tag{3.37}$$

where $\|\nu - \pi\|_V = \sup_{|g|_V \le 1} |\int_D g(x)(\nu(\mathrm{d}x) - \pi(\mathrm{d}x))|$. This seems to be the first explicit error bound of S_{n,n_0} for integrands f which belong to L_V. If the transition kernel is reversible, then V-uniform ergodicity with (α, M) is equivalent to the existence of an L_2-spectral gap (see [RR97a, RT01]). Furthermore if $V \in L_p$ for some $p > 2$ then $L_V \subset L_p$ and the error bound of Theorem 3.41 can also be applied. However, in general Theorem 3.41 cannot be used in this setting.

The paper of Joulin and Ollivier [JO10] based on [Oll09] follows a new idea. Let $(D, \mathbf{d}\mathrm{ist})$ be a metric, complete, separable state space, with metric $\mathbf{d}\mathrm{ist}$, and let K be a transition kernel with stationary distribution π on $(D, \mathcal{B}(D))$. Let $\mathcal{P}_{\mathbf{d}\mathrm{ist}}(D)$ be the set of probability measures μ on $(D, \mathcal{B}(D))$ for which there exists an $x_0 \in D$ such that

$$\int_D \mathbf{d}\mathrm{ist}(x_0, y)\, \mu(\mathrm{d}y) < \infty.$$

Then we define the *Wasserstein distance* between $\mu_1, \mu_2 \in \mathcal{P}_{\mathbf{d}\mathrm{ist}}(D)$ by

$$W_1(\mu_1, \mu_2) = \inf_{\xi \in \Pi(\mu_1, \mu_2)} \int_D \int_D \mathbf{d}\mathrm{ist}(x, y)\, \xi(\mathrm{d}x, \mathrm{d}y),$$

where $\Pi(\mu_1, \mu_2)$ is the set of probability measures ξ on $(D^2, \mathcal{B}(D^2))$ with marginals μ_1 and μ_2. If there exists a $\kappa > 0$ such that

$$W_1(K(x, \cdot), K(y, \cdot)) \le (1 - \kappa)\, \mathbf{d}\mathrm{ist}(x, y), \quad x, y \in D, \tag{3.38}$$

then we say that the transition kernel K has *positive Ricci curvature* κ. Let the function $f\colon D \to \mathbb{R}$ be integrable with respect to π and let

$$\|f\|_{\mathrm{Lip}} = \sup_{x,y \in D,\, x \ne y} \frac{|f(x) - f(y)|}{\mathbf{d}\mathrm{ist}(x, y)}.$$

The *coarse diffusion constant* $\sigma(x)$ for $x \in D$ of the transition kernel is defined by

$$\sigma(x)^2 = \frac{1}{2}\int_D \int_D \mathbf{d}\mathrm{ist}(y, z)^2\, K(x, \mathrm{d}y)\, K(x, \mathrm{d}z),$$

and the *local dimension* n_x for $x \in D$ is defined by

$$n_x = \inf_{\|f\|_{\mathrm{Lip}} = 1} \frac{2\sigma(x)^2}{\int_D \int_D |f(y) - f(z)|^2\, K(x, \mathrm{d}z)\, K(x, \mathrm{d}y)}.$$

If the transition kernel has positive Ricci curvature, then by [JO10, Proposition 1, p. 2423, and Theorem 2, p. 2424] one obtains

$$e_{\delta_x}(S_{n,n_0},f)^2 \le \left(\frac{1}{\kappa^2 n}+\frac{1}{\kappa^3 n^2}\right)\|f\|_{\mathrm{Lip}}^2 \sup_{x\in D}\frac{\sigma(x)^2}{n_x} + \frac{(1-\kappa)^{2(n_0+1)}}{\kappa^4 n^2}\|f\|_{\mathrm{Lip}}^2\left(\int_D \mathbf{d}\mathrm{ist}(x,y)\,K(x,\mathrm{d}y)\right)^2.$$

The estimate is reasonable for any deterministic initial state $x \in D$, the initial distribution is δ_x. For further estimates and details we refer to [JO10]. Let $p \in (2,\infty]$, $\|f\|_{\mathrm{Lip}} < \infty$ and assume that there exists an $x_0 \in D$ such that $\|\mathbf{d}\mathrm{ist}(\cdot,x_0)\|_p < \infty$. Then $f \in L_p$, in particular

$$\|f\|_p \le 2^{\frac{p-1}{p}}(\|f\|_{\mathrm{Lip}}\|\mathbf{d}\mathrm{ist}(\cdot,x_0)\|_p + |f(x_0)|).$$

If the transition kernel is reversible with respect to π and $\|\sigma\|_2 < \infty$, then one can show that a positive Ricci curvature $\kappa > 0$ of K implies an L_2-spectral gap of P, namely $1-\beta \ge \kappa$ (see [Oll09, Proposition 30, p. 831]). In this setting Theorem 3.41 can be applied when the initial distribution ν belongs to $\mathcal{M}_{\max\{2,p/(p-2)\}}$.

A regenerative Markov chain Monte Carlo algorithm for the approximation of $S(f)$ is studied in [ŁMN09]. Roughly speaking, if one has information on a certain small set, then one can explicitly estimate the mean square error of this regenerative estimator for uniformly and V-uniformly ergodic Markov chains (see [ŁMN09] for details).

The literature also provides confidence estimates for S_{n,n_0}. One can apply Lemma 2.27 if an upper bound of $e_\nu(S_{n,n_0},f)^2$ is available. These estimates can be boosted by a median trick explained in [NP09] and applied in [ŁN11]. However, exponential inequalities such as Hoeffding or Chernoff bounds for Markov chain Monte Carlo are better (see [Krü98, Lez01, GO02, JO10, Mia10]). Asymptotic confidence estimates are discussed in [FJ11].

Let us provide a conclusion. There are different explicit error bounds of the mean square error for S_{n,n_0} on general state spaces. In some situations these estimates could be improved. It seems that the error bound with respect to $\|\cdot\|_2$ is not known so far. Let us recall that we assumed that the Markov chain considered is L_1-exponentially convergent and reversible with respect to π. If we only assume that the Markov chain has an L_2-spectral gap, then we showed an estimate of the error that is uniform with respect to $\|\cdot\|_p$ for $p \in (2,\infty]$. Upper error bounds with respect to $\|\cdot\|_\infty$ are known but with respect to $\|\cdot\|_p$ seem to be new. In this setting it is not assumed that the Markov chain is reversible with respect to π, but we require that π is the stationary distribution. The suggested burn-in n_0 of Theorem 3.45 works well and also appears to be new. All error bounds hold for bounded and unbounded state spaces whenever estimates of the crucial parameters, for example Λ, β or (α, M), are available.

4. Applications

In numerous applications one wants to compute for $D \subset \mathbb{R}^d$ an integral of the form

$$\int_D f(x) \cdot c\varrho(x)\,\mathrm{d}x, \tag{4.1}$$

with density $c\varrho$, where the number c is unknown. Of course c can be defined by

$$\frac{1}{c} = \int_D \varrho(x)\,\mathrm{d}x.$$

However, it is desirable to have algorithms that are able to compute (4.1) without any precomputation of c. Let $\mathcal{F}(D)$ be a class of tuples of the form (f, ϱ), where $\varrho\colon D \to [0,\infty)$ is a possibly unnormalized density with $\int_D \varrho(x)\,\mathrm{d}x > 0$ and $f \cdot \varrho$ is integrable with respect to the Lebesgue measure. Then the goal is to compute

$$S(f,\varrho) = \frac{\int_D f(x)\varrho(x)\,\mathrm{d}x}{\int_D \varrho(x)\,\mathrm{d}x} \quad \text{for } (f,\varrho) \in \mathcal{F}(D). \tag{4.2}$$

The solution operator S is linear in f but not in ϱ. Hence S is a nonlinear functional.

We assume that there are two procedures, Or_f and Or_ϱ, which provide information on f and ϱ, respectively. These procedures are considered as "black boxes" and we call them oracles. Let Or_f be a procedure which returns for an input $x \in D$ the function value $f(x)$, i.e. $\mathrm{Or}_f(x) = f(x)$. Unless stated otherwise we also assume that $\mathrm{Or}_\varrho(x) = \varrho(x)$ for $x \in D$. We assume that an oracle call is much more expensive than arithmetic operations. Hence we count the total number of oracle calls which are needed to approximate $S(f,\varrho)$.

