Discrepancy estimates for variance bounding Markov chain quasi-Monte Carlo

Josef Dick[*]   Daniel Rudolf[†]

Abstract

Markov chain Monte Carlo (MCMC) simulations are modeled as driven by true random numbers. We consider variance bounding Markov chains driven by a deterministic sequence of numbers. The star-discrepancy provides a measure of efficiency of such Markov chain quasi-Monte Carlo methods. We define a pull-back discrepancy of the driver sequence and state a close relation to the star-discrepancy of the Markov chain-quasi Monte Carlo samples. We prove that there exists a deterministic driver sequence such that the discrepancies decrease almost with the Monte Carlo rate $n^{-1/2}$. As for MCMC simulations, a burn-in period can also be taken into account for Markov chain quasi-Monte Carlo to reduce the influence of the initial state. In particular, our discrepancy bound leads to an estimate of the error for the computation of expectations. To illustrate our theory we provide an example for the Metropolis algorithm based on a ball walk. Furthermore, under additional assumptions we prove the existence of a driver sequence such that the discrepancy of the corresponding deterministic Markov chain sample decreases with order $n^{-1+\delta}$ for every $\delta > 0$.

1 Introduction

Markov chain Monte Carlo (MCMC) simulations are used in different branches of statistics and science to estimate an expected value with respect to a probability measure, say π, by the sample average of the Markov chain. This procedure is of advantage if random numbers with distribution π are difficult to construct.

When sampling the Markov chain the transitions are usually modeled as driven by i.i.d. $\mathcal{U}(0,1)^s$ random variables for some $s \geq 1$. But in simulations the driver sequences are pseudo-random numbers. In many applications, if one uses a carefully constructed random number generator, this works well. Instead of modeling the Markov chain with random numbers, or imitating random numbers, the idea of Markov chain quasi-Monte Carlo is to construct a finite, deterministic sequence of numbers, $(u_i)_{0 \leq i \leq n}$ in $[0,1]^s$ for

---

[*]The University of New South Wales, AU. E-mail: `josef.dick@unsw.edu.au`
[†]Friedrich Schiller University Jena, DE. E-mail: `daniel.rudolf@uni-jena.de`

all $n \in \mathbb{N}$, to generate a deterministic Markov chain sample and to use it to estimate the desired mean.

The motivation of this conceptual change is that carefully constructed sequences may lead to more accurate sample averages. For example, quasi-Monte Carlo (QMC) points lead to higher order of convergence compared to plain Monte Carlo, which is a special case of MCMC. Numerical experiments for QMC versions of MCMC also show promising results [17, 19, 23, 27, 30]. In particular, Owen and Tribble [23] and Tribble [30] report an improvement by a factor of up to 10^3 and a better convergence rate for a Gibbs sampler problem.

In the work of Chen, Dick and Owen [6] and Chen [5] the first theoretical justification for Markov chain quasi-Monte Carlo on continuous state spaces is provided. The authors show a consistency result if a contraction assumption is satisfied and the random sequence is substituted by a deterministic 'completely uniformly distributed' sequence, see [6, 7, 31]. Thus the sample average converges to the expected value but we do not know how fast this convergence takes place.

Recently, in [8] another idea appears. Namely, the question is considered whether there exists a good driver sequence such that an explicit error bound is satisfied. It is shown that if the Markov chain is uniformly ergodic, then for any initial state a deterministic sequence exists such that the sample average converges to the mean almost with the Monte Carlo rate.

However, in [6] and [8] rather strong conditions, the contraction assumption and uniform ergodicity, are imposed on the Markov chain. We substantially extend the results of [8] to Markov chains which satisfy a much weaker convergence condition. Namely, we consider variance bounding Markov chains, introduced by Roberts and Rosenthal in [24], and show existence results of good driver sequences. We also show what property pseudo-random number generators need to satisfy in order to improve the performance of Markov chain quasi-Monte Carlo algorithms, see Definition 3.1 below. This property has not been studied in the literature before. In the following we describe the setting in detail and explain our main contributions.

1.1 Main results

Let $(X_n)_{n \in \mathbb{N}}$ be a Markov chain with transition kernel K and initial distribution ν on $(G, \mathcal{B}(G))$ with $G \subseteq \mathbb{R}^d$. For $f \colon G \to \mathbb{R}$ let $\mathbb{E}_\pi(f) = \int_G f(x)\pi(\mathrm{d}x)$ be the desired mean and $Pf(x) = \int_G f(y)K(x,\mathrm{d}y)$ be the Markov operator induced by the transition kernel K. We assume that the transition kernel is reversible with respect to the distribution π and that it is variance bounding, see [24]. Roughly, a Markov chain is variance bounding if the asymptotic variances for functionals with unit stationary variance are uniformly bounded. Equivalent to this is the assumption that $\Lambda < 1$ with

$$\Lambda = \sup\{\lambda \in \mathrm{spec}(P - \mathbb{E}_\pi \mid L_2)\} \tag{1.1}$$

where $L_2 = L_2(\pi)$ is the Hilbert space of functions $f \colon G \to \mathbb{R}$ with

$$\|f\|_2 = \left(\int_G |f(x)|^2 \, \pi(\mathrm{d}x) \right)^{1/2} < \infty$$

and $\mathrm{spec}(P - \mathbb{E}_\pi \mid L_2)$ denotes the spectrum of $P - \mathbb{E}_\pi$ on L_2. Let us point out that the Markov chain does not need to be uniformly or geometrically ergodic. In fact, a variance bounding Markov chain may even be periodic. Hence the distribution of X_i, for i arbitrarily large, is not necessarily close to π.

Let $\varphi \colon G \times [0,1]^s \to G$ be an arbitrary *update function* of K and $\psi \colon [0,1]^s \to G$ be an arbitrary *generator function* of ν for some $s \in \mathbb{N}$. This means that the Markov chain

$(X_n)_{n \in \mathbb{N}}$ permits the representation

$$X_1 = \psi(U_0),$$
$$X_{i+1} = \varphi(X_i, U_i), \quad i \geq 1,$$

where $(U_n)_{n \in \mathbb{N}}$ are i.i.d. with $U_i \sim \mathcal{U}[0,1]^s$. Using a deterministic sequence $(u_i)_{i \geq 0}$ we generate the deterministic Markov chain $(x_i)_{i \geq 1}$ with $x_1 = \psi(u_0)$ and $x_{i+1} = \varphi(x_i, u_i)$ where $i \geq 1$. The efficiency of this procedure is measured by the star-discrepancy, a generalized Kolmogorov-Smirnov test, between the stationary measure π and the empirical distribution $\widehat{\pi}_n(A) = \frac{1}{n} \sum_{i=1}^{n} 1_{x_i \in A}$, where $1_{x_i \in A}$ is the indicator function of a set $A \subseteq G$. For a certain set $\mathscr{A}$ of subsets of G we define the star-discrepancy $D^*_{\mathscr{A}, \pi}$ of $S_n = \{x_1, \ldots, x_n\}$ as the supremum of $|\pi(A) - \widehat{\pi}_n(A)|$ over all $A \in \mathscr{A}$, i.e.

$$D^*_{\mathscr{A}, \pi}(S_n) = \sup_{A \in \mathscr{A}} |\widehat{\pi}_n(A) - \pi(A)|.$$

By inverting the iterates of the update function we also define a pull-back discrepancy of the driver sequence (the test sets are pulled back). We show that for large $n \in \mathbb{N}$ both discrepancies are close to each other.

The main result, in a general setting, is an estimate of $D^*_{\mathscr{A}, \pi}(S_n)$ (Theorem 4.5) under the assumption that we have an approximation of $\mathscr{A}$, for any $\delta > 0$, given by a so-called δ-cover Γ_δ of $\mathscr{A}$ with respect to π (Definition 4.1). The proof of the main result is based on a Hoeffding inequality for Markov chains. After that we prove that a sufficiently good δ-cover exists if π is absolutely continuous with respect to the Lebesgue measure and the set of test sets is the set of open boxes restricted to G anchored at $-\infty$, i.e. we consider the set of test sets

$$\mathscr{B} = \{(-\infty, x) \cap G \colon x \in \mathbb{R}^d\},$$

with $(-\infty, x) = \Pi_{i=1}^{d}(-\infty, x_i)$. By the Koksma-Hlawka inequality (Theorem 5.5) we have

$$\left| \mathbb{E}_\pi(f) - \frac{1}{n} \sum_{i=1}^{n} f(x_i) \right| \leq \|f\|_{H_1} D^*_{\mathscr{B}, \pi}(S_n),$$

with $\|f\|_{H_1}$ defined in (5.5) below. Thus a bound on the discrepancy leads to an error bound for the approximation of $\mathbb{E}_\pi(f)$.

We show for any update function $\varphi \colon G \times [0,1]^s \to G$ of K, any generator function $\psi \colon [0,1]^s \to G$ of ν, and for all $n \geq 16$ that there exists a driver sequence $u_0, \ldots, u_{n-1} \in [0,1]^s$ such that $S_n = \{x_1, \ldots, x_n\}$ given by

$$x_1 = \psi(u_0)$$
$$x_{i+1} = \varphi(x_i; u_i), \quad i = 1, \ldots, n-1,$$

satisfies

$$D^*_{\mathscr{B}, \pi}(S_n) \leq \sqrt{\frac{1 + \Lambda_0}{1 - \Lambda_0}} \cdot \frac{\sqrt{2} \left(\log \left\| \frac{d\nu}{d\pi} \right\|_2 + d \log n + 3d^2 \log(5d) \right)^{1/2}}{\sqrt{n}} + \frac{8}{n^{3/4}}, \tag{1.2}$$

where $\frac{d\nu}{d\pi}$ is the density of $\nu = \mathbb{P}_\psi$ (the probability measure induced by ψ) with respect to π and $\Lambda_0 = \max\{\Lambda, 0\}$ with Λ defined in (1.1). For the details we refer to Corollary 5.4 below. This implies, by the Koksma-Hlawka inequality, that the sample average converges to the mean with $\mathcal{O}(n^{-1/2}(\log n)^{1/2})$.

Additionally we might take a burn-in period of n_0 steps into account to reduce the dependence of the initial state in the discrepancy bound. Roughly, the idea is to generate a sequence $x_1, \ldots, x_{n_0+n}$ by the Markov chain quasi-Monte Carlo procedure and to

consider the discrepancy of the point set $S_{[n_0,n]} = \{x_{n_0+1}, \ldots, x_{n_0+n}\}$. Under suitable convergence conditions on the Markov chain, for example the existence of an absolute L_2-spectral gap (see Definition 2.1), the density $\frac{d(\nu P^{n_0})}{d\pi}$ is close to 1, see Subsection 4.3.

If we further assume that one can reach every state from every other state within one step of the Markov chain, then we prove that there exists a driver sequence such that the discrepancy converges with $\mathcal{O}(n^{-1}(\log_2 n)^{(3d+1)/2})$. We call the additional assumption 'anywhere-to-anywhere' condition. The result shows that in principle a higher order of convergence for Markov chain quasi-Monte Carlo is possible. Note that, many well studied Markov chains satisfy such a condition, for example the hit-and-run algorithm, the independent Metropolis sampler or the slice sampler, see for example [20]. From our work it is not immediately clear how to obtain suitable driver sequences which yield such an improvement. However, what our results here show is that the main quality criterion is the pull-back discrepancy (see Definition 3.1 below) of the driver sequence. Since this has previously not been known, the pull-back discrepancy of explicit constructions of quasi-Monte Carlo point sets or pseudo-random number generators has not been studied so far. The task of future work is therefore to explicitly construct point sets with small pull-back discrepancy. We leave it is an interesting and challenging problem for further research. We provide an outline of our work in the following.

1.2 Outline

In the next section the necessary background information on Markov chains is stated. Section 3 is devoted to the study of the relation of the discrepancies. The Monte Carlo rate of convergence for Markov chain-quasi Monte Carlo is shown in Section 4. There we also provide results for the case when a burn-in period is taken into account. Section 5 deals with the set of test sets which consists of axis parallel boxes, see $\mathscr{B}$ above. We show the existence of a good δ-cover and how the discrepancy bounds can be used to obtain bounds on the error for the computation of expected values of smooth functions. This yields a Koksma-Hlawka inequality for Markov chains. To illustrate our results, we provide an example of a Metropolis algorithm with ball walk proposal on the Euclidean unit ball. A special situation arises when the update function of the Markov chain has an 'anywhere-to-anywhere' property, see Section 6. In this situation we show that a convergence rate of order almost n^{-1} can be obtained.

2 Background on Markov chains

Let $G \subseteq \mathbb{R}^d$ and let $\mathcal{B}(G)$ denote the Borel σ-algebra of G. In the following we provide a brief introduction to Markov chains on $(G, \mathcal{B}(G))$. We assume that $K : G \times \mathcal{B}(G) \to [0,1]$ is a transition kernel on $(G, \mathcal{B}(G))$, i.e. for each $x \in G$ the mapping $A \in \mathcal{B}(G) \mapsto K(x,A)$ is a probability measure and for each $A \in \mathcal{B}(G)$ the mapping $x \in G \mapsto K(x,A)$ is a $\mathcal{B}(G)$-measurable real-valued function. Further let ν be a probability measure on $(G, \mathcal{B}(G))$.

Then let $(X_n)_{n \in \mathbb{N}}$, with X_n mapping from some probability space into $(G, \mathcal{B}(G))$, be a Markov chain with transition kernel K and initial distribution ν. This might be interpreted as follows: Let $X_1 = x_1 \in G$ be chosen with ν on $(G, \mathcal{B}(G))$ and let $i \in \mathbb{N}$. Then for a given $X_i = x_i$, the random variable X_{i+1} has distribution $K(x_i, \cdot)$, that is, for all $A \in \mathcal{B}(G)$, the probability that $X_{i+1} \in A$ is given by $K(x_i, A)$.

Let π be a probability measure on $(G, \mathcal{B}(G))$. We assume that the transition kernel K is *reversible with respect to π*, i.e.

$$\int_A K(x,B)\,\pi(\mathrm{d}x) = \int_B K(x,A)\,\pi(\mathrm{d}x)$$

for all $A, B \in \mathcal{B}(G)$. This implies that π is a *stationary distribution* of the transition

kernel K, i.e.

$$\int_G K(x, A)\, \pi(\mathrm{d}x) = \pi(A) \tag{2.1}$$

for all $A \in \mathcal{B}(G)$. We assume that the stationary distribution π is unique. Let $L_2 = L_2(\pi)$ be the set of all functions $f \colon G \to \mathbb{R}$ with

$$\|f\|_2 = \left(\int_G |f(x)|^2\, \pi(\mathrm{d}x) \right)^{1/2} < \infty.$$

The transition kernel K induces an operator acting on functions and an operator acting on measures. For $x \in G$ and $A \in \mathcal{B}(G)$ the operators are given by

$$P f(x) = \int_G f(y)\, K(x, \mathrm{d}y), \quad \text{and} \quad \nu P(A) = \int_G K(x, A)\, \nu(\mathrm{d}x),$$

where $f \in L_2$ and ν is a signed measure on $(G, \mathcal{B}(G))$ with a density $\frac{d\nu}{d\pi} \in L_2$. By the reversibility with respect to π we have that $P \colon L_2 \to L_2$ is self-adjoint and π-almost everywhere holds $P(\frac{d\nu}{d\pi})(x) = \frac{d(\nu P)}{d\pi}(x)$. For details we refer to [26].

In the following we introduce two convergence properties of transition kernels. Let the expectation with respect to π be denoted by $\mathbb{E}_\pi(f) = \int_G f(y)\pi(\mathrm{d}x)$. Let $L_2^0 = \{f \in L_2 \colon \mathbb{E}_\pi(f) = 0\}$ and note that L_2^0 is a closed subspace of L_2. We have

$$\|P - \mathbb{E}_\pi\|_{L_2 \to L_2} = \|P\|_{L_2^0 \to L_2^0} = \sup_{f \in L_2^0,\, \|f\|_2 \le 1} \|P f\|_2,$$

for details see [26, Lemma 3.16, p. 44].

Definition 2.1 (absolute L_2-spectral gap). *We say that a transition kernel K, and its corresponding Markov operator P, has an* absolute L_2-spectral gap *if*

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

and the absolute spectral gap is $1 - \beta$.

Let us introduce the *total variation distance* of two probability measures ν_1, ν_2 on $(G, \mathcal{B}(G))$ by

$$\|\nu_1 - \nu_2\|_{\mathrm{tv}} = \sup_{A \in \mathcal{B}(G)} |\nu_1(A) - \nu_2(A)|.$$

Note that for a Markov chain $(X_n)_{n \in \mathbb{N}}$ with transition kernel K and initial distribution ν holds $\mathbb{P}_{\nu, K}(X_n \in A) = \nu P^{n-1}(A)$, where ν and K in $\mathbb{P}_{\nu, K}$ indicate the initial distribution and transition kernel. Then we obtain the following relation between the absolute L_2-spectral gap and the total variation distance. The result is an application of [26, Corollary 3.15 and Lemma 3.21].

Proposition 2.2. *Let ν be a distribution on $(G, \mathcal{B}(G))$ and assume that there exists a density $\frac{d\nu}{d\pi} \in L_2$. Then*

$$\|\nu P^n - \pi\|_{\mathrm{tv}} \le \beta^n \left\| \frac{d\nu}{d\pi} - 1 \right\|_2, \quad n \in \mathbb{N},$$

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

The next convergence property is weaker than the existence of an absolute spectral gap.

Definition 2.3 (Variance bounding or L_2-spectral gap). *We say that a reversible transition kernel K, and its corresponding Markov operator P, is* variance bounding *or has an* L_2-spectral gap *if*

$$\Lambda = \sup\{\lambda \in \operatorname{spec}(P \mid L_2^0)\} < 1, \tag{2.2}$$

where $\operatorname{spec}(P \mid L_2^0)$ denotes the spectrum of $P \colon L_2^0 \to L_2^0$.

For a motivation of the term variance bounding and a general treatment we refer to [24]. In particular, by [24, Theorem 14] under the assumption of reversibility our definition is equivalent to the one stated by Roberts and Rosenthal. Note that the existence of an absolute L_2-spectral gap implies variance bounding, since

$$\|P\|_{L_2^0 \to L_2^0} = \sup_{\lambda \in \mathrm{spec}(P|L_2^0)} |\lambda|.$$

We have the following relation between variance bounding and the total variation distance.

Lemma 2.4. *Let the transition kernel K be reversible with respect to π and let $n \in \mathbb{N}$ with $n \geq 2$. Further, let P be variance bounding. Then the Markov operator $P_n = \frac{1}{n} \sum_{j=0}^{n-1} P^j$ has an absolute L_2-spectral gap. In particular, if ν is a distribution on $(G, \mathcal{B}(G))$ with $\frac{d\nu}{d\pi} \in L_2$, then*

$$\|\nu P_n - \pi\|_{\mathrm{tv}} \leq \frac{1 - \Lambda_0^n}{n \cdot (1 - \Lambda_0)} \left\| \frac{d\nu}{d\pi} - 1 \right\|_2,$$

with $\Lambda_0 = \max\{0, \Lambda\}$, see (2.2).

Proof. By the spectral theorem for bounded self-adjoint operators we have for a polynomial $F \colon \mathrm{spec}(P \mid L_2^0) \to \mathbb{R}$ that

$$\|F(P)\|_{L_2^0 \to L_2^0} = \max_{\alpha \in \mathrm{spec}(P|L_2^0)} |F(\alpha)|.$$

For details see for example [25] or [15, Theorem 9.9-2]. In our case $F(\lambda) = \frac{1}{n} \sum_{i=0}^{n-1} \lambda^i$ so that $F(P) = \frac{1}{n} \sum_{i=0}^{n-1} P^i$. Thus

$$\left\| \frac{1}{n} \sum_{i=0}^{n-1} P^i \right\|_{L_2^0 \to L_2^0} = \max_{\lambda \in \mathrm{spec}(P|L_2^0)} \left| \frac{1 - \lambda^n}{n \cdot (1 - \lambda)} \right| \leq \frac{1 - \Lambda_0^n}{n \cdot (1 - \Lambda_0)}.$$

The last inequality is proven by $\mathrm{spec}(P|L_2^0) \subseteq [-1, 1]$ and the following facts: For $\lambda \in [-1, 0]$ holds $\frac{1-\lambda^n}{n \cdot (1-\lambda)} \leq \frac{1}{n}$ and for $\lambda \in [0, 1]$ the function $\frac{1-\lambda^n}{n \cdot (1-\lambda)} = \frac{1}{n} \sum_{j=0}^{n-1} \lambda^j$ is increasing. The estimate of the total variation distance follows by Proposition 2.2. $\square$

The next part deals with an update function, say φ, of a given transition kernel K. We state the crucial properties of the transition kernel in terms of an update function. This is partially based on [8].

Definition 2.5 (Update function)**.** *Let $\varphi : G \times [0,1]^s \to G$ be a measurable function and*

$$B : G \times \mathcal{B}(G) \to \mathcal{B}([0,1]^s),$$
$$B(x, A) = \{u \in [0,1]^s : \varphi(x; u) \in A\}.$$

Let λ_s denote the Lebesgue measure on $\mathbb{R}^s$. Then the function φ is an update function for the transition kernel K if and only if

$$K(x, A) = \mathbb{P}(\varphi(x; U) \in A) = \lambda_s(B(x, A)), \tag{2.3}$$

where $\mathbb{P}$ is the probability measure for the uniform distribution in $[0,1]^s$.

Note that for any transition kernel on $(G, \mathcal{B}(G))$ there exists an update function with $s = 1$, see for example [14, Lemma 2.22, p. 34]. For $x \in G$ and $A \in \mathcal{B}(G)$ the set $B(x, A)$ is the set of all random numbers $u \in [0,1]^s$ which take x into the set A using the update function φ with arguments x and u.

We consider the iterated application of an update function. Let $\varphi_1(x; u) = \varphi(x; u)$ and for $i > 1$ with $i \in \mathbb{N}$ let

$$\varphi_i : G \times [0, 1]^{is} \to G,$$

$$\varphi_i(x; u_1, u_2, \dots, u_i) = \varphi(\varphi_{i-1}(x; u_1, u_2, \dots, u_{i-1}); u_i).$$

Thus, $x_{i+1} = \varphi_i(x; u_1, u_2, \dots, u_i) \in G$ is the point obtained via i updates using the sequence $u_1, u_2, \dots, u_i \in [0, 1]^s$, where the starting point is $x \in G$.