Let $\mathbf{A}\mathrm{lg}_n$ be the class of all randomized algorithms which use at most n calls to the oracle Or_f and n calls to the oracle Or_ϱ. More precisely $A_n \in \mathbf{A}\mathrm{lg}_n$ is a mapping described by a function $\varphi_{2n} : \mathbb{R}^{2n} \to \mathbb{R}$ such that

$$A_n(f,\varrho) = \varphi_{2n}(\mathrm{Or}_f(X_1),\dots,\mathrm{Or}_f(X_n),\mathrm{Or}_\varrho(X_1),\dots,\mathrm{Or}_\varrho(X_n)).$$

The sample $(X_1,\dots,X_n) \in D^n$ is determined as follows: Let $\omega = (\omega_1,\dots,\omega_n)$ be a random element with some distribution W. Then

$$\begin{aligned} X_1 &= X_1(\omega_1),\\ X_i &= X_i(\mathrm{Or}_f(X_1),\dots,\mathrm{Or}_f(X_{i-1}),\mathrm{Or}_\varrho(X_1),\dots,\mathrm{Or}_\varrho(X_{i-1}),\omega_i), \quad i = 2,\dots,n. \end{aligned}$$

The *individual error* of $A_n \in \mathbf{A}\mathrm{lg}_n$ applied to $(f,\varrho) \in \mathcal{F}(D)$ is, as in the previous chapters, measured in the mean square sense:

$$e(A_n,(f,\varrho)) = (\mathbf{E}|S(f,\varrho) - A_n(f,\varrho)|^2)^{1/2},$$

where the expectation is taken with respect to W. The *overall error* on $\mathcal{F}(D)$ is

$$e(A_n,\mathcal{F}(D)) = \sup_{(f,\varrho)\in\mathcal{F}(D)} e(A_n,(f,\varrho)).$$

The *complexity* of the problem (4.2) on $\mathcal{F}(D)$ is given by

$$\mathbf{c}\mathrm{omp}(\varepsilon,d,\mathcal{F}(D)) = \min\{n \mid \text{there exists } A_n \in \mathbf{A}\mathrm{lg}_n \text{ with } e(A_n,\mathcal{F}(D)) \le \varepsilon\}.$$

Note that d is the dimension of the domain D. We want to quantify the complexity of a problem with respect to the dimension d. The integration problem (4.2) for the class

$\mathcal{F}(D)$ is called *polynomially tractable* if there exist non-negative numbers c, q_1 and q_2 such that

$$\mathbf{c}\mathrm{omp}(\varepsilon, d, \mathcal{F}(D)) \le c\,\varepsilon^{-q_1} d^{q_2} \quad \text{for all } d \in \mathbb{N},\ \varepsilon \in (0,1).$$

Roughly speaking, this says that the complexity of computing (4.2) increases at most polynomially in the precision ε^{-1} and the dimension d. For details of the concept of tractability we refer to Novak and Woźniakowski [NW08, NW10].

Let us provide a result which motivates introducing a modified notion of tractability. We consider the class of functions

$$\mathcal{F}_C(D) = \left\{ (f,\varrho) \;\middle|\; \|f\|_\infty \le 1,\ \frac{\sup \varrho}{\inf \varrho} \le C \right\}.$$

In some applications C can be very large, such as $C = 10^{20}$. Observe that always $S(\mathcal{F}_C(D)) = [-1,1]$, hence the problem is scaled properly. In [MN07] Mathé and Novak proved a lower error bound (see [MN07, Theorem 1, p. 678]).

THEOREM 4.1. *For any $A_n \in \mathbf{A}\mathrm{lg}_n$,*

$$e(A_n, \mathcal{F}_C(D)) \ge \frac{\sqrt{2}}{6} \begin{cases} \sqrt{\dfrac{C}{2n}}, & 2n \ge C-1, \\ \dfrac{3C}{C+2n-1}, & 2n < C-1. \end{cases}$$

For an upper error bound Mathé and Novak consider the simple Monte Carlo algorithm: Evaluate the numerator and denominator on a common independent sample according to the uniform distribution, say $(X_1, \dots, X_n) \in D^n$, and compute

$$A_n^{\mathrm{simple}}(f,\varrho) = \frac{\sum_{j=1}^n f(X_j)\varrho(X_j)}{\sum_{j=1}^n \varrho(X_j)}.$$

Note that every X_j is uniformly distributed. It is essential that one can sample according to the uniform distribution on D. This might be a restrictive assumption. In [MN07, Theorem 2, p. 680] the following upper error bound is proven.

THEOREM 4.2. *For all $n \in \mathbb{N}$ we have*

$$e(A_n^{\mathrm{simple}}, \mathcal{F}_C(D)) \le 2\min\{1, \sqrt{2C/n}\}.$$

From Theorems 4.1 and 4.2 one finds that the complexity $\mathbf{c}\mathrm{omp}(\varepsilon, d, \mathcal{F}_C(D))$ of (4.2) is linear in C and A_n^{simple} is almost optimal; moreover, for all $\varepsilon \in (0, 1/(2\sqrt{2}))$,

$$0.02\,C\varepsilon^{-2} \le \mathbf{c}\mathrm{omp}(\varepsilon, d, \mathcal{F}_C(D)) \le 8\,C\varepsilon^{-2}.$$

Hence all algorithms are bad if $C = 10^{20}$. Mathé and Novak suggest considering a smaller class of densities. The main goal is to have also tractability with respect to C on a class of functions, say $\widetilde{\mathcal{F}}_C(D)$, where the possibly unnormalized densities satisfy $\sup\varrho/\inf\varrho \le C$. More precisely, the integration problem (4.2) is called *tractable with respect to C* if there exist non-negative numbers c, q_1, q_2 and q_3 such that

$$\mathbf{c}\mathrm{omp}(\varepsilon, d, \widetilde{\mathcal{F}}_C(D)) \le c\,\varepsilon^{-q_1} d^{q_2} [\log C]^{q_3} \tag{4.3}$$

for all $\varepsilon \in (0,1)$, $d \in \mathbb{N}$ and $C > 1$ (see [NW10, p. 541]).

With Markov chain Monte Carlo algorithms one can achieve this goal on certain classes of functions. Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with transition kernel K and initial distribution ν. Assume that the transition kernel has stationary distribution π_ϱ, where

$$\pi_\varrho(A) = \frac{\int_A \varrho(x)\,\mathrm{d}x}{\int_D \varrho(x)\,\mathrm{d}x}, \quad A \in \mathcal{B}(D), \quad \text{so that} \quad S(f,\varrho) = \int_D f(x)\,\pi_\varrho(\mathrm{d}x).$$

Under suitable assumptions on the Markov chain and on $(f,\varrho) \in \widetilde{\mathcal{F}}_C(D)$, the algorithm

$$S_{n,n_0}(f,\varrho) = \frac{1}{n}\sum_{j=1}^{n} f(X_{j+n_0})$$

is an approximation of $S(f,\varrho)$. Suppose that for each step of the Markov chain we use a single call to Or_ϱ. Then it follows that S_{n,n_0} needs $n+n_0$ calls to Or_ϱ and n calls to Or_f. Consequently, $S_{n,n_0} \in \mathbf{A}\mathrm{lg}_{n+n_0}$.

4.1. Integration with respect to log-concave densities. Let $r > 0$ and let $B(x,r)$ be the d-dimensional Euclidean ball with radius r around $x \in \mathbb{R}^d$. Furthermore let $B^d = B(0,1)$ and $rB^d = B(0,r)$. The goal is to compute

$$S(f,\varrho) = \frac{\int_{rB^d} f(x)\varrho(x)\,\mathrm{d}x}{\int_{rB^d} \varrho(x)\,\mathrm{d}x} \tag{4.4}$$

for (f,ϱ) in a certain class of functions. Let us define the class of functions on a convex body $D \subset \mathbb{R}^d$ rather than on rB^d. We assume that the state space D is equipped with the Borel σ-algebra $\mathcal{B}(D)$. We consider functions (f,ϱ) with the following properties:

- ϱ is strictly positive and log-concave, i.e. for all $x,y \in D$ and $0 < \lambda < 1$,
$$\varrho(\lambda x + (1-\lambda)y) \geq \varrho(x)^\lambda \cdot \varrho(y)^{1-\lambda}.$$
- The logarithm of ϱ is Lipschitz continuous, i.e. there exists an $\mathrm{L} \geq 0$ such that
$$|\log \varrho(x) - \log \varrho(y)| \leq \mathrm{L}\|x-y\|_{\mathrm{E}}, \quad x,y \in D,$$
where $\|\cdot\|_{\mathrm{E}}$ denotes the Euclidean norm.
- The integrand f satisfies $\|f\|_p \leq 1$.

For $D = rB^d$ one obtains $\sup\varrho/\inf\varrho \leq e^{2\mathrm{L}r}$. Hence $C = e^{2\mathrm{L}r}$ and to have tractability with respect to C (see (4.3)), the goal is to show an error bound which depends polynomially on $\mathrm{L}\,r$. In general one has the classes of functions

$$\mathcal{F}_p^{\mathrm{L}}(D) = \{(f,\varrho) \mid \varrho \in \mathcal{R}^{\mathrm{L}}(D),\ \|f\|_p \leq 1\},$$

where

$$\mathcal{R}^{\mathrm{L}}(D) = \{\varrho > 0 \mid \varrho \text{ is log-concave, } |\log\varrho(x) - \log\varrho(y)| \leq \mathrm{L}\|x-y\|_{\mathrm{E}}\}.$$

The idea is to apply the Metropolis algorithm to obtain a Markov chain with stationary distribution π_ϱ (see Section 3.4). The candidate transition kernel on $(D,\mathcal{B}(D))$ is given by the ball walk. This random walk is used in [MN07, Rud09] and studied in different references on volume computation (see e.g. [LS93, Vem05]).

The transition kernel of the δ *ball walk* is given by

$$Q_\delta(x,A) = \frac{\mathbf{vol}_d(B(x,\delta)\cap A)}{\mathbf{vol}_d(\delta B^d)} + \left(1 - \frac{\mathbf{vol}_d(B(x,\delta)\cap D)}{\mathbf{vol}_d(\delta B^d)}\right)\mathbf{1}_A(x), \quad x\in D,\ A\in\mathcal{B}(D),$$

where $\delta > 0$ and $\mathbf{vol}_d(A)$ denotes the d-dimensional Lebesgue measure of $A \in \mathcal{B}(D)$. Schematically, a single step of the δ ball walk from state x may be viewed as in the procedure Ball-Walk(x,δ).

Procedure Ball-Walk(x,δ)

input: current state x, radius δ.
output: next state y.

Choose y uniformly distributed in $B(x,\delta)$;
if $y \in D$ **then**
  Return y;
else
  Return x;
end

Let us state some well known properties.

LEMMA 4.3 (see [MN07, Proposition 1, p. 685]). *The transition kernel Q_δ is reversible with respect to the uniform distribution on D.*

The *local conductance* of the ball walk is defined by

$$l(x) = \frac{\mathbf{vol}_d(B(x,\delta)\cap D)}{\mathbf{vol}_d(\delta B^d)}, \quad x\in D.$$

We call l a *lower bound of the local conductance* if $l(x) \geq l$ for all $x \in D$. Note that l might be very small. For $D = [0,1]^d$, the d-dimensional unit cube, one finds even for small δ that $l = 2^{-d}$. However, one can show for $D = rB^d$ and $\delta \leq r/\sqrt{d+1}$ that $l = 0.3$ is a lower bound of the local conductance.

LEMMA 4.4. *Let Q_δ be the transition kernel of the ball walk on $D = rB^d$ for $r > 0$. If $\delta \leq r/\sqrt{d+1}$, then $l = 0.3$ is a lower bound of the local conductance of the ball walk.*

Proof. The assertion follows by the same arguments as in [MN07, Lemma 7, p. 687] (see also [Rud07]). The only difference is that rB^d is a ball with radius r instead of being the unit ball. ∎

The *Metropolis transition kernel* based on the δ ball walk is

$$K_{\varrho,\delta}(x,A) = \int_A \theta(x,y)\, Q_\delta(x,\mathrm{d}y) + \mathbf{1}_A(x)\left(\int_D (1-\theta(x,y))\, Q_\delta(x,\mathrm{d}y)\right),$$

where the acceptance probability is $\theta(x,y) = \min\{1, \varrho(y)/\varrho(x)\}$ for $x, y \in D$ and $A \in \mathcal{B}(D)$. The lazy version of $K_{\varrho,\delta}$ is denoted by $\widetilde{K}_{\varrho,\delta}$. The transition kernel $\widetilde{K}_{\varrho,\delta}$ is reversible with respect to π_ϱ. In Algorithm 1 we present the integration algorithm S^δ_{n,n_0} which

uses the lazy version of the Metropolis transition kernel with the suggested transition kernel Q_δ.

Algorithm 1: S^δ_{n,n_0}

input: n, n_0, δ, (f, ϱ).
output: $S^\delta_{n,n_0}(f,\varrho)$.

Choose X_1 uniformly distributed in D;
for $k = 1$ **to** $n + n_0$ **do**
  if `rand()` > 0.5 **then**
    $X_{k+1} := X_k$;
  else
    $Y :=$Ball-Walk(X_k, δ);
    if $\varrho(Y)/\varrho(X_i) \geq$ `rand()` **then**
      $X_{i+1} := Y$;
    else
      $X_{i+1} := X_i$;
    end
  end
end
Compute

$$S^\delta_{n,n_0}(f,\varrho) := \frac{1}{n}\sum_{i=1}^{n} f(X_{i+n_0}).$$

It is convenient to use the notation $P_K = P$, $\beta_K = \beta$ and $\Lambda_K = \Lambda$ to indicate the transition kernel K. The following lemma provides a lower bound of the L_2-spectral gap of $P_{\widetilde{K}_{\varrho,\delta}}$. The lemma follows from a result of Mathé and Novak [MN07, Theorem 4, p. 690], where an estimate of the conductance of $K_{\varrho,\delta}$ is shown.

PROPOSITION 4.5. *For $r > 0$ let $D \subset \mathbb{R}^d$ be a convex body with*

$$\mathbf{d}\mathrm{iam}(D) = \sup\{\|x - y\|_{\mathrm{E}} \mid x, y \in D\} \leq 2r.$$

Let l be a lower bound of the local conductance of the δ ball walk. Then, for all $\varrho \in \mathcal{R}^{\mathrm{L}}(D)$, for the lazy version of the Metropolis transition kernel based on a δ ball walk, given by $\widetilde{K}_{\varrho,\delta}$, one has

$$1 - \beta_{\widetilde{K}_{\varrho,\delta}} \geq \frac{l^2 e^{-2\mathrm{L}\delta}}{256} \min\left\{\frac{\pi}{8}\frac{l^2\delta^2}{r^2(d+1)}, 1\right\}.$$

Proof. One has $\beta_{\widetilde{K}_{\varrho,\delta}} = \Lambda_{\widetilde{K}_{\varrho,\delta}} = \frac{1}{2}(1 + \Lambda_{K_{\varrho,\delta}})$. The conductance of $K_{\varrho,\delta}$ is defined by

$$\varphi(K_{\varrho,\delta}, \pi_\varrho) = \inf_{0<\pi_\varrho(A)\leq 1/2} \frac{\int_A K_{\varrho,\delta}(x, A^c)\, \pi_\varrho(\mathrm{d}x)}{\pi_\varrho(A)}.$$

One can use the Cheeger inequality (see Proposition A.7)

$$1 - \Lambda_{K_{\varrho,\delta}} \geq \varphi(K_{\varrho,\delta}, \pi_\varrho)^2/2.$$

Altogether one obtains

$$1-\beta_{\widetilde{K}_{\varrho,\delta}} = 1/2(1-\Lambda_{K_{\varrho,\delta}}) \geq \varphi(K_{\varrho,\delta},\pi_\varrho)^2/4. \tag{4.5}$$

In [MN07, Theorem 4, p. 690] it is shown that

$$\varphi(K_{\varrho,\delta},\pi_\varrho) \geq \frac{le^{-\mathrm{L}\delta}}{8}\min\left\{\sqrt{\frac{\pi}{8}}\,\frac{l\delta}{r\sqrt{d+1}},1\right\}.$$

Plugging this lower bound into (4.5) proves the assertion. ∎

In the previous result one can see that the lower bound of the local conductance is crucial. This motivates considering $D = rB^d$, since by Lemma 4.4 a lower bound of the local conductance is provided. An immediate consequence of the last proposition is

COROLLARY 4.6. *For $r > 0$ let $D = rB^d$, assume that $\varrho \in \mathcal{R}^{\mathrm{L}}(rB^d)$ and set $\delta^* = \min\{1/\mathrm{L}, r/\sqrt{d+1}\}$. Then*

$$1-\beta_{\widetilde{K}_{\varrho,\delta^*}} \geq \frac{1.69\cdot 10^{-6}}{d+1}\min\left\{\frac{1}{r^2\,\mathrm{L}^2},\frac{1}{d+1}\right\}.$$

Proof. This follows from Proposition 4.5 and Lemma 4.4. ∎

In particular one deduces that the lazy version of the ball walk has an L_2-spectral gap, since one can consider constant densities where $\mathrm{L} = 0$.

COROLLARY 4.7. *For $r > 0$ let $D = rB^d$ and $\delta = r/\sqrt{d+1}$. Then the lazy version $\widetilde{Q}_\delta$ of the transition kernel of the ball walk obeys*

$$1-\beta_{\widetilde{Q}_\delta} \geq \frac{1.69\cdot 10^{-6}}{(d+1)^2}.$$

Now we can apply the error bounds of Section 3.2. The next theorem states an error bound for $S^{\delta^*}_{n,n_0}(f,\varrho)$ where $(f,\varrho) \in \mathcal{F}^{\mathrm{L}}_p(rB^d)$.