Lemma 2.6. *Let $i, j \in \mathbb{N}$ and $i \geq j$. For any $u_1, \dots, u_i \in [0, 1]^s$ and $x \in G$ we have*

$$\varphi_i(x; u_1, \dots, u_i) = \varphi_{i-j}(\varphi_j(x; u_1, \dots, u_j); u_{j+1}, \dots, u_i). \tag{2.4}$$

Proof. The proof follows by induction on i. $\qquad\qquad\square$

For $i \geq 1$ with $i \in \mathbb{N}$ let

$$B_i : G \times \mathcal{B}(G) \to \mathcal{B}([0, 1]^{is}),$$

$$B_i(x, A) = \{(u_1, u_2, \dots, u_i) \in [0, 1]^{is} : \varphi_i(x; u_1, u_2, \dots, u_i) \in A\}.$$

Note that $B_1(x, A) = B(x, A)$. For $x \in G$ and $A \in \mathcal{B}(G)$ the set $B_i(x, A)$ is the set of all random numbers $u_1, u_2, \dots, u_i \in [0, 1]^s$ which take x into the set A after the ith iteration of the update function φ, i.e. φ_i with arguments x and $u_1, u_2, \dots, u_i$.

In [8] we considered the case where the initial state is deterministically chosen. The following definition is useful to work with general initial distributions.

Definition 2.7. *For a probability measure ν on $(G, \mathcal{B}(G))$ we call a measurable function $\psi\colon [0, 1]^s \to G$ generator function if*

$$\nu(A) = \mathbb{P}(\psi(U) \in A), \quad A \in \mathcal{B}(G),$$

where $\mathbb{P}$ is the uniform distribution in $[0, 1]^s$.

Let ν be a probability measure on $(G, \mathcal{B}(G))$ and let $\psi\colon [0, 1]^s \to G$ be its generator function.

Then, for $i \geq 1$ with $i \in \mathbb{N}$ and $A \in \mathcal{B}(G)$, let

$$\begin{aligned} C_{i,\psi}(A) &= \{(u_0, u_1, \dots, u_i) \in [0, 1]^{(i+1)s} : \varphi_i(\psi(u_0); u_1, \dots, u_i) \in A\} \\ &= \{(u_0, u_1, \dots, u_i) \in [0, 1]^{(i+1)s} : (u_1, \dots, u_i) \in B_i(\psi(u_0), A)\} \end{aligned} \tag{2.5}$$

and $C_{0,\psi}(A) = \{u_0 \in [0, 1]^s : \psi(u_0) \in A\}$. The set $C_{i,\psi}(A) \subseteq [0, 1]^{(i+1)s}$ is the set of possible sequences to get into the set A with starting point $\psi(u_0)$ and i updates of the update function.

The next lemma is important to understand the relation between the update function, generator function, transition kernel and initial distribution.

Lemma 2.8. *Let K be a transition kernel and ν be a distribution on $(G, \mathcal{B}(G))$. Let $(X_n)_{n \in \mathbb{N}}$ be a Markov chain with transition kernel K and initial distribution ν. Let us assume that $i \in \mathbb{N}$ and $F\colon G^i \to \mathbb{R}$. Then, for any update function $\varphi\colon G \times [0, 1]^s \to G$ of the transition kernel K and any generator function $\psi\colon [0, 1]^s \to G$ of ν the expectation of F with respect to the joint distribution of $X_1, \dots, X_i$ satisfies*

$$\begin{aligned} &\mathbb{E}_{\nu,K}(F(X_1, \dots, X_i)) \\ &= \int_{[0,1]^{is}} F(\psi(u_0), \varphi_1(\psi(u_0), u_1), \dots, \varphi_{i-1}(\psi(u_0), u_1, \dots, u_{i-1})) \\ &\qquad\qquad \times \mathrm{d}u_0 \, \mathrm{d}u_1 \dots \mathrm{d}u_{i-1}, \end{aligned} \tag{2.6}$$

whenever one of the integrals exist.

Proof. First, note that

$$\mathbb{E}_{\nu,K}(F(X_1,\ldots,X_i))$$
$$= \underbrace{\int_G \ldots \int_G}_{i\text{-times}} F(x_1,\ldots,x_i)\,K(x_{i-1},\mathrm{d}x_i)\ldots K(x_1,\mathrm{d}x_2)\,\nu(\mathrm{d}x_1).$$

By the fact that ψ is a generator function of ν we have

$$\int_{[0,1]^{is}} F(\psi(u_0),\varphi_1(\psi(u_0),u_1),\ldots,\varphi_{i-1}(\psi(u_0),u_1,\ldots,u_{i-1}))\,\mathrm{d}u_0\,\mathrm{d}u_1\ldots\mathrm{d}u_{i-1}$$
$$= \int_G \int_{[0,1]^{(i-1)s}} F(x_1,\varphi_1(x_1,u_1),\ldots,\varphi_{i-1}(x_1,u_1,\ldots,u_{i-1}))\,\mathrm{d}u_1\ldots\mathrm{d}u_{i-1}\,\nu(\mathrm{d}x_1),$$

and by Lemma 2.6 we obtain

$$\int_G \int_{[0,1]^{(i-1)s}} F(x_1,\varphi_1(x_1,u_1),\ldots,\varphi_{i-1}(x_1,u_1,\ldots,u_{i-1}))\,\mathrm{d}u_1\ldots\mathrm{d}u_{i-1}\,\nu(\mathrm{d}x_1)$$
$$= \int_G \int_G \int_{[0,1]^{(i-2)s}} F(x_1,x_2,\varphi_1(x_2,u_2),\ldots,\varphi_{i-1}(x_2,u_2,\ldots,u_{i-1}))$$
$$\times \mathrm{d}u_2\ldots\mathrm{d}u_{i-1}\,K(x_1,\mathrm{d}x_2)\,\nu(\mathrm{d}x_1).$$

By iterating the application of Lemma 2.6 the assertion is proven. $\square$

Note that the right-hand-side of (2.6) is the expectation with respect to the uniform distribution in $[0,1]^{is}$.

Corollary 2.9. *Assume that the conditions of Lemma 2.8 are satisfied. Then, for $A \in \mathcal{B}(G)$, we have*
$$\nu P^i(A) = \lambda_{(i+1)s}(C_{i,\psi}(A)), \tag{2.7}$$
and $\nu P^0(A) = \nu(A) = \lambda_s(C_{0,\psi}(A))$.

Proof. By Lemma 2.8 we have

$$\nu P^i(A) = \int_G K^i(x,A)\,\nu(\mathrm{d}x)$$
$$= \int_G \underbrace{\int_G \ldots \int_G}_{i\text{-times}} 1_{x_{i+1}\in A}\,K(x_i,\mathrm{d}x_{i+1})\ldots K(x_1,\mathrm{d}x_2)\,\nu(\mathrm{d}x_1)$$
$$= \int_{[0,1]^{(i+1)s}} 1_{\varphi_i(\psi(u_0),u_1,\ldots,u_i)\in A}\,\mathrm{d}u_0\,\mathrm{d}u_1\ldots\mathrm{d}u_i$$
$$= \int_{[0,1]^{(i+1)s}} 1_{(u_0,u_1,\ldots,u_i)\in C_{i,\psi}(A)}\,\mathrm{d}u_0\,\mathrm{d}u_1\ldots\mathrm{d}u_i = \lambda_{(i+1)s}(C_{i,\psi}(A)),$$

which completes the proof. $\square$

3 On the pull-back discrepancy

Let $\mathscr{A} \subseteq \mathcal{B}(G)$ be a set of test sets. Then the star-discrepancy of a point set $S_n = \{x_1,\ldots,x_n\} \subseteq G$ with respect to the distribution π is given by

$$D^*_{\mathscr{A},\pi}(S_n) = \sup_{A\in\mathscr{A}} \left| \frac{1}{n}\sum_{i=1}^n 1_{x_i\in A} - \pi(A) \right|.$$

Assume that $u_0, u_1, \ldots, u_{n-1} \in [0,1]^s$ is a finite deterministic sequence. We call this finite sequence *driver sequence*. Further, let $\varphi \colon G \times [0,1]^s \to G$ and $\psi \colon [0,1]^s \to G$ be measurable functions. Then let $S_n = \{x_1, \ldots, x_n\} \subseteq G$ be given by

$$x_{i+1} = x_{i+1}(x_1) = \varphi(x_i; u_i) = \varphi_i(x_1; u_1, \ldots, u_i), \quad i = 1, \ldots, n-1, \qquad (3.1)$$

where $x_1 = \psi(u_0)$. Note that ψ might be considered a generator function and φ might be considered an update function. We now define a discrepancy measure on the driver sequence. We call it *pull-back discrepancy*. Below we show how this pull-back discrepancy is related to the star-discrepancy of S_n.

Definition 3.1 (Pull-back discrepancy). *Let* $\mathcal{U}_n = \{u_0, u_1, \ldots, u_{n-1}\} \subset [0,1]^s$ *and let* $C_{i,\psi}(A)$ *for* $A \in \mathcal{B}(G)$ *and* $i \in \mathbb{N} \cup \{0\}$ *be defined as in (2.5). Define the local discrepancy function by*

$$\Delta_{n,A,\psi,\varphi}^{\mathrm{loc}}(\mathcal{U}_n) = \frac{1}{n} \sum_{i=0}^{n-1} \left[1_{(u_0,\ldots,u_i) \in C_{i,\psi}(A)} - \lambda_{(i+1)s}(C_{i,\psi}(A)) \right].$$

Let $\mathscr{A} \subseteq \mathcal{B}(G)$ *be a set of test sets. Then we define the discrepancy of the driver sequence by*

$$D_{\mathscr{A},\psi,\varphi}^{*}(\mathcal{U}_n) = \sup_{A \in \mathscr{A}} \left| \Delta_{n,A,\psi,\varphi}^{\mathrm{loc}}(\mathcal{U}_n) \right|.$$

We call $D_{\mathscr{A},\psi,\varphi}^{*}(\mathcal{U}_n)$ *pull-back discrepancy of* $\mathcal{U}_n$.

The discrepancy of the driver sequence $D_{\mathscr{A},\psi,\varphi}^{*}(\mathcal{U}_n)$ is a 'pull-back discrepancy' since the test sets $C_{i,\psi}(A)$ are derived from the test sets $A \in \mathscr{A}$ from the star-discrepancy $D_{\mathscr{A},\pi}^{*}(S_n)$ via inverting the update function and the generator.

The following theorem provides a relation between the star-discrepancy of S_n and the pull-back discrepancy of $\mathcal{U}_n$, this is similar to [8, Theorem 1].