THEOREM 4.8. *For $r > 0$ let $D = rB^d$ and let ν be the uniform distribution on rB^d. Let $\varrho \in \mathcal{R}^{\mathrm{L}}(rB^d)$ and $\delta^* = \min\{1/\mathrm{L}, r/\sqrt{d+1}\}$. Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with transition kernel $\widetilde{K}_{\varrho,\delta^*}$ and initial distribution ν. The approximation of $S(f,\varrho)$ is*

$$S^{\delta^*}_{n,n_0}(f,\varrho) = \frac{1}{n}\sum_{i=1}^{n} f(X_{i+n_0}).$$

For $p \in (2,\infty]$ recall that

$$\mathcal{F}^{\mathrm{L}}_p(rB^d) = \{(f,\varrho) \mid \varrho \in \mathcal{R}^{\mathrm{L}}(rB^d),\ \|f\|_p \leq 1\}.$$

Let $n_0(p)$ be the smallest natural number (including zero) greater than or equal to

$$5.92\cdot 10^6\,(d+1)\max\{r^2\,\mathrm{L}^2, d+1\}\cdot\begin{cases}\dfrac{p}{p-2}\left(\mathrm{L}\,r + 0.5\log\dfrac{32p}{p-2}\right), & p\in(2,4),\\ 2\mathrm{L}\,r + 4.16, & p\in[4,\infty].\end{cases}$$

Then

$$e(S^{\delta^*}_{n,n_0(p)},\mathcal{F}^{\mathrm{L}}_p(rB^d)) \leq \frac{1089}{\sqrt{n}}\sqrt{d+1}\max\{r\,\mathrm{L},\sqrt{d+1}\} + \frac{8.38\cdot 10^5}{n}(d+1)\max\{r^2\,\mathrm{L}^2, d+1\}.$$

Proof. The initial distribution obeys

$$\nu(A) = \frac{\mathbf{vol}_d(A)}{\mathbf{vol}_d(rB^d)} = \frac{1}{\mathbf{vol}_d(rB^d)} \int_A \int_{rB^d} \frac{\varrho(y)}{\varrho(x)} \,\mathrm{d}y \,\pi_\varrho(\mathrm{d}x), \quad A \in \mathcal{B}(rB^d).$$

Since $\log \varrho$ is Lipschitz continuous with Lipschitz constant L we obtain

$$e^{-2\mathrm{L}r} \le \frac{\varrho(y)}{\varrho(x)} \le e^{2\mathrm{L}r}, \quad x, y \in rB^d,$$

so that

$$\left\| \frac{d\nu}{d\pi_\varrho} - 1 \right\|_p \le \left\| \frac{d\nu}{d\pi_\varrho} - 1 \right\|_\infty \le \max\{1, e^{2\mathrm{L}r}\} = e^{2\mathrm{L}r}.$$

By Corollary 4.6 we have the crucial lower bound for the spectral gap $1 - \beta_{\widetilde{K}_{\varrho,\delta^*}}$ and consequently Theorem 3.45 (ii) can be applied, which proves the assertion. ∎

Note that $p \in (2, \infty]$ is necessary to apply Theorem 3.45 (ii). An essential consequence of the last theorem is the following result concerning the tractability of (4.4).

THEOREM 4.9. *For the integration problem $S(f, \varrho)$ defined over $\mathcal{F}_p^{\mathrm{L}}(rB^d)$ with $r > 0$ and $p > 2$ we have*

$$\mathbf{comp}(\varepsilon, d, \mathcal{F}_p^{\mathrm{L}}(rB^d)) \le (d+1) \max\{r^2 \,\mathrm{L}^2, d+1\}$$
$$\cdot \left[4.8 \cdot 10^6 \,\varepsilon^{-2} + 1.2 \cdot 10^6 \cdot \begin{cases} \dfrac{p}{p-2} \left(\mathrm{L}\, r + 0.5 \log \dfrac{32p}{p-2} \right), & p \in (2,4) \\ 2\mathrm{L}\, r + 4.16, & p \in [4, \infty] \end{cases} \right]$$

for all $\varepsilon \in (0,1)$ and $d \in \mathbb{N}$.

The last theorem states that the problem (4.4) is polynomially tractable. Roughly speaking, for fixed p one obtains

$$\mathbf{comp}(\varepsilon, d, \mathcal{F}_p^{\mathrm{L}}(rB^d)) \prec d \max\{r^2 \,\mathrm{L}^2, d\}(\varepsilon^{-2} + \mathrm{L}\, r),$$

so that the dependence on L, the precision ε, dimension d and r is polynomial. We also have tractability with respect to $C = e^{2r\mathrm{L}}$: inequality (4.3) holds with $q_1 = 2$, $q_2 = 2$ and $q_3 = 3$. For $p \in [4, \infty]$ the complexity can be bounded independently of p, and for fixed $p \in (2, \infty]$ we showed that the integration problem on $\mathcal{F}_p^{\mathrm{L}}(rB^d)$ is polynomially tractable in the sense of (4.3).

4.2. Integration over a convex body. The goal is to compute

$$S(f, A) = \frac{1}{\mathbf{vol}_d(A)} \int_A f(x) \,\mathrm{d}x \tag{4.6}$$

for $A \subset \mathbb{R}^d$. In other words, $S(f, A)$ is the expectation of f with respect to the uniform distribution, say μ_A, on $A \subset \mathbb{R}^d$. The domain A and the function f are the input quantities. This fits in the class of problems described by (4.2) if we assume that $A \subset D$. Then μ_A might be considered as given by a density which is an indicator function.

For some domains A it is indeed simple to generate uniformly distributed random points, e.g. the Euclidean unit ball or the unit cube. Then one can approximate $S(f, A)$ by Monte Carlo methods with an i.i.d. sample. However, here A is part of the input to the algorithm, thus the problem $S(f, A)$ shall be solved uniformly for a class of state

spaces, where we cannot assume that sampling with respect to the uniform distribution is possible.

Let $r \geq 1$ and let

$$\mathcal{S}_d(r) = \{A \subset \mathbb{R}^d \,\text{convex} \mid B^d \subset A \subset rB^d\}.$$

If $A \in \mathcal{S}_d(r)$ then A is a convex bounded set with non-empty interior which contains the origin. The class of input parameters is given by

$$\mathcal{F}_p(r,d) = \{(f,A) \mid \|f\|_p \leq 1,\, A \in \mathcal{S}_d(r)\}.$$

We assume that for any $A \in \mathcal{S}_d(r)$ there exists an oracle $\mathrm{Or}_A(\ell)$ which returns for an arbitrary line ℓ a uniformly distributed random point on $A \cap \ell$.

Let us comment on this assumption. Assume that we have a membership oracle of $A \in \mathcal{S}_d(r)$ which is given by $\widetilde{\mathrm{Or}}_A(x) = \mathbf{1}_A(x)$ for any $x \in rB^d$. The oracle Or_A can be implemented by using the membership oracle. Let $[x,y] = \{x + ty \mid t \in [0,1]\}$ be the segment of $x, y \in \mathbb{R}^d$ with Euclidean distance $\|x-y\|_{\mathrm{E}}$. By the convexity of A it follows that $A \cap \ell$ is a single segment, hence there exist $a_1, a_2 \in \mathbb{R}^d$ such that $[a_1,a_2] = A \cap \ell$. Suppose that $\ell = \{\tilde{x} + t\,\mathbf{d}\mathrm{ir} \mid t \in \mathbb{R}\}$ with $\tilde{x} \in A$ and assume that there is a positive number ε_0 such that $\|a_1 - a_2\|_{\mathrm{E}} \geq \varepsilon_0$. We use that $A \in \mathcal{S}_d(r)$ and $\tilde{x} \in A$. By a bisection method one can find with at most $3\log(2r/\varepsilon_0) + 2$ calls to $\widetilde{\mathrm{Or}}_A$ a segment $[b_1,b_2]$ with $b_1, b_2 \in \mathbb{R}^d$ and $[a_1,a_2] \subset [b_1,b_2]$ such that

$$\tfrac{1}{6}\|b_1 - b_2\|_{\mathrm{E}} \leq \|a_1 - a_2\|_{\mathrm{E}} \leq \|b_1 - b_2\|_{\mathrm{E}}.$$

Then choose a uniformly distributed random point in $[b_1,b_2]$ and accept it if it is in A, otherwise reject it and repeat the acceptance/rejection procedure. This procedure gives a uniformly distributed random point in $A \cap \ell$ and works reasonably fast, since the acceptance probability is $1/6$. Altogether a call to Or_A requires at most an expected number of $3\log(2r/\varepsilon_0) + 8$ calls to $\widetilde{\mathrm{Or}}_A$. In the analysis of the error we count the calls to Or_A and the function evaluations of f, i.e. the calls to Or_f.

Now let us provide a Markov chain on the measurable space $(A, \mathcal{B}(A))$ with stationary distribution μ_A. We consider the hit-and-run algorithm, also called hypersphere directions algorithm (see [Smi84]). The algorithm is studied and analyzed in [Lov99, LV06]. The work of Vempala [Vem05] provides an introduction to geometric random walks.