Theorem 3.2. *Let* K *be a transition kernel and* ν *be a distribution on* $(G, \mathcal{B}(G))$. *Let* $\mathscr{A} \subseteq \mathcal{B}(G)$ *be a set of test sets. Then, for any update function* $\varphi \colon G \times [0,1]^s \to G$ *of* K *and any generator function* $\psi \colon [0,1]^s \to G$ *of* ν *we have, with driver sequence* $\mathcal{U}_n = \{u_0, u_1, \ldots, u_{n-1}\} \subset [0,1]^s$ *and* S_n *given by (3.1), that*

$$\left| D_{\mathscr{A},\pi}^{*}(S_n) - D_{\mathscr{A},\psi,\varphi}^{*}(\mathcal{U}_n) \right| \leq \sup_{A \in \mathscr{A}} \left| \frac{1}{n} \sum_{i=0}^{n-1} \nu P^i(A) - \pi(A) \right|.$$

Proof. For any $A \in \mathscr{A}$ we have by (2.7) that $\lambda_{(i+1)s}(C_{i,\psi}(A)) = \nu P^i(A)$. Thus

$$\left| \frac{1}{n} \sum_{i=1}^{n} 1_{x_i \in A} - \pi(A) \right|$$

$$= \left| \frac{1}{n} \sum_{i=0}^{n-1} \left[1_{(u_0,\ldots,u_i) \in C_{i,\psi}(A)} - \nu P^i(A) + \nu P^i(A) - \pi(A) \right] \right|$$

$$\leq \left| \frac{1}{n} \sum_{i=0}^{n-1} \left[1_{(u_0,\ldots,u_i) \in C_{i,\psi}(A)} - \lambda_{(i+1)s}(C_{i,\psi}(A)) \right] \right| + \left| \frac{1}{n} \sum_{i=0}^{n-1} \nu P^i(A) - \pi(A) \right|.$$

Hence

$$D_{\mathscr{A},\pi}^{*}(S_n) \leq D_{\mathscr{A},\psi,\varphi}^{*}(\mathcal{U}_n) + \sup_{A \in \mathscr{A}} \left| \frac{1}{n} \sum_{i=0}^{n-1} \nu P^i(A) - \pi(A) \right|.$$

The inequality

$$D_{\mathscr{A},\psi,\varphi}^{*}(\mathcal{U}_n) \leq D_{\mathscr{A},\pi}^{*}(S_n) + \sup_{A \in \mathscr{A}} \left| \frac{1}{n} \sum_{i=0}^{n-1} \nu P^i(A) - \pi(A) \right|$$

follows by the same arguments. $\qquad\qquad\square$

Corollary 3.3. *Assume that the conditions of Theorem 3.2 are satisfied. By P denote the Markov operator of K. Further, let K be reversible with respect to π, let P be variance bounding and let $\frac{d\nu}{d\pi} \in L_2$. Then, for any update function $\varphi \colon G \times [0,1]^s \to G$ of K and any generator function $\psi \colon [0,1]^s \to G$ of ν we have, with driver sequence $\mathcal{U}_n = \{u_0, u_1, \ldots, u_{n-1}\} \subset [0,1]^s$ and S_n given by (3.1), that*

$$\left| D^*_{\mathscr{A},\pi}(S_n) - D^*_{\mathscr{A},\psi,\varphi}(\mathcal{U}_n) \right| \leq \frac{1 - \Lambda_0^n}{n \cdot (1 - \Lambda_0)} \left\| \frac{d\nu}{d\pi} - 1 \right\|_2,$$

where $\Lambda_0 = \max\{0, \Lambda\}$ and Λ is defined in (2.2).

Proof. With $P_n = \frac{1}{n} \sum_{i=0}^{n-1} P^i$ we have

$$\sup_{A \in \mathscr{A}} \left| \frac{1}{n} \sum_{i=0}^{n-1} \nu P^i(A) - \pi(A) \right| \leq \| \nu P_n - \pi \|_{\mathrm{tv}}.$$

Thus, the assertion follows by Lemma 2.4 and Theorem 3.2. $\qquad\square$

Remark 3.4. For the moment let us assume that we can sample with respect to π. For any initial distribution ν with $\frac{d\nu}{d\pi} \in L_2$, for all $x \in G$ and $A \in \mathcal{B}(G)$ we set $K(x, A) = \pi(A)$, hence $\Lambda = 0$. Thus, for any update function φ of K and generator function ψ of ν we have

$$\left| D^*_{\mathscr{A},\pi}(S_n) - D^*_{\mathscr{A},\psi,\varphi}(\mathcal{U}_n) \right| \leq \frac{1}{n} \left\| \frac{d\nu}{d\pi} - 1 \right\|_2.$$

Note that the discrepancies do not coincide. The reason for this is that the initial state is taken into account in the average computation. However, if $\nu = \pi$, then for any reversible transition kernel with respect to π we obtain $D^*_{\mathscr{A},\pi}(P_n) = D^*_{\mathscr{A},\psi,\varphi}(\mathcal{U}_n)$.

4 Monte Carlo rate of convergence

In this section we show for any update function $\varphi \colon G \times [0,1]^s \to G$ of a variance bounding transition kernel K and any generator function $\psi \colon [0,1]^s \to G$ of a distribution ν the existence of finite sequences $\mathcal{U}_n = \{u_0, u_1, \ldots, u_{n-1}\} \subset [0,1]^s$, which define S_n by (3.1), such that

$$D^*_{\mathscr{A},\psi,\varphi}(\mathcal{U}_n) \quad \text{and} \quad D^*_{\mathscr{A},\pi}(S_n)$$

converge to 0 approximately with order $n^{-1/2}$. The main result is proven for $D^*_{\mathscr{A},\pi}(S_n)$. The result with respect to $D^*_{\mathscr{A},\psi,\varphi}(\mathcal{U}_n)$ holds by Theorem 3.2.

4.1 Useful tools: delta-cover and Hoeffding inequality

The concept of a δ-cover will be useful (cf. [11] for a discussion of δ-covers, bracketing numbers and Vapnik-Červonenkis dimension).

Definition 4.1. *Let $\mathscr{A} \subseteq \mathcal{B}(G)$ be a set of test sets. A finite subset $\Gamma_\delta \subseteq \mathcal{B}(G)$ is called a δ-cover of $\mathscr{A}$ with respect to π if for every $A \in \mathscr{A}$ there are sets $C, D \in \Gamma_\delta$ such that*

$$C \subseteq A \subseteq D$$

and

$$\pi(D \setminus C) \leq \delta.$$

We assume that $\emptyset \in \Gamma_\delta$.

The following result is well known for the uniform distribution, see [13, Section 2.1] (see also [8, Remark 3] for the particular case below).

Proposition 4.2. *Let $\mathscr{A} \subseteq \mathcal{B}(G)$ be a set of test sets. Let Γ_δ be a δ-cover of $\mathscr{A}$ with respect to π. Then, for any point set $Z_n = \{z_1, \ldots, z_n\} \subseteq G$, we have*

$$D^*_{\mathscr{A},\pi}(Z_n) \leq \max_{C \in \Gamma_\delta} \left| \frac{1}{n} \sum_{i=1}^n 1_{z_i \in C} - \pi(C) \right| + \delta.$$

Instead of considering the supremum over the possibly infinite set of test sets $\mathscr{A}$ in the star-discrepancy we use a finite set Γ_δ and take the maximum over $C \in \Gamma_\delta$ by paying the price of adding δ.

For variance bounding Markov chains on discrete state spaces a Hoeffding inequality is proven in [18]. In [22] this is extended to non-reversible Markov chains on general state spaces. The following Hoeffding inequality for reversible, variance bounding Markov chains follows by [22, Theorem 3.3 and the remark after (3.4)].

Proposition 4.3 (Hoeffding inequality for Markov chains)**.** *Let K be a reversible transition kernel with respect π and let ν be a distribution on $(G, \mathcal{B}(G))$ with $\frac{d\nu}{d\pi} \in L_2$. Let us assume that the Markov operator of K is variance bounding. Further, let $(X_n)_{n \in \mathbb{N}}$ be a Markov chain with transition kernel K and initial distribution ν. Then, for any $A \in \mathcal{B}(G)$ and $c > 0$, we obtain*

$$\mathbb{P}_{\nu,K}\left[\left| \frac{1}{n} \sum_{i=1}^n 1_{X_i \in A} - \pi(A) \right| \geq c \right] \leq 2 \left\| \frac{d\nu}{d\pi} \right\|_2 \exp\left(-\frac{1-\Lambda_0}{1+\Lambda_0} c^2 n \right), \tag{4.1}$$

with $\Lambda_0 = \max\{0, \Lambda\}$ and where Λ is defined in (2.2).

We provide a lemma to state the Hoeffding inequality for Markov chains in terms of the driver sequence. To do so, let $\varphi\colon G \times [0,1]^s \to G$ and $\psi\colon [0,1]^s \to G$. We need the following notation. Let $\Delta_{n,A,\varphi,\psi}\colon [0,1]^{ns} \to [-1,1]$ be given by

$$\Delta_{n,A,\varphi,\psi}(u_0, \ldots, u_{n-1}) = \frac{1}{n} \sum_{i=0}^{n-1} \left[1_{(u_0,\ldots,u_i) \in C_{i,\psi}(A)} - \pi(A) \right]. \tag{4.2}$$

Lemma 4.4. *Let K be a transition kernel and ν be a distribution on $(G, \mathcal{B}(G))$. Let $(X_n)_{n \in \mathbb{N}}$ be a Markov chain with transition kernel K and initial distribution ν. Then, for any update function $\varphi\colon G \times [0,1]^s \to G$ of K, any generator function $\psi\colon [0,1]^s \to G$, any $A \in \mathcal{B}(G)$ and $c > 0$, we have*

$$\mathbb{P}[|\Delta_{n,A,\varphi,\psi}| \geq c] = \mathbb{P}_{\nu,K}\left[\left| \frac{1}{n} \sum_{i=1}^n 1_{X_i \in A} - \pi(A) \right| \geq c \right], \tag{4.3}$$

where $\mathbb{P}$ denotes the uniform distribution in $[0,1]^{ns}$ and $\mathbb{P}_{\nu,K}$ denotes the joint distribution of $X_1, \ldots, X_n$.

Proof. Let $J(A,c) = \left\{ (z_1, \ldots, z_n) \in G^n \colon \left| \frac{1}{n} \sum_{i=1}^n 1_{z_i \in A} - \pi(A) \right| \geq c \right\}$ and let

$$F(x_1, \ldots, x_n) = 1_{(x_1,\ldots,x_n) \in J(A,c)} = \begin{cases} 1 & \left| \frac{1}{n} \sum_{i=1}^n 1_{x_i \in A} - \pi(A) \right| \geq c, \\ 0 & \text{otherwise.} \end{cases}$$

By $\mathbb{E}_{\nu,K}(F(X_1, \ldots, X_n)) = \mathbb{P}_{\nu,K}(J(A,c))$, Lemma 2.8 and

$$1_{(\psi(u_0),\varphi_1(\psi(u_0),u_1),\ldots,\varphi_{n-1}(\psi(u_0),u_1,\ldots,u_{n-1})) \in J(A,c)}$$
$$= \begin{cases} 1 & \left| \frac{1}{n} \sum_{i=0}^{n-1} \left[1_{(u_0,\ldots,u_i) \in C_{i,\psi}(A)} - \pi(A) \right] \right| \geq c, \\ 0 & \text{otherwise,} \end{cases}$$

the assertion follows. $\qquad\square$

4.2 Discrepancy bounds

We show that for any $s \in \mathbb{N}$, for any update function $\varphi \colon G \times [0,1]^s \to G$ of the transition kernel K, for any generator function $\psi \colon [0,1]^s \to G$ of initial distribution ν with $\frac{d\nu}{d\pi} \in L_2$ and every natural number n there exists a finite sequence $u_0, u_1, \dots, u_{n-1} \in [0,1]^s$ such that the star-discrepancy of S_n, given by (3.1), converges approximately with order $n^{-1/2}$. The main idea to prove the existence result is to use probabilistic arguments. We apply a Hoeffding inequality for variance bounding Markov chains and show that for a fixed test set the probability of point sets with small $\Delta_{n,A,\varphi,\psi}$, see (4.2), is large. We then extend this result to all sets in the δ-cover using the union bound and finally to all test sets. The result shows that if the finite driver sequence is chosen at random from the uniform distribution, most choices satisfy the Monte Carlo rate of convergence of the discrepancy for the induced point set S_n.

Theorem 4.5. *Let K be a reversible transition kernel with respect to π and ν be a distribution on $(G, \mathcal{B}(G))$ with $\frac{d\nu}{d\pi} \in L_2$. Assume that P, the Markov operator of K, is variance bounding. Let $\mathscr{A} \subseteq \mathcal{B}(G)$ be a set of test sets and for every $\delta > 0$ assume that there exists a set $\Gamma_\delta \subseteq \mathcal{B}(G)$ with $|\Gamma_\delta| < \infty$ such that Γ_δ is a δ-cover of $\mathscr{A}$ with respect to π.*

Then, for any update function $\varphi \colon G \times [0,1]^s \to G$ of K, and any generator function $\psi \colon [0,1]^s \to G$ of ν, there exists a driver sequence $u_0, u_1, \dots, u_{n-1} \in [0,1]^s$ such that $S_n = \{x_1, \dots, x_n\}$ given by $x_1 = \psi(u_0)$ and

$$x_{i+1} = x_i(x_1) = \varphi(x_i; u_i) = \varphi_i(x_1; u_1, \dots, u_i), \quad i = 1, \dots, n-1,$$

satisfies

$$D^*_{\mathscr{A},\pi}(S_n) \leq \sqrt{\frac{1 + \Lambda_0}{1 - \Lambda_0}} \cdot \frac{\sqrt{2 \log(|\Gamma_\delta|^2 \left\| \frac{d\nu}{d\pi} \right\|_2)}}{\sqrt{n}} + \delta, \tag{4.4}$$

with $\Lambda_0 = \max\{0, \Lambda\}$ and Λ defined in (2.2).

Remark 4.6. In Lemma 5.3 in Section 5.1 we show for the set of test sets of axis parallel boxes that for any $\delta > 0$ there exists a δ-cover with $|\Gamma_\delta| = \mathcal{O}(\delta^{-d/(1-\varepsilon)})$ for any $\varepsilon > 0$. Hence, for instance, by choosing $\delta = n^{-3/4}$, we obtain that $|\Gamma_{n^{-3/4}}| = \mathcal{O}(n^d)$, where we used $\varepsilon = 1/4$.

Proof. Let $A \in \mathcal{B}(G)$. By Lemma 4.4 and Proposition 4.3 we have for any $c_n \geq 0$ that

$$\mathbb{P}\left[|\Delta_{n,A,\varphi,\psi}| \leq c_n\right] \geq 1 - 2 \left\| \frac{d\nu}{d\pi} \right\|_2 \exp\left(-\frac{1 - \Lambda_0}{1 + \Lambda_0} c_n^2 n\right). \tag{4.5}$$

Let

$$\widehat{\Gamma}_\delta = \{D \setminus C : C \subseteq A \subseteq D, \text{ and } C, D \in \Gamma_\delta\}.$$

If for all $A \in \widehat{\Gamma}_\delta$ we have

$$\mathbb{P}\left[|\Delta_{n,A,\varphi,\psi}| \leq c_n\right] > 1 - \frac{1}{|\widehat{\Gamma}_\delta|}, \tag{4.6}$$

then there exists a finite sequence $u_0, \dots, u_{n-1} \in [0,1]^s$ such that

$$\max_{A \in \widehat{\Gamma}_\delta} |\Delta_{n,A,\varphi,\psi}(u_0, \dots, u_{n-1})| \leq c_n. \tag{4.7}$$

For

$$c_n = \sqrt{\frac{1 + \Lambda_0}{1 - \Lambda_0}} \cdot \frac{\sqrt{2 \log(2 |\widehat{\Gamma}_\delta| \left\| \frac{d\nu}{d\pi} \right\|_2)}}{\sqrt{n}}$$

we obtain by (4.5) that (4.6) holds and that there exists a finite sequence $u_0, \ldots, u_{n-1} \in [0,1]^s$ such that (4.7) is satisfied.

Now we extend the result from $\widehat{\Gamma}_\delta$ to $\mathscr{A}$. By the δ-cover we have for $A \in \mathscr{A}$, that there are $C, D \in \Gamma_\delta$ such that $C \subseteq A \subseteq D$ and $\pi(D \setminus C) \leq \delta$. Hence

$$
\left| \frac{1}{n} \sum_{i=0}^{n-1} \left[1_{(u_0,\ldots,u_i) \in C_{i,\psi}(A)} - \pi(A) \right] \right|
$$

$$
= \left| \frac{1}{n} \sum_{i=0}^{n-1} \left[1_{(u_0,\ldots,u_i) \in C_{i,\psi}(D)} - \pi(D) \right] - \frac{1}{n} \sum_{i=0}^{n-1} \left[1_{(u_0,\ldots,u_i) \in C_{i,\psi}(D \setminus A)} - \pi(D \setminus A) \right] \right|
$$

$$
\leq \left| \frac{1}{n} \sum_{i=0}^{n-1} \left[1_{(u_0,\ldots,u_i) \in C_{i,\psi}(D)} - \pi(D) \right] \right|
$$

$$
+ \left| \frac{1}{n} \sum_{i=0}^{n-1} \left[1_{(u_0,\ldots,u_i) \in C_{i,\psi}(D \setminus A)} - \pi(D \setminus A) \right] \right|.
$$

Set

$$
I_1 = \left| \frac{1}{n} \sum_{i=0}^{n-1} \left[1_{(u_0,\ldots,u_i) \in C_{i,\psi}(D)} - \pi(D) \right] \right|
$$

and

$$
I_2 = \left| \frac{1}{n} \sum_{i=0}^{n-1} \left[1_{(u_0,\ldots,u_i) \in C_{i,\psi}(D \setminus A)} - \pi(D \setminus A) \right] \right|.
$$

Since $\emptyset \in \Gamma_\delta$ we have $D = D \setminus \emptyset \in \widehat{\Gamma}_\delta$ and therefore

$$
I_1 \leq \max_{A \in \widehat{\Gamma}_\delta} |\Delta_{n,A,\varphi,\psi}| \leq c_n.
$$

Furthermore

$$
\begin{aligned}
I_2 &= \left| \frac{1}{n} \sum_{i=0}^{n-1} 1_{(u_0,\ldots,u_i) \in C_{i,\psi}(D \setminus A)} - \pi(D \setminus C) + \pi(D \setminus C) - \pi(D \setminus A) \right| \\
&\leq \left| \frac{1}{n} \sum_{i=0}^{n-1} \left[1_{(u_0,\ldots,u_i) \in C_{i,\psi}(D \setminus C)} - \pi(D \setminus C) \right] \right| + |\pi(D \setminus C) - \pi(D \setminus A)| \\
&\leq c_n + \delta.
\end{aligned}
$$

The last inequality follows by the δ-cover property, (4.7) and the fact that $D \setminus C \in \widehat{\Gamma}_\delta$. Finally note that $|\widehat{\Gamma}_\delta| \leq |\Gamma_\delta|^2/2$, which completes the proof. $\square$

Remark 4.7. We did not impose any regularity conditions on the update functions. In particular, for any transition kernel K on $(G, \mathcal{B}(G))$ there exists an update function $\varphi \colon G \times [0,1] \to G$, with $s = 1$, see for example [14, Lemma 2.22, p. 34]. Thus, there exists a driver sequence $U_n = \{u_0, \ldots, u_{n-1}\} \subset [0,1]$ such that S_n driven by U_n satisfies (4.4).

Remark 4.8. The proof of Theorem 4.5 shows that with probability greater than 0, there is a driver sequence $u_0, u_1, \ldots, u_{n-1} \in [0,1]^s$ which yields a Markov chain quasi-Monte Carlo point set satisfying the discrepancy bound. By increasing the constant in the discrepancy bound (4.4), we can increase this probability to $> 1/2$. Assume now we are given two different Markov chains with different transition kernels satisfying the assumptions of Theorem 4.5. Since with probability $> 1/2$ there is a driver sequence for each transition kernel satisfying the conclusion of Theorem 4.5, it follows that there is a single driver sequence $u_0, u_1, \ldots, u_{n-1}$ such that the discrepancy bound (4.4) holds for both Markov chain-quasi Monte Carlo point sets simultaneously.

By Corollary 3.3 and Theorem 4.5 we can also state an upper bound on the pull-back discrepancy.

Theorem 4.9. *Let the assumptions of Theorem 4.5 be satisfied. Then, for any update function $\varphi\colon G \times [0,1]^s \to G$ of K and for any generator function $\psi\colon [0,1]^s \to G$ of ν there exists a driver sequence $\mathcal{U}_n = \{u_0, u_1, \ldots, u_{n-1}\} \subset [0,1]^s$ such that*

$$D^*_{\mathscr{A},\psi,\varphi}(\mathcal{U}_n) \leq \sqrt{\frac{1+\Lambda_0}{1-\Lambda_0}} \cdot \frac{\sqrt{2\log(|\Gamma_\delta|^2 \left\|\frac{d\nu}{d\pi}\right\|_2)}}{\sqrt{n}} + \frac{1-\Lambda_0^n}{n \cdot (1-\Lambda_0)}\left\|\frac{d\nu}{d\pi}-1\right\|_2 + \delta,$$

with $\Lambda_0 = \max\{0, \Lambda\}$ and Λ defined in (2.2).

We refer to Remark 4.6 and Lemma 5.3 for a relation between δ and $|\Gamma_\delta|$. Thus, we showed the existence of a driver sequence with small pull-back discrepancy. Note that by using Corollary 3.3 one could also argue the other way around: If one can construct a sequence with small pull-back discrepancy then the star-discrepancy of S_n is also small.

Remark 4.10. Let us consider a special case of Theorem 4.5 and Theorem 4.9. Namely, let us assume that we can sample with respect to π. Thus, we set $\nu = \pi$ and $K(x, A) = \pi(A)$ for any $x \in G$, $A \in \mathcal{B}(G)$. Then, for any update function φ of K and generator function ψ of π we have

$$D^*_{\mathscr{A},\pi}(S_n) = D^*_{\mathscr{A},\psi,\varphi}(\mathcal{U}_n) \leq \frac{\sqrt{2\log|\Gamma_\delta|^2}}{\sqrt{n}} + \delta,$$

since $\Lambda_0 = \Lambda = 0$. This is essentially the same as Theorem 1 in [13] in their setting. However, it is not as elaborate as Theorem 4 in [13], which is based on results by Talagrand [29] and Haussler [12]. We do not know a version of these results which apply to Markov chains (such a result could yield an improvement of Theorems 4.5 and 4.9).

4.3 Burn-in period

For Markov chain Monte Carlo a burn-in period is used to reduce the bias of the initial distribution. We show how a burn-in changes the discrepancy bound of Theorem 4.9.