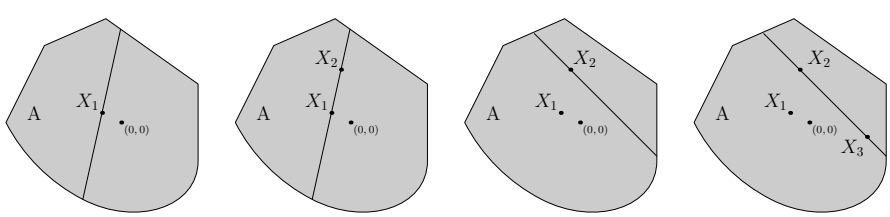

Fig. 1. Illustration of the generation of X_3 and X_2 by the hit-and-run algorithm given state X_1

The algorithm is as follows. Suppose that the current position is $X_i \in A$ with $i \in \mathbb{N}$. Then choose a uniformly distributed direction, say $\mathbf{d}\mathrm{ir}_i$, and consider the line which is defined by $\ell^{(i)} = \{X_i + t\,\mathbf{d}\mathrm{ir}_i \mid t \in \mathbb{R}\}$. Apply $\mathrm{Or}_A(\ell^{(i)})$, which gives the next state X_{i+1}

chosen uniformly distributed in $\ell^{(i)} \cap A$. Then, again, a uniformly distributed direction, say $\mathbf{dir}_{i+1}$, is generated and the next state is chosen uniformly distributed on $\ell^{(i+1)} \cap A$ by $\mathrm{Or}_A(\ell^{(i+1)})$. Two consecutive steps of the hit-and-run algorithm are illustrated in Figure 1. Recall that the Euclidean unit ball is denoted by B^d and its boundary is denoted by ∂B^d. Schematically, a single step of the hit-and-run algorithm from $x \in A$ is presented in the Procedure Hit-and-Run(x).

Procedure Hit-and-Run(x)

input: current state x.
output: next state y.

Choose a direction $\mathbf{dir}$ uniformly distributed on ∂B^d;
Choose y uniformly distributed on

$$A \cap \{x + t\,\mathbf{dir} \mid t \in \mathbb{R}\};$$

Return y.

The transition kernel of the hit-and-run algorithm follows. For any $x, y \in \mathbb{R}^d$ let

$$\mathbf{Int}(x,y) = \left\{ \lambda \in \mathbb{R} \;\middle|\; x + \lambda \frac{y-x}{\|y-x\|_{\mathrm{E}}} \in A \right\}.$$

Since A is convex, $\mathbf{Int}(x,y)$ is an interval. Let

$$\lambda_1(x,y) = \min\{\alpha \mid \alpha \in \mathbf{Int}(x,y)\} \quad \text{and} \quad \lambda_2(x,y) = \max\{\alpha \mid \alpha \in \mathbf{Int}(x,y)\},$$

which implies that $\mathbf{Int}(x,y) = [\lambda_1(x,y), \lambda_2(x,y)]$. The length of the chord $\mathbf{Int}(x,y)$ is given by $\ell(x,y) = \lambda_2(x,y) - \lambda_1(x,y)$. Let $U(x,y)$ be a uniformly distributed random variable in the interval $\mathbf{Int}(x,y)$. Then the *hit-and-run transition kernel* H of the hit-and-run algorithm is

$$\begin{aligned}
H(x,C) &= \frac{\int_{\partial B^d} \Pr[x + U(x,x+\theta)\theta \in C]\,\mathrm{d}\theta}{\mathbf{vol}_{d-1}(\partial B^d)} \\
&= \frac{1}{\mathbf{vol}_{d-1}(\partial B^d)} \int_{\partial B^d} \int_{\lambda_1(x,x+\theta)}^{\lambda_2(x,x+\theta)} \frac{\mathbf{1}_C(x+\lambda\theta)}{\ell(x,x+\theta)}\,\mathrm{d}\lambda\,\mathrm{d}\theta \\
&= \frac{1}{\mathbf{vol}_{d-1}(\partial B^d)} \int_{\partial B^d} \int_{\lambda_1(x,x+\theta)}^{0} \frac{\mathbf{1}_C(x+\lambda\theta)}{\ell(x,x+\theta)}\,\mathrm{d}\lambda\,\mathrm{d}\theta \\
&\quad + \frac{1}{\mathbf{vol}_{d-1}(\partial B^d)} \int_{\partial B^d} \int_{0}^{\lambda_2(x,x+\theta)} \frac{\mathbf{1}_C(x+\lambda\theta)}{\ell(x,x+\theta)}\,\mathrm{d}\lambda\,\mathrm{d}\theta \\
&= \frac{2}{\mathbf{vol}_{d-1}(\partial B^d)} \int_C \frac{1\,\mathrm{d}y}{\ell(x,y)\|x-y\|_{\mathrm{E}}^{d-1}},
\end{aligned} \tag{4.7}$$

where $x \in A$ and $C \in \mathcal{B}(A)$. The last equality follows by the integral transformation formula

$$\int_{\mathbb{R}^d} h(y)\,\mathrm{d}y = \int_{\partial B^d} \int_0^\infty h(g(\lambda,\theta))\lambda^{d-1}\,\mathrm{d}\lambda\,\mathrm{d}\theta$$

with

$$h(y) = \frac{\mathbf{1}_C(y)}{\ell(x,y)\|x-y\|_{\mathrm{E}}^{d-1}}$$

and either $g(\lambda,\theta) = x + \lambda\theta$ or $g(\lambda,\theta) = x - \lambda\theta$.

LEMMA 4.10. *The hit-and-run transition kernel H, given by* (4.7)*, is reversible with respect to μ_A on A.*

Proof. Let $k(x,y)$ be a symmetric transition density of a transition kernel K, i.e. $k(x,y) = k(y,x)$ for all $x, y \in A$. Then it follows by Fubini's theorem that

$$\begin{aligned}\int_B K(x,C)\,\mu_A(\mathrm{d}x) &= \int_B \int_C k(x,y)\,\mu_A(\mathrm{d}y)\,\mu_A(\mathrm{d}x) = \int_C \int_B k(x,y)\,\mu_A(\mathrm{d}x)\,\mu_A(\mathrm{d}y)\\ &= \int_C \int_B k(y,x)\,\mu_A(\mathrm{d}x)\,\mu_A(\mathrm{d}y) = \int_C K(x,B)\,\mu_A(\mathrm{d}x),\ B, C \in \mathcal{B}(A).\end{aligned}$$

Hence the transition kernel K is reversible with respect to μ_A. Since $\ell(x,y) = \ell(y,x)$, the transition kernel H has a symmetric density and this implies that it is reversible with respect to μ_A. ∎

The lazy version of H is denoted by $\widetilde{H}$. In Algorithm 2 we present the integration algorithm S^{har}_{n,n_0} which uses the lazy version of the hit-and-run transition kernel. We

Algorithm 2: S^{har}_{n,n_0}

input: n, n_0, (f, A).
output: $S^{\mathrm{har}}_{n,n_0}(f,A)$.

Choose X_1 uniformly distributed in B^d;
for $k = 1$ **to** $n + n_0$ **do**
  if `rand()` > 0.5 **then**
    $X_{k+1} := X_k$;
  else
    X_{k+1} :=Hit-and-Run(X_k);
  end
end
Compute

$$S^{\mathrm{har}}_{n,n_0}(f,A) := \frac{1}{n}\sum_{i=1}^{n} f(X_{i+n_0}).$$

use the notation $P_K = P$, $\beta_K = \beta$ and $\Lambda_K = \Lambda$ to indicate the transition kernel K. The following lemma provides a lower bound of the L_2-spectral gap of $P_{\widetilde{H}}$. The lemma is a straightforward implication of a result of Lovász and Vempala [LV06, Theorem 4.2, p. 993]. Lovász and Vempala show an estimate of the conductance of H.

PROPOSITION 4.11. *Let $r \geq 1$. Then, for all $A \in \mathcal{S}_d(r)$, for the lazy version of the hit-and-run transition kernel, given by $\widetilde{H}$, one has*

$$1 - \beta_{\widetilde{H}} \geq 2^{-52}(dr)^{-2}.$$

Proof. In [LV06, Theorem 4.2, p. 993] it is proven that

$$\varphi(H, \mu_A) \geq 2^{-25}(dr)^{-1}.$$

Then one follows the same arguments as in the proof of Lemma 4.5. ∎

Now we can apply the error bounds of Section 3.2 to obtain the following.