Let us introduce the following notation. Let $\varphi\colon G \times [0,1]^s \to G$ and $\psi\colon [0,1]^s \to G$ be measurable functions. Let $n_0, n \in \mathbb{N}$, let

$$\mathcal{U}_{n_0,n} = \{u_0, \ldots, u_{n_0}, u_{n_0+1}, \ldots, u_{n_0+n-1}\} \subset [0,1]^s$$

and assume that $S_{[n_0,n]} = \{x_{n_0+1}, \ldots, x_{n_0+n}\} \subseteq G$ is given by (3.1), i.e.

$$x_{i+1} = x_{i+1}(x_1) = \varphi(x_i; u_i) = \varphi_i(x_1; u_1, \ldots, u_i), \quad i = 1, \ldots, n_0 + n - 1,$$

where $x_1 = \psi(u_0)$. As before ψ might be considered as a generator function and φ might be considered as an update function. We now define a discrepancy measure on the driver sequence where the burn-in period is taken into account. We call it *pull-back discrepancy with burn-in*.

Definition 4.11 (Pull-back discrepancy with burn-in)**.** *Let $C_{i,\psi}(A)$ for $A \in \mathcal{B}(G)$ and $i \in \mathbb{N} \cup \{0\}$ be defined as in* (2.5)*. Define the local discrepancy function with burn-in by*

$$\Delta^{\mathrm{loc}}_{n_0,n,A,\psi,\varphi}(\mathcal{U}_{n_0,n}) = \frac{1}{n}\sum_{i=n_0}^{n_0+n-1}\left[1_{(u_0,\ldots,u_i)\in C_{i,\psi}(A)} - \lambda_{(i+1)s}(C_{i,\psi}(A))\right].$$

Let $\mathscr{A} \subseteq \mathcal{B}(G)$ be a set of test sets. Then we define the discrepancy of the driver sequence by

$$D^*_{n_0,\mathscr{A},\psi,\varphi}(\mathcal{U}_{n_0,n}) = \sup_{A\in\mathscr{A}}\left|\Delta^{\mathrm{loc}}_{n_0,n,A,\psi,\varphi}(\mathcal{U}_{n_0,n})\right|.$$

We call $D^*_{n_0,\mathscr{A},\psi,\varphi}(\mathcal{U}_{n_0,n})$ pull-back discrepancy with burn-in of $\mathcal{U}_{n_0,n}$.

By adapting Proposition 4.3 and Lemma 4.4 to the setting with burn-in we obtain, by the same steps as in the proof of Theorem 4.5, a bound on the star-discrepancy for $S_{[n_0,n]}$. Further, adapting Theorem 3.2 and Corollary 3.3 to the burn-in leads to a bound on $D^*_{n_0,\mathscr{A},\psi,\varphi}(\mathcal{U}_{n_0,n})$ for a certain set $\mathcal{U}_{n_0,n}$.

Theorem 4.12. *Let the assumptions of Theorem 4.5 be satisfied. Then, for any update function $\varphi\colon G \times [0,1]^s \to G$ of K and any generator function $\psi\colon [0,1]^s \to G$ of ν there exists a driver sequence*

$$\mathcal{U}_{n_0,n} = \{u_0, u_1, \ldots, u_{n_0+n-1}\} \subset [0,1]^s$$

such that

$$D^*_{\mathscr{A},\pi}(S_{[n_0,n]}) \leq \sqrt{\frac{1+\Lambda_0}{1-\Lambda_0}} \cdot \frac{\sqrt{2\log\left(|\Gamma_\delta|^2 \left\|\frac{d(\nu P^{n_0})}{d\pi}\right\|_2\right)}}{\sqrt{n}} + \delta,$$

with $\Lambda_0 = \max\{0,\Lambda\}$ and Λ defined in (2.2). If the Markov operator P has an absolute L_2-spectral gap we have

$$\begin{aligned}
D^*_{n_0,\mathscr{A},\psi,\varphi}(\mathcal{U}_{n_0,n}) \leq{}& \sqrt{\frac{1+\Lambda_0}{1-\Lambda_0}} \cdot \frac{\sqrt{2\log(|\Gamma_\delta|^2(1+\beta^{n_0}\left\|\frac{d\nu}{d\pi}-1\right\|_2))}}{\sqrt{n}} \\
&+ \frac{(1-\Lambda_0^n)\beta^{n_0}}{n\cdot(1-\Lambda_0)}\left\|\frac{d\nu}{d\pi}-1\right\|_2 + \delta,
\end{aligned} \tag{4.8}$$

with $\beta = \|P\|_{L_2^0 \to L_2^0}$, see Definition 2.1. In particular, by $\Lambda \leq \Lambda_0 \leq \beta < 1$ and $|\Lambda| \leq \beta$, we deduce

$$D^*_{n_0,\mathscr{A},\psi,\varphi}(\mathcal{U}_{n_0,n}) \leq \frac{4\sqrt{\log\left(|\Gamma_\delta|^2(1+\beta^{n_0}\left\|\frac{d\nu}{d\pi}-1\right\|_2)\right)}}{\sqrt{n\cdot(1-\beta)}} + \frac{2\beta^{n_0}\left\|\frac{d\nu}{d\pi}-1\right\|_2}{n\cdot(1-\beta)} + \delta. \tag{4.9}$$

Equations (4.8) and (4.9) reveal that the burn-in n_0 can eliminate the influence of the initial state induced by ψ under the assumption that there exists an absolute L_2-spectral gap. A variance bounding transition kernel is not enough, since it could be periodic and then νP^{n_0} would not converge to π at all.

5 Application

We consider the set of test sets $\mathscr{B}$ which consists of all axis parallel boxes anchored at $-\infty$ restricted to $G \subseteq \mathbb{R}^d$, i.e.

$$\mathscr{B} = \{(-\infty,x)_G \colon x \in \mathbb{R}^d\},$$

with $(-\infty,x)_G = (-\infty,x) \cap G$ and $(-\infty,x) = \Pi_{i=1}^d(-\infty,x_i)$. In the following we study the size of δ-covers with respect to such rectangular boxes.

We then focus on the application of Theorem 4.5 and state the relation between the discrepancy and the error of the computation of expectations. The Metropolis algorithm with ball walk proposal provides an example where one can see that the existence result shows an error bound which depends polynomially on the dimension d.

5.1 Delta-cover with respect to distributions

We now use an explicit version of a result due to Beck [3], for a proof and further details we refer to [1, Theorem 1]. We state it as a lemma.

Lemma 5.1. *Let* $([0,1]^d, \mathcal{B}([0,1]^d), \mu)$ *be a probability space. Let the set of test sets* $\mathscr{A} = \{[0,y) \mid y \in [0,1]^d\}$, *with* $[a,b) = \Pi_{j=1}^d [a_j, b_j)$ *for* $a,b \in \mathbb{R}^d$, *be the set of anchored boxes. Let* $\mathrm{supp}\mu$ *be the closure of*

$$\{x \in [0,1]^d : \forall \text{open neighborhoods } B \text{ of } x : \mu(B) > 0\}$$

Then, for any $r \in \mathbb{N}$ *there exists a set* $Z_r = \{z_1, \dots, z_r\}$ *with* $z_1, \dots, z_r \in \mathrm{supp}\mu$ *such that*

$$D^*_{\mathscr{A},\mu}(Z_r) \leq 63\sqrt{d} \, \frac{(2 + \log_2 r)^{(3d+1)/2}}{r}. \tag{5.1}$$

Note that $\log_2$ denotes the dyadic and $\log$ the natural logarithm.

Proof. The assertion follows by [1, Theorem 3] with $\mathbb{P} = \mu$, $X = [0,1]^d \cap \mathrm{supp}\,\mu$, $\mathscr{C} = \{[0,y) \cap \mathrm{supp}\,\mu \mid y \in \mathbb{Q}^d\}$. This implies a version of [1, Corollary 1], thus a version of [1, Theorem 1], with $x_1, \dots, x_N \in \mathrm{supp}\mu$. □

By a linear transformation we extend the result to general, bounded state spaces $G \subset \mathbb{R}^d$.

Corollary 5.2. *Let* $G \subset \mathbb{R}^d$ *be a bounded, measurable set and let* $(G, \mathcal{B}(G), \pi)$ *be a probability space. Let the set of test sets* $\mathscr{B} = \{(-\infty, x)_G \mid x \in \mathbb{R}^d\}$. *Then, for any* $r \in \mathbb{N}$ *there exists a set* $S_r = \{x_1, \dots, x_r\} \subseteq G$ *such that*

$$D^*_{\mathscr{B},\pi}(S_r) \leq 63\sqrt{d} \, \frac{(2 + \log_2 r)^{(3d+1)/2}}{r}.$$

Proof. Since G is bounded there exist $a, b \in \mathbb{R}^d$ such that $G \subseteq \prod_{j=1}^d [a_j, b_j]$. There is a linear transformation $T \colon \prod_{j=1}^d [a_j, b_j] \to [0,1]^d$ which induces a probability measure μ on $([0,1]^d, \mathcal{B}([0,1]^d))$ with $\pi(A) = \mu(T(A))$ for $A \in \mathcal{B}(G)$. In particular, for $A \in \mathcal{B}([0,1]^d \setminus T(G))$ we have $\mu(A) = 0$.

By Lemma 5.1 we have that there exists a set $Z_r = \{z_1, \dots, z_r\} \subseteq \mathrm{supp}\,\mu$ such that (5.1) is satisfied. Let $x_i = T^{-1}(z_i)$ for $i = 1, \dots, r$ and for $z \in [0,1]^d$ let $x = T^{-1}(z)$. Then

$$\frac{1}{r}\sum_{i=1}^r 1_{(-\infty,x)_G}(x_i) - \pi((-\infty,x)_G) = \frac{1}{r}\sum_{i=1}^r 1_{[0,z) \cap T(G)}(z_i) - \mu([0,z] \cap T(G)).$$

Since $z_1, \dots, z_r \in \mathrm{supp}\mu \subset T(G)$ and $\mu(A) = 0$ for $A \in \mathcal{B}([0,1]^d \setminus T(G))$ we have

$$\frac{1}{r}\sum_{i=1}^r 1_{(-\infty,x)_G}(x_i) - \pi((-\infty,x)_G) = \frac{1}{r}\sum_{i=1}^r 1_{[0,z)}(z_i) - \mu([0,z]).$$

By taking the supremum over the test sets on the right-hand side and using (5.1) the assertion follows. □

As in [8, Lemma 4] a point set which satisfies a discrepancy bound can be used to construct a δ-cover. The idea is to define for each subset of the point set a minimal and maximal set for the δ-cover, see [8, Lemma 4]. To simplify the bound of Corollary 5.2, for any $r \in \mathbb{N}$ and $0 < \varepsilon < 1$ we have

$$\frac{(2 + \log_2 r)^{(3d+1)/2}}{r} \leq r^{\varepsilon-1} C_{\varepsilon,d},$$

where

$$C_{\varepsilon,d} = \max_{x \geq 1} \frac{(2 + \log_2 x)^{(3d+1)/2}}{x^\varepsilon} = 4^\varepsilon \left(\frac{3d+1}{2\mathrm{e}\varepsilon \log 2}\right)^{(3d+1)/2}. \tag{5.2}$$

With this notation we obtain the following result.

Lemma 5.3. *Let $G \subset \mathbb{R}^d$ be a bounded measurable set and let π be a probability measure on $(G, \mathcal{B}(G))$ which is absolutely continuous with respect to the Lebesgue measure. For the set $\mathscr{B} = \{(-\infty, x)_G \mid x \in \mathbb{R}^d\}$, any $0 < \delta \leq 1$ and $0 < \varepsilon < 1$, there is a δ-cover Γ_δ of $\mathscr{B}$ with respect to π with*

$$|\Gamma_\delta| \leq \left(2 + \left\lceil (2C_{\varepsilon,d}\delta^{-1})^{1/(1-\varepsilon)} \right\rceil \right)^d,$$

where $C_{\varepsilon,d}$ is given by (5.2).