THEOREM 4.12. *Let ν be the uniform distribution on B^d. Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with transition kernel $\widetilde{H}$ and initial distribution ν. The approximation of $S(f, A)$ is*

$$S^{\mathrm{har}}_{n,n_0}(f, A) = \frac{1}{n}\sum_{i=1}^{n} f(X_{i+n_0}).$$

For $r \geq 1$ and $p > 2$ recall that

$$\mathcal{F}_p(r, d) = \{(f, A) \mid \|f\|_p \leq 1,\ A \in \mathcal{S}_d(r)\}.$$

Let $n_0(p)$ be the smallest natural number (including zero) greater than or equal to

$$4.51 \cdot 10^{15}\, d^2\, r^2 \cdot \begin{cases} \dfrac{p}{2(p-2)}\left(d\log r + \log\dfrac{32p}{p-2}\right), & p \in (2, 4), \\ d\log r + 4.16, & p \in [4, \infty]. \end{cases}$$

Then

$$e(S^{\mathrm{har}}_{n,n_0(p)}, \mathcal{F}_p(r, d)) \leq 9.5 \cdot 10^7\, \frac{dr}{\sqrt{n}} + 6.4 \cdot 10^{15}\, \frac{d^2\, r^2}{n}.$$

Proof. Note that the initial distribution ν is well defined, since for $A \in \mathcal{S}_d(r)$ one has $B^d \subset A \subset rB^d$. Furthermore, it follows that

$$\nu(C) = \frac{1}{\mathbf{vol}_d(B^d)} \int_C \mathbf{1}_{B^d}(x)\,\mathrm{d}x = \frac{1}{\mathbf{vol}_d(A)} \int_C \mathbf{1}_{B^d}(x) \frac{\mathbf{vol}_d(A)}{\mathbf{vol}_d(B^d)}\,\mathrm{d}x, \quad C \in \mathcal{B}(A).$$

One obtains

$$\left\| \frac{d\nu}{d\mu_A} - 1 \right\|_p \leq \left\| \frac{d\nu}{d\mu_A} - 1 \right\|_\infty \underset{r\in[1,\infty)}{\leq} \frac{\mathbf{vol}_d(rB^d)}{\mathbf{vol}_d(B^d)} = r^d.$$

By Lemma 4.11 we have the crucial lower bound for the spectral gap $1 - \beta_{\widetilde{H}}$ and consequently Theorem 3.45(ii) can be applied. Hence the assertion follows. ∎

Note that $p > 2$ is necessary to apply Theorem 3.45(ii). A consequence of the last theorem is the following result concerning the tractability of the integration problem (4.6).

THEOREM 4.13. *For the integration problem $S(f, A)$ defined over $\mathcal{F}_p(r, d)$ with $r \geq 1$ and $p > 2$ we have*

$$\mathbf{c}\mathrm{omp}(\varepsilon, \mathcal{F}_p(r, d)) \leq d^2 r^2 \left[4 \cdot 10^{16}\, \varepsilon^{-2} + 5 \cdot 10^{15} \begin{cases} \dfrac{p}{2(p-2)}\left(d\log r + \log\dfrac{32p}{p-2}\right), & p \in (2, 4) \\ d\log r + 4.16, & p \in [4, \infty] \end{cases} \right]$$

for all $\varepsilon \in (0, 1)$ and $d \in \mathbb{N}$.

The last theorem states that (4.6) is polynomially tractable. Roughly speaking, for fixed p one obtains

$$\mathbf{c}\mathrm{omp}(\varepsilon, \mathcal{F}_p(r,d)) \asymp d^2 r^2 (\varepsilon^{-2} + d \log r),$$

so that the dependence on the precision ε, dimension d and r is polynomial. For $p \in [4, \infty]$ the complexity can be bounded independently of p, and for fixed $p > 2$ we showed that the integration problem is polynomially tractable on $\mathcal{F}_d(r,p)$.

4.3. Notes and remarks. Let us briefly summarize the features of the last sections and provide additional results from the literature. In Section 4.1 elementary state spaces were considered, namely balls, and the distribution π_ϱ determined by ϱ could be complicated. In Section 4.2 the distribution of interest was simple, namely the uniform one, and the state space was possibly complicated.

The problem of integration (4.1), stated in the form

$$S(f,\varrho) = \frac{\int_D f(x)\varrho(x)\,\mathrm{d}x}{\int_D \varrho(x)\,\mathrm{d}x},$$

is formulated as in the work of Mathé and Novak [MN07]. There the authors also proved an asymptotic error bound for the Metropolis algorithm based on the ball walk on $\mathcal{F}_2^{\mathrm{L}}(B^d)$. They studied the algorithm $S_{n,0}^{\delta^*}$ and for $\delta^* = \min\{(d+1)^{-1/2}, \mathrm{L}^{-1}\}$ it is shown in [MN07, Theorem 5, p. 693] that

$$\lim_{n\to\infty} ne(S_{n,0}^{\delta^*}, \mathcal{F}_2^{\mathrm{L}}(B^d))^2 \le 594700 \cdot (d+1)\max\{d+1, \mathrm{L}^2\}.$$

The first non-asymptotic error bound is proven in [Rud09] for the class $\mathcal{F}_\infty^{\mathrm{L}}(B^d)$. It states that for $n_0 \ge 1.28 \cdot 10^6 \cdot \mathrm{L}(d+1)\max\{d+1, \mathrm{L}^2\}$ the error obeys

$$e(S_{n,n_0}^{\delta^*}, \mathcal{F}_\infty^{\mathrm{L}}(B^d)) \le \frac{8000}{\sqrt{n}}\sqrt{d+1}\max\{\sqrt{d+1}, \mathrm{L}\}.$$

Theorem 4.8 extends this result. The integrands f belong to L_p for $p > 2$ and we considered the domain rB^d. The constants in the error bound are of the same order of magnitude and the dependence on the dimension d, the Lipschitz constant L and the precision ε is the same. The problem is tractable in the sense of (4.3).

Apart from the asymptotic result of [MN07, Theorem 5, p. 693] it is always assumed that the integrand f belongs to L_p for $p > 2$. The case of $f \in L_2$ has not been covered so far. To apply Theorem 3.34 it is sufficient to have a transition kernel which is reversible with respect to the desired distribution and uniformly ergodic with (α, M). It is well known that the ball walk, the Metropolis algorithm based on the ball walk and the hit-and-run algorithm are uniformly ergodic (see [Smi84, KS98, MN07]). However, as far as we know there is no estimate for the parameters $\alpha \in [0,1)$ and $M < \infty$ of the uniform ergodicity, guaranteeing polynomial tractability. We get polynomial tractability if there exist non-negative numbers c and q such that $(1-\alpha)^{-1} \le c\, d^q$.

Let $D = B^d$ and $\delta = 2/\sqrt{d+1}$. Then the ball walk Q_δ is uniformly ergodic with (α, M), where

$$\alpha = 1 - \frac{0.15}{\sqrt{d+1}((d+1)2^{d+1})^{\sqrt{d+1}}} \quad \text{and} \quad M = 100.$$

Unfortunately the crucial quantity $(1-\alpha)^{-1}$ is exponentially bad in d. Hence, this is not enough to prove polynomial tractability. It is not clear if one can get a significantly better α.

The hit-and-run algorithm is studied in various references on volume computation and optimization. However, as far as we know it has not yet been applied to integration problems of the form of (4.6). There is an immediate generalization of the hit-and-run algorithm which can be used to sample a distribution given by a log-concave density (see for example [LV06, p. 987]). This might be used to obtain further error bounds for other classes of functions.

Appendix

Some aspects of functional analysis are fundamental for the understanding of the error of Markov chain Monte Carlo. We present the spectral theorem for self-adjoint bounded linear operators. Then we state the interpolation theorem of Riesz–Thorin for operators acting on L_p. Afterwards the conductance and the Cheeger inequality are recalled.

A.1. Spectral theorem. We state the spectral theorem for self-adjoint bounded linear operators. For further reading, proofs and details we refer to [KG82, Rud91, Tri92]. For an introduction see [Kre89].

Let H be a real or complex Hilbert space and let $\mathcal{L}(H)$ be the space of all bounded linear operators from H to H. Let $\mathcal{B}(\mathbb{R})$ be the Borel σ-algebra over $\mathbb{R}$.

Definition A.1 (spectral measure). A *spectral measure* or a *projection-valued measure* is a mapping $E\colon \mathcal{B}(\mathbb{R}) \to \mathcal{L}(H)$ with the following properties:

(i) for all $A \in \mathcal{B}(\mathbb{R})$ the operator E_A is an orthogonal projection,
(ii) $E_\emptyset = 0$, $E_{\mathbb{R}} = I$, where I is the identity,
(iii) for pairwise disjoint $A_1, A_2, \ldots \in \mathcal{B}(\mathbb{R})$ and for any $g \in H$,

$$\sum_{i=1}^{\infty} E_{A_i}(g) = E_{\bigcup_{i=1}^{\infty} A_i}(g).$$

If there exists a compact set $K \subset \mathbb{R}$ with $E_K = I$, then we say that the spectral measure has *compact support*.