Proof. The proof of the assertion follows essentially by the same steps as the proof of [8, Lemma 4]. The only difference is that we use the discrepancy bound of Corollary 5.2 instead of [13, Theorem 4]. ☐

The dependence of the size of the δ-cover on δ is arbitrarily close to order δ^{-d} in Lemma 5.3, whereas in [8, Lemma 4] it is of order δ^{-2d}. Furthermore, the constant in Lemma 5.3 is fully explicit (one can choose $0 < \varepsilon < 1$ to obtain the best bound on the size of the δ-cover).

By Theorem 4.5 and Lemma 5.3 we obtain the following result.

Corollary 5.4. *Let $G \subset \mathbb{R}^d$ be a bounded set. Let K be a reversible transition kernel with respect to π and ν be a distribution on $(G, \mathcal{B}(G))$ with $\frac{d\nu}{d\pi} \in L_2$. Assume that P, the Markov operator of K, is variance bounding. Further, let $\mathscr{B} = \{(-\infty, x)_G \mid x \in \mathbb{R}^d\}$ be the set of test sets.*

Then, for any update function $\varphi \colon G \times [0,1]^s \to G$ of K, any generator function $\psi \colon [0,1]^s \to G$ of ν and for all $n \geq 16$, there exists a driver sequence $u_0, \ldots, u_{n-1} \in [0,1]^s$ such that $S_n = \{x_1, \ldots, x_n\}$ given by (3.1) satisfies

$$D^*_{\mathscr{B},\pi}(S_n) \leq \sqrt{\frac{1 + \Lambda_0}{1 - \Lambda_0}} \cdot \frac{\sqrt{2}\,(\log\left\|\frac{d\nu}{d\pi}\right\|_2 + d\log n + 3d^2\log(5d))^{1/2}}{\sqrt{n}} + \frac{8}{n^{3/4}}, \qquad (5.3)$$

with $\Lambda_0 = \max\{\Lambda, 0\}$.

Proof. Let $\varepsilon = 1/4$. Thus $C_{1/4,d} = \sqrt{2}(\frac{6d+2}{\mathrm{e}\log 2})^{(3d+1)/2}$ and $|\Gamma_\delta| \leq (16\delta^{-4/3}(5d)^{3d})^d$. By $\delta = 8/n^{3/4}$ and Theorem 4.5 the assertion follows. ☐

Let us discuss the result. The factor depending on Λ_0 is the penalty for the convergence of the Markov chain. The term $\log\left\|\frac{d\nu}{d\pi}\right\|_2$ shows the dependence on ψ and the additional summand $\frac{8}{n^{3/4}}$ comes from the δ-cover approximation. The rest is basically as in [13, Theorem 1].

5.2 Integration error

In this section we state a relation between a reproducing kernel Hilbert space and the star-discrepancy. As in [8, Appendix B] we define a reproducing kernel Q by

$$Q(x,y) = 1 + \int_{\mathbb{R}^d} 1_{(-\infty,z)_G}(x)\, 1_{(-\infty,z)_G}(y)\, \rho(\mathrm{d}z),$$

where ρ is a finite measure on $\mathbb{R}^d$, i.e. $\int_{\mathbb{R}^d} \rho(\mathrm{d}z) < \infty$.

The function Q uniquely defines a reproducing kernel Hilbert space $H_2 = H_2(Q)$ of functions defined on $\mathbb{R}^d$. Reproducing kernel Hilbert spaces were studied in detail in [2]. It is also known that the functions f in H_2 permit the representation

$$f(x) = f_0 + \int_{\mathbb{R}^d} 1_{(-\infty,z)_G}(x)\, \widetilde{f}(z)\, \rho(\mathrm{d}z), \qquad (5.4)$$

for some $f_0 \in \mathbb{C}$ and function $\widetilde{f} \in L_2(\mathbb{R}^d, \rho)$, see for instance [28, Theorem 4.21, p. 121] or follow the same arguments as in [4, Appendix A]. The inner product in H_2 is given by

$$\langle f, g \rangle = f_0 \overline{g_0} + \int_{\mathbb{R}^d} \widetilde{f}(z) \overline{\widetilde{g}(z)} \, \rho(\mathrm{d}z).$$

With these definitions we have the reproducing property

$$\langle f, Q(\cdot, y) \rangle = f_0 + \int_{\mathbb{R}^d} \widetilde{f}(z) 1_{(-\infty,z)_G}(y) \rho(\mathrm{d}z) = f(y).$$

For $1 \leq q \leq \infty$ we also define the space H_q of functions of the form (5.4) for which $\widetilde{f} \in L_q(G, \rho)$, with finite norm

$$\|f\|_{H_q} = \left(|f_0|^q + \int_{\mathbb{R}^d} |\widetilde{f}(z)|^q \rho(\mathrm{d}z) \right)^{1/q}. \tag{5.5}$$

The following result concerning the integration error in H_q is proven in [8, Theorem 3].

Theorem 5.5. *Let $G \subseteq \mathbb{R}^d$ and π be a probability measure on G. Further let $\mathscr{B} = \{(-\infty, x)_G : x \in \mathbb{R}^d\}$. We assume that $1 \leq p, q \leq \infty$ with $1/p + 1/q = 1$. Then for $Z_n = \{z_1, z_2, \ldots, z_n\} \subseteq G$ and for all $f \in H_q$ we have*

$$\left| \int_G f(z) \pi(\mathrm{d}z) - \frac{1}{n} \sum_{i=1}^n f(z_i) \right| \leq \|f\|_{H_q} D_{p,\mathscr{B},\pi}^*(Z_n),$$

where

$$D_{p,\mathscr{B},\pi}^*(Z_n) = \left(\int_{\mathbb{R}^d} \left| \int_G 1_{(-\infty,z)_G}(y) \pi(\mathrm{d}y) - \frac{1}{n} \sum_{i=1}^n 1_{(-\infty,z)_G}(z_i) \right|^p \rho(\mathrm{d}z) \right)^{1/p},$$

and for $p = \infty$ let

$$D_{\mathscr{B},\pi}^*(Z_n) := D_{\infty,\mathscr{B},\pi}^*(Z_n) = \sup_{z \in \mathbb{R}^d} \left| \int_G 1_{(-\infty,z)_G}(y) \pi(\mathrm{d}y) - \frac{1}{n} \sum_{i=1}^n 1_{(-\infty,z)_G}(z_i) \right|.$$

Corollary 5.6 (Markov chain Koksma-Hlawka inequality)**.** *Assume that the conditions of Corollary 3.3 are satisfied. Further let $\mathscr{B} = \{(-\infty, x)_G : x \in \mathbb{R}^d\}$. Let H_1 denote the space of functions $f : \mathbb{R}^d \to \mathbb{C}$ with finite norm given by (5.5). Then, for any update function $\varphi \colon G \times [0,1]^s \to G$ of K and any generator function $\psi \colon [0,1]^s \to G$ of ν we have, with driver sequence $\mathcal{U}_n = \{u_0, u_1, \ldots, u_{n-1}\} \subset [0,1]^s$ and S_n given by (3.1), that*

$$\left| \int_G f(x) \pi(\mathrm{d}z) - \frac{1}{n} \sum_{i=1}^n f(x_i) \right|$$
$$\leq \left(D_{\mathscr{B},\psi,\varphi}^*(\mathcal{U}_n) + \frac{1 - \Lambda_0^n}{n \cdot (1 - \Lambda_0)} \left\| \frac{d\nu}{d\pi} - 1 \right\|_2 \right) \|f\|_{H_1},$$

with $\Lambda_0 = \max\{0, \Lambda\}$, where Λ is defined in (2.2).

In the spirit of Remark 3.4 we obtain for $K(x, A) = \pi(A)$ that $\Lambda = 0$. Further, if $\nu = \pi$ we have the Koksma-Hlawka inequality (cf. [16, p. 151, Theorem 5.5])

$$\left| \int_G f(x) \pi(\mathrm{d}x) - \frac{1}{n} \sum_{i=1}^n f(x_i) \right| \leq D_{\mathscr{B},\psi,\varphi}^*(\mathcal{U}_n) \|f\|_{H_1}.$$

5.3 Metropolis algorithm with ball walk proposal

The goal of this subsection is the application of the previously developed theory to an example. Let us assume that $G = \mathbb{B}_d$ is the Euclidean unit ball, i.e. $\mathbb{B}_d = \{x \in \mathbb{R}^d \mid \|x\| := (\sum_{i=1}^{d} |x_i|^2)^{1/2} \leq 1\}$. Let $\rho \colon \mathbb{B}_d \to (0, \infty)$ be integrable with respect to the Lebesgue measure. We define the distribution π_ρ on $(\mathbb{B}_d, \mathcal{B}(\mathbb{B}_d))$ by

$$\pi_\rho(A) = \frac{\int_A \rho(x)\,\mathrm{d}x}{\int_{\mathbb{B}_d} \rho(x)\,\mathrm{d}x}.$$

The goal is to compute

$$\mathbb{E}_{\pi_\rho}(f) = \int_{\mathbb{B}_d} f(x)\,\pi_\rho(\mathrm{d}x) = \frac{\int_{\mathbb{B}_d} f(x)\rho(x)\,\mathrm{d}x}{\int_{\mathbb{B}_d} \rho(x)\,\mathrm{d}x},$$

for functions $f \colon \mathbb{B}_d \to \mathbb{R}$ which are integrable with respect to π_ρ. Note that for an approximation of $\mathbb{E}_{\pi_\rho}(f)$ the functions f and ρ are part of the input of a possible approximation scheme. We assume that sampling directly with respect to π_ρ is not feasible. We use the Metropolis algorithm with ball walk proposal to sample approximately according to π_ρ.

Let $\gamma > 0$, $x \in \mathbb{B}_d$ and $C \in \mathcal{B}(\mathbb{B}_d)$, then the transition kernel of the γ ball walk is given by

$$W_\gamma(x, C) = \frac{\lambda_d(C \cap D_\gamma(x))}{\lambda_d(D_\gamma(0))} + 1_{x \in A} \left[1 - \frac{\lambda_d(\mathbb{B}_d \cap D_\gamma(x))}{\lambda_d(D_\gamma(0))}\right],$$

where λ_d denotes the d-dimensional Lebesgue measure and $D_\gamma(x) = \{y \in \mathbb{R}^d \mid \|x - y\| \leq \gamma\}$ denotes the Euclidean ball with radius γ around $x \in \mathbb{R}^d$. The transition kernel of the Metropolis algorithm with ball walk proposal is given by

$$M_{\rho,\gamma}(x, C) = \int_C \theta(x, y)\, W_\gamma(x, \mathrm{d}y) + 1_{x \in A} \left[1 - \int_{\mathbb{B}_d} \theta(x, y)\, W_\gamma(x, \mathrm{d}y)\right],$$

where $\theta(x, y) = \min\{1, \rho(y)/\rho(x)\}$ is the so-called acceptance probability. The transition kernel $M_{\rho,\gamma}$ is reversible with respect to π_ρ.