For $f, g \in H$ a signed measure is defined on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ by

$$\omega(A) = \langle E_A f, g\rangle, \quad A \in \mathcal{B}(\mathbb{R}).$$

If $f = g$, then the measure ω is non-negative. Let $P \in \mathcal{L}(H)$ be a self-adjoint operator and denote its spectrum by $\mathbf{spec}(P)$. Furthermore let

$$\lambda = \inf_{\|g\|=1} \langle Pg, g\rangle \quad \text{and} \quad \Lambda = \sup_{\|g\|=1} \langle Pg, g\rangle.$$

The spectrum of P is closed and $\mathbf{spec}(P) \subset [\lambda, \Lambda]$. Additionally $\lambda, \Lambda \in \mathbf{spec}(P)$, thus

$$\lambda = \inf\{\alpha \mid \alpha \in \mathbf{spec}(P)\} \quad \text{and} \quad \Lambda = \sup\{\alpha \mid \alpha \in \mathbf{spec}(P)\}.$$

Now we state the spectral theorem for self-adjoint bounded linear operators. It is an analogue to the finite-dimensional spectral theorem for matrices.

PROPOSITION A.2 (Spectral Theorem). *Let $P \in \mathcal{L}(H)$ be self-adjoint and $k \in \mathbb{N}$. Then there exists a unique spectral measure E with support $\mathbf{spec}(P)$ such that*

$$\langle P^k f, g\rangle = \int_\lambda^\Lambda \alpha^k \,\mathrm{d}\langle E_{\{\alpha\}} f, g\rangle, \quad f, g \in H. \tag{A.1}$$

Let $F\colon [\lambda, \Lambda] \to \mathbb{R}$ be a continuous function. Then by the continuous functional calculus one has a self-adjoint operator $F(P) \in \mathcal{L}(H)$ with

$$\langle F(P) f, g\rangle = \int_\lambda^\Lambda F(\alpha) \,\mathrm{d}\langle E_{\{\alpha\}} f, g\rangle, \quad f, g \in H, \tag{A.2}$$

and

$$\|F(P)\|_{H\to H} = \max_{\alpha \in \mathbf{spec}(P)} |F(\alpha)|.$$

REMARK A.3. Mostly in the literature the case where H is a complex Hilbert space is considered. [KG82] handles both real and complex Hilbert spaces. Note that the integral in (A.1) and (A.2) is defined with respect to a signed measure.

A.2. Interpolation theorem. We state a version of the Riesz–Thorin theorem. For a proof and further details let us refer to [BL76, BS88]. Let $L_p = L_p(D, \pi)$ for a probability measure π on a measurable space $(D, \mathfrak{D})$.

PROPOSITION A.4 (Riesz–Thorin Theorem). *Let $1 \le p, q_1, q_2 \le \infty$. Assume that $\theta \in (0, 1)$ and*

$$\frac{1}{p} = \frac{1-\theta}{q_1} + \frac{\theta}{q_2}.$$

Further let T be a linear operator from L_{q_1} to L_{q_1} and at the same time from L_{q_2} to L_{q_2} with

$$\|T\|_{L_{q_1}\to L_{q_1}} \le M_1 \quad \text{and} \quad \|T\|_{L_{q_2}\to L_{q_2}} \le M_2.$$

Then

$$\|T\|_{L_p \to L_p} \le 2 M_1^{1-\theta} M_2^{\theta}.$$

REMARK A.5. We can replace the function spaces L_p, L_{q_1}, L_{q_2} in the last proposition by the sequence spaces ℓ_p, ℓ_{q_1}, ℓ_{q_2} and the result remains the same.

REMARK A.6. Note that we consider real-valued functions. For functions which map into the complex numbers, the same result holds true and the factor of two is not needed.

A.3. Conductance and the Cheeger inequality. Let $(D, \mathfrak{D})$ be a measurable space. Assume K is a transition kernel defined on $(D, \mathfrak{D})$ which is reversible with respect to a probability measure π. The *conductance* of the transition kernel K is defined by

$$\varphi(K, \pi) = \inf_{0 < \pi(A) \le 1/2} \frac{\int_A K(x, A^c)\, \pi(\mathrm{d}x)}{\pi(A)}.$$

Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with transition kernel K and initial distribution π. Then the numerator of the ratio above is the probability that $X_1 \in A$ and $X_2 \in A^c$. Hence

$$\Pr(X_2 \in A^c \mid X_1 \in A) = \frac{\int_A K(x, A^c)\,\pi(\mathrm{d}x)}{\pi(A)}.$$

The conductance of K is the infimum over sets $A \in \mathfrak{D}$, with $0 < \pi(A) \le 1/2$, of the probability that $X_2 \in A^c$ under the condition that $X_1 \in A$.

The Markov operator P given by $Pf(x) = \int_D f(y)\,K(x, \mathrm{d}y)$ is self-adjoint on $L_2 = L_2(D, \pi)$. For $f \in L_2$ let $S(f) = \int_D f(x)\,\pi(\mathrm{d}x)$ and let

$$L_2^0 = \{f \in L_2 \mid S(f) = 0\}.$$

Furthermore define

$$\Lambda = \sup\{\alpha \mid \alpha \in \mathbf{spec}(P|L_2^0)\}.$$

The Cheeger inequality provides a relation between Λ and the conductance $\varphi(K, \pi)$.

Proposition A.7 (Cheeger inequality). *Let the transition kernel K be reversible with respect to the probability measure π. Then*

$$1 - \Lambda \ge \varphi(K, \pi)^2/2. \tag{A.3}$$

For a proof on finite state spaces we refer to [Beh00, Theorem 11.3, p. 93]. The Cheeger inequality for general state spaces is proven by Lawler and Sokal in [LS88, Theorem 3.5, p. 570] and by Lovász and Simonovits in [LS93, Lemma 1.7, p. 374]. Lawler and Sokal provide different types of inequalities for Markov chains and Markov processes.

References

[Ald87] D. Aldous, *On the Markov chain simulation method for uniform combinatorial distributions and simulated annealing*, Probab. Engrg. Inform. Sci. 1 (1987), 33–46.

[AG10] S. Asmussen and P. Glynn, *Harris recurrence and MCMC: A simplified approach*, submitted (2010).

[Bax05] P. Baxendale, *Renewal theory and computable convergence rates for geometrically ergodic Markov chains*, Ann. Appl. Probab. 15 (2005), 700–738.

[BD06] F. Bassetti and P. Diaconis, *Examples comparing importance sampling and the Metropolis algorithm*, Illinois J. Math. 50 (2006), 67–91.

[BR95] J. Baxter and J. Rosenthal, *Rates of convergence for everywhere-positive Markov chains*, Statist. Probab. Lett. 22 (1995), 333–338.

[Beh00] E. Behrends, *Introduction to Markov Chains with Special Emphasis on Rapid Mixing*, Vieweg, Braunschweig, 2000.

[BC09] A. Belloni and V. Chernozhukov, *On the computational complexity of MCMC-based estimators in large samples*, Ann. Statist. 37 (2009), 2011–2055.

[BS88] C. Bennett and R. Sharpley, *Interpolation of Operators*, Pure Appl. Math. 129, Academic Press, Boston, MA, 1988.

[BL76] J. Bergh and J. Löfström, *Interpolation Spaces. An Introduction*, Grundlehren Math. Wiss. 223, Springer, Berlin, 1976.

[Bré99] P. Brémaud, *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*, Texts Appl. Math. 31, Springer, New York, 1999.

[BGJM11] S. Brooks, A. Gelman, G. Jones, and X. Meng, *Handbook of Markov Chain Monte Carlo*, Chapman & Hall, 2011.

[Che05] M. Chen, *Eigenvalues, Inequalities, and Ergodic Theory*, Springer, London, 2005.

[Dia09] P. Diaconis, *The Markov chain Monte Carlo revolution*, Bull. Amer. Math. Soc. (N.S.) 46 (2009), 179–205.

[DS91] P. Diaconis and D. Stroock, *Geometric bounds for eigenvalues of Markov chains*, Ann. Appl. Probab. 1 (1991), 36–61.

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

[FHY92] A. Frigessi, C. Hwang, and L. Younes, *Optimal spectral structure of reversible stochastic matrices, Monte Carlo methods and the simulation of Markov random fields*, Ann. Appl. Probab. 2 (1992), 610–628.

[Gey92] C. Geyer, *Practical Markov chain Monte Carlo*, Statist. Sci. 7 (1992), 473–483.

[GRS96] W. Gilks, S. Richardson, and D. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Chapman & Hall, 1996.

[Gil98] D. Gillman, *A Chernoff bound for random walks on expander graphs*, SIAM J. Comput. 27 (1998), 1203–1220.

[GO02] P. Glynn and D. Ormoneit, *Hoeffding's inequality for uniformly ergodic Markov chains*, Statist. Probab. Lett. 56 (2002), 143–146.

[Häg02] O. Häggström, *Finite Markov Chains and Algorithmic Applications*, London Math. Soc. Student Texts 52, Cambridge Univ. Press, Cambridge, 2002.

[Has70] W. Hastings, *Monte Carlo sampling methods using Markov chains and their applications*, Biometrika 57 (1970), 97–109.

[JS89] M. Jerrum and A. Sinclair, *Approximating the permanent*, SIAM J. Comput. 18 (1989), 1149–1178.

[JHCN06] G. Jones, M. Haran, B. Caffo, and R. Neath, *Fixed-width output analysis for Markov chain Monte Carlo*, J. Amer. Statist. Assoc. 101 (2006), 1537–1547.

[JO10] A. Joulin and Y. Ollivier, *Curvature, concentration and error estimates for Markov chain Monte Carlo*, Ann. Probab. 38 (2010), 2418–2442.

[Kal02] O. Kallenberg, *Foundations of Modern Probability*, 2nd ed., Springer, New York, 2002.

[KS98] D. Kaufman and R. Smith, *Direction choice for accelerated convergence in hit-and-run sampling*, Oper. Res. 46 (1998), 84–95.

[KG82] A. Kirillov and A. Gvishiani, *Theorems and Problems in Functional Analysis*, Springer, New York, 1982.

[Kre89] E. Kreyszig, *Introductory Functional Analysis with Applications*, Wiley, 1989.

[Krü98] R. Krüger, *Gleichmäßige Ergodizität von Markovketten Monte Carlo-Verfahren*, Herbert Utz, München, 1998.

[ŁMN09] K. Łatuszyński, B. Miasojedow, and W. Niemiro, *Nonasymptotic bounds on the estimation error for regenerative MCMC algorithms*, ArXiv e-prints (2009).

[ŁN11] K. Łatuszyński and W. Niemiro, *Rigorous confidence bounds for MCMC under a geometric drift condition*, J. Complexity 27 (2011), 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), 557–580.

[LP04] C. León and F. Perron, *Optimal Hoeffding bounds for discrete reversible Markov chains*, Ann. Appl. Probab. 14 (2004), 958–970.

[LPW09] D. Levin, Y. Peres, and E. Wilmer, *Markov Chains and Mixing Times*, Amer. Math. Soc., Providence, RI, 2009.

[Lez98] P. Lezaud, *Chernoff-type bound for finite Markov chains*, Ann. Appl. Probab. 8 (1998), 849–867.

[Lez01] P. Lezaud, *Chernoff and Berry–Esséen inequalities for Markov processes*, European Series in Applied and Industrial Mathematics. Probab. Statist. 5 (2001), 183–201.

[Liu08] J. Liu, *Monte Carlo Strategies in Scientific Computing*, Springer, New York, 2008.

[Lov99] L. Lovász, *Hit-and-run mixes fast*, Math. Program. Ser. A 86 (1999), 443–461.

[LS93] L. Lovász and M. Simonovits, *Random walks in a convex body and an improved volume algorithm*, Random Structures Algorithms 4 (1993), 359–412.

[LV06] L. Lovász and S. Vempala, *Hit-and-run from a corner*, SIAM J. Comput. 35 (2006), 985–1005.

[Mar99] F. Martinelli, *Lectures on Glauber dynamics for discrete spin models*, in: Lectures on Probability Theory and Statistics (Saint-Flour, 1997), Lecture Notes in Math. 1717, Springer, Berlin, 1999, 93–191.

[Mat99] P. Mathé, *Numerical integration using Markov chains*, Monte Carlo Methods Appl. 5 (1999), 325–343.

[Mat04] P. Mathé, *Numerical integration using V-uniformly ergodic Markov chains*, J. Appl. Probab. 41 (2004), 1104–1112.

[MN07] P. Mathé and E. Novak, *Simple Monte Carlo and the Metropolis algorithm*, J. Complexity 23 (2007), 673–696.

[Mei99] C. Meise, *On spectral gap estimates of a Markov chain via hitting times and coupling*, J. Appl. Probab. 36 (1999), 310–319.

[MT96] K. Mengersen and R. Tweedie, *Rates of convergence of the Hastings and Metropolis algorithms*, Ann. Statist. 24 (1996), 101–121.

[MR$^+$53] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, *Equation of state calculations by fast computing machines*, J. Chemical Phys. 21 (1953), 1087–1092.

[MT09] S. Meyn and R. Tweedie, *Markov Chains and Stochastic Stability*, 2nd ed., Cambridge Univ. Press, 2009.

[Mia10] B. Miasojedow, *Bounds on the estimation error for MCMC algorithms*, Ph.D. thesis, University of Warsaw, 2010 (in Polish).

[NP09] W. Niemiro and P. Pokarowski, *Fixed precision MCMC estimation by median of products of averages*, J. Appl. Probab. 46 (2009), 309–329.

[NW08] E. Novak and H. Woźniakowski, *Tractability of Multivariate Problems. Vol. 1: Linear Information*, EMS Tracts Math. 6, Eur. Math. Soc., Zürich, 2008.

[NW10] E. Novak and H. Woźniakowski, *Tractability of Multivariate Problems. Vol. 2: Standard Information for Functionals*, EMS Tracts Math. 12, Eur. Math. Soc., Zürich, 2010.

[Num84] E. Nummelin, *General Irreducible Markov Chains and Non-Negative Operators*, Cambridge Univ. Press, 1984.

[Oll09] Y. Ollivier, *Ricci curvature of Markov chains on metric spaces*, J. Funct. Anal. 256 (2009), 810–864.

[PW96] J. Propp and D. Wilson, *Exact sampling with coupled Markov chains and applications to statistical mechanics*, in: Proceedings of the Seventh International Conference on Random Structures and Algorithms (Atlanta, GA, 1995), vol. 9, 1996, 223–252.

[Rev84] D. Revuz, *Markov Chains*, 2nd ed., North-Holland Math. Library 11, North-Holland, Amsterdam, 1984.

[RR97a] G. Roberts and J. Rosenthal, *Geometric ergodicity and hybrid Markov chains*, Electron. Comm. Probab. 2 (1997), 13–25.

[RR97b] G. Roberts and J. Rosenthal, *Shift-coupling and convergence rates of ergodic averages*, Comm. Statist. Stochastic Models 13 (1997), 147–165.

[RR04] G. Roberts and J. Rosenthal, *General state space Markov chains and MCMC algorithms*, Probab. Surveys 1 (2004), 20–71.

[RR08] G. Roberts and J. Rosenthal, *Variance bounding Markov chains*, Ann. Appl. Probab. 18 (2008), 1201–1214.

[RT99] G. Roberts and R. Tweedie, *Bounds on regeneration times and convergence rates for Markov chains*, Stochastic Process. Appl. 80 (1999), 211–229.

[RT01] G. Roberts and R. Tweedie, *Geometric L^2 and L^1 convergence are equivalent for reversible Markov chains*, J. Appl. Probab. 38A (2001), 37–41.

[Ros95] J. Rosenthal, *Convergence rates for Markov chains*, SIAM Rev. 37 (1995), 387–405.

[Ros03] J. Rosenthal, *Asymptotic variance and convergence rates of nearly-periodic Markov chain Monte Carlo algorithms*, J. Amer. Statist. Assoc. 98 (2003), 169–177.

[Rud91] W. Rudin, *Functional Analysis*, 2nd ed., McGraw-Hill, New York, 1991.

[Rud07] D. Rudolf, *Monte-Carlo-Integration mit einer unbekannten Dichte*, Diplomarbeit, University Jena, 2007 (in German).

[Rud09] D. Rudolf, *Explicit error bounds for lazy reversible Markov chain Monte Carlo*, J. Complexity 25 (2009), 11–24.

[Rud10] D. Rudolf, *Error bounds of computing the expectation by Markov chain Monte Carlo*, Monte Carlo Methods Appl. 16 (2010), 323–342.

[SC04] L. Saloff-Coste, *Random Walks on Finite Groups*, Encyclopaedia Math. Sci. 110, Springer, Berlin, 2004.

[Sen06] E. Seneta, *Non-Negative Matrices and Markov Chains*, Springer, New York, 2006 (revised reprint of the second (1981) edition).

[Smi84] R. Smith, *Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions*, Oper. Res. 32 (1984), 1296–1308.

[Sok97] A. Sokal, *Monte Carlo methods in statistical mechanics: foundations and new algorithms*, in: Functional Integration (Cargèse, 1996), NATO Adv. Sci. Inst. Ser. B Phys. 361, Plenum, New York, 1997, 131–192.

[Str05] D. Stroock, *An Introduction to Markov Processes*, Grad. Texts in Math. 230, Springer, Berlin, 2005.

[Tri92] H. Triebel, *Higher Analysis*, Johann Ambrosius Barth, Leipzig, 1992.

[Vem05] S. Vempala, *Geometric random walks: A survey*, Combin. Comput. Geom. 52 (2005), 573–612.

[Yue00] W. Yuen, *Applications of geometric bounds to the convergence rate of Markov chains on $\mathbf{R}^n$*, Stochastic Process. Appl. 87 (2000), 1–23.