Now we provide update functions of the ball walk and the Metropolis algorithm with ball walk proposal. Let $\mathbb{S}^{d-1} = \{x \in \mathbb{R}^d \mid \|x\| = 1\}$ be the unit sphere in $\mathbb{R}^d$. Let $\widetilde{\psi} \colon [0,1]^{d-1} \to \mathbb{S}^{d-1}$ be a generator for the uniform distribution on the sphere, see for instance [10]. Then, $\psi_\gamma \colon [0,1]^d \to D_\gamma(0)$ given by

$$\psi_\gamma(\bar{u}) = \gamma\, v_d^{1/d}\, \widetilde{\psi}(v_1, \ldots, v_{d-1}), \tag{5.6}$$

with $\bar{u} = (v_1, \ldots, v_d) \in [0,1]^d$, is a generator for the uniform distribution in $D_\gamma(0)$ (the Euclidean ball with radius γ around 0). Thus, an update function $\varphi_{W,\gamma} \colon \mathbb{B}_d \times [0,1]^d \to \mathbb{B}_d$ of the γ ball walk, with $\bar{u} = (v_1, \ldots, v_d) \in [0,1]^d$, is

$$\varphi_{W,\gamma}(x, \bar{u}) = \begin{cases} x + \psi_\gamma(\bar{u}) & x + \psi_\gamma(\bar{u}) \in \mathbb{B}_d \\ x & \text{otherwise.} \end{cases}$$

This leads to an update function $\varphi_{M,\gamma,\rho} \colon \mathbb{B}_d \times [0,1]^{d+1} \to \mathbb{B}_d$ of the Metropolis algorithm with ball walk proposal. Let

$$A(x; \bar{u}) = \min\{1, \rho(\varphi_{W,\gamma}(x, \bar{u}))/\rho(x)\},$$

then an update function for the Metropolis algorithm with ball walk proposal is

$$\varphi_{M,\gamma,\rho}(x, u) = \begin{cases} \varphi_{W,\gamma}(x, v_1, \ldots, v_d) & v_{d+1} \leq A(x, v_1, \ldots, v_d) \\ x & v_{d+1} > A(x, v_1, \ldots, v_d), \end{cases} \tag{5.7}$$

where $u = (v_1, \ldots, v_{d+1}) \in [0,1]^{d+1}$ and $x \in \mathbb{B}_d$. Thus, we have an update function of W_γ. For the convenience of the reader we provide a transition of the Metropolis algorithm with ball walk proposal from x to y with driving point $(v_1, \ldots, v_{d+1}) \in [0,1]^{d+1}$ in algorithmic form:

Algorithm 1. Metropolis algorithm with ball walk proposal

Input: driving point $(v_1, \ldots, v_{d+1}) \in [0,1]^{d+1}$, and
          current state $x \in \mathbb{B}_d$;
Output: next state $y \in \mathbb{B}_d$;

1. Compute $z := \gamma\, v_d^{1/d}\, \widetilde{\psi}(v_1, \ldots, v_{d-1})$ where $\widetilde{\psi}$ is a generator function for the uniform distribution on $\mathbb{S}^{d-1}$.

2.   a) If $x + z \in \mathbb{B}_d$ and $v_{d+1} \leq \min\{1, \rho(x+z)/\rho(x)\}$, then $y := x + z$.

   b) Otherwise $y := x$.

3. Return y.

   We assume that the functions $f\colon \mathbb{B}_d \to \mathbb{R}$ and $\rho\colon \mathbb{B}_d \to (0, \infty)$ have some additional structure. Let $f \in H_1$ with $\|f\|_{H_1} \leq 1$, where H_1 is defined in Subsection 5.2. For $\alpha > 0$ let $\rho \in \mathcal{R}_{\alpha,d}$ if the following conditions are satisfied:

(i) ρ is log-concave, i.e. for all $\lambda \in (0,1)$ and for all $x, y \in \mathbb{B}_d$ holds

$$\rho(\lambda x + (1 - \lambda)y) \geq \rho(x)^\lambda \rho(y)^{1-\lambda}.$$

(ii) ρ is log-Lipschitz continuous with α, i.e.

$$|\log \rho(x) - \log \rho(y)| \leq \alpha\|x - y\|.$$

Thus

$$\mathcal{R}_{\alpha,d} = \{\rho\colon \mathbb{B}_d \to (0, \infty) \mid \rho \text{ log-concave}, |\log \rho(x) - \log \rho(y)| \leq \alpha\|x - y\|\}. \qquad (5.8)$$

Next we provide a lower bound for $\Lambda_{\gamma,\rho}$, defined as in (2.2) for the transition kernel $M_{\gamma,\rho}$, where the density ρ is log-concave and log-Lipschitz. The result follows by [21, Corollary 1, Lemma 13].

Proposition 5.7. *Let us assume that $\rho \in \mathcal{R}_{\alpha,d}$. Further let*

$$\gamma^* = \min\{1/\sqrt{d+1}, 1/\alpha\}.$$

Then

$$1 - \Lambda_{\gamma^*,\rho} \geq \frac{3.125 \cdot 10^{-6}}{d+1} \min\left\{\frac{1}{d+1}, \frac{1}{\alpha}\right\}. \qquad (5.9)$$

   The combination of Proposition 5.7, Theorem 5.5, Lemma 5.3 and Corollary 5.4 lead to the following error bound for the computation of $\mathbb{E}_{\pi_\rho}(f)$ for $f \in H_1$ and $\rho \in \mathcal{R}_{\alpha,d}$.

Theorem 5.8. *Let ν be the uniform distribution on $(\mathbb{B}_d, \mathcal{B}(\mathbb{B}_d))$ with generator function ψ_1, see (5.6). Let*

$$\gamma^* = \min\{1/\sqrt{d+1}, 1/\alpha\}$$

and recall that $\varphi_{M,\gamma^*,\rho}$ is an update function of the Metropolis algorithm with ball walk proposal, see (5.7).

Then, for all $n \geq 16$ and any $\rho \in \mathcal{R}_{\alpha,d}$ there exists a driver sequence $u_0, u_1, \ldots, u_{n-1} \in [0,1]^{d+1}$ such that $S_n = \{x_1, \ldots, x_n\}$ given by

$$x_1 = \psi_1(\bar{u}_0)$$
$$x_{i+1} = \varphi_{M,\gamma^*,\rho}(x_i; u_i), \qquad i = 1, \ldots, n-1,$$

with $\bar{u}_0 = (v_1, \ldots, v_d)$ where $u_0 = (v_1, \ldots, v_d, v_{d+1})$, satisfies

$$\sup_{f \in H_1, \|f\|_{H_1} \leq 1} \left| \mathbb{E}_{\pi_\rho}(f) - \frac{1}{n} \sum_{i=1}^{n} f(x_i) \right|$$

$$\leq \frac{5000\sqrt{d} \max\{\sqrt{2d}, \sqrt{\alpha}\} \left(\alpha + d \log n + 3d^2 \log(5d)\right)^{1/2}}{\sqrt{n}} + \frac{8}{n^{3/4}}.$$

Proof. By

$$\frac{d\nu}{d\pi_\rho}(x) = \frac{\int_{\mathbb{B}_d} \rho(y)\,\mathrm{d}y}{\lambda_d(\mathbb{B}_d)\rho(x)},$$

and by $\rho(x)/\rho(y) \leq \exp(2\alpha)$ for any $x, y \in \mathbb{B}_d$ we have $\|\frac{d\nu}{d\pi_\rho}\|_2 \leq \exp\alpha$. Further, by Proposition 5.7 we obtain

$$1 - \Lambda_0 \geq \frac{3.125 \cdot 10^{-6}}{d+1} \min\left\{\frac{1}{d+1}, \frac{1}{\alpha}\right\}.$$

Thus by Corollary 5.4 and Theorem 5.5 the assertion follows. $\qquad\square$

Let us emphasize that the theorem shows that for any $\rho \in \mathcal{R}_{\alpha,d}$ there exist a deterministic algorithm where the error depends only polynomially on the dimension d and the Log-Lipschitz constant α.

6 Beyond the Monte Carlo rate

In the previous sections we have seen that there exist deterministic driver sequences which yield almost the Monte Carlo rate of convergence of $n^{-1/2}$. Roughly speaking, the proof of Theorem 4.5 reveals that, if the driver sequence is chosen at random from the uniform distribution the discrepancy bound of (4.4) is satisfied with high probability. In this section we use a stronger assumption to achieve a better rate of convergence. Again this result is an existence result. We want to point out that the proof of the result does not reveal any information on how to find driver sequences which lead to good discrepancy bounds.

Its proof is based on an additional regularity condition of the update function, the 'anywhere-to-anywhere' condition, and Corollary 5.2.

Definition 6.1. *Let $\varphi : G \times [0,1]^s \to G$ be an update function of a transition kernel K. We say that φ satisfies the 'anywhere-to-anywhere' condition if for all $x, y \in G$ there exists a $u \in [0,1]^s$ such that*

$$\varphi(x; u) = y.$$

Now we use the 'anywhere-to-anywhere' condition to reformulate Corollary 5.2. We obtain a bound on the star-discrepancy for the Markov chain quasi-Monte Carlo construction.

Corollary 6.2. *Let $G \subset \mathbb{R}^d$ be a bounded, measurable set and let $(G, \mathcal{B}(G), \pi)$ be a probability space. Let the set of test sets $\mathscr{B} = \{(-\infty, x) \cap G \mid x \in \mathbb{R}^d\}$ be the set of anchored boxes intersected with G.*

Then, for any update function $\varphi \colon G \times [0,1]^s \to G$ of the transition kernel K which satisfies the 'anywhere-to-anywhere' condition, any surjective function $\psi \colon [0,1]^s \to G$

and for any $n \in \mathbb{N}$ there exists a driver sequence $u_0, u_1, \ldots u_{n-1} \in [0,1]^s$ such that $S_n = \{x_1, \ldots, x_n\}$ given by $x_1 = \psi(u_0)$ and

$$x_i = \varphi(x_{i-1}; u_i), \qquad i = 1, \ldots, n-1,$$

satisfies

$$D^*_{\mathscr{B},\pi}(S_n) \le 63\sqrt{d}\, \frac{(2 + \log_2 n)^{(3d+1)/2}}{n}.$$

The corollary states that if the 'anywhere-to-anywhere' condition is satisfied, in principle, we can get the same discrepancy for the Markov chain quasi-Monte Carlo construction as without using any Markov chain. If the update function and underlying Markov operator P satisfies the conditions of Corollary 3.3, then a similar discrepancy bound as in Corollary 6.2 also holds for the driver sequence $\mathcal{U}_n = \{u_0, u_1, \ldots, u_{n-1}\}$. Namely

$$D^*_{\mathscr{B},\psi,\varphi}(\mathcal{U}_n) \le 63\sqrt{d}\, \frac{(2 + \log_2 n)^{(3d+1)/2}}{n} + \frac{1 - \Lambda_0^n}{n \cdot (1 - \Lambda_0)} \left\| \frac{d\nu}{d\pi} - 1 \right\|_2.$$

7 Concluding remarks

Let us point out that the discrepancy results of Subsection 4.2 and Subsection 4.3, in particular, also hold for local Markov chains which do not satisfy the 'anywhere to anywhere' condition and the proof of this bound reveals that a uniformly i.i.d. driver sequence satisfies the discrepancy estimate with high probability. In other words, there are many driver sequences which satisfy the discrepancy bound of order $(\log n)^{1/2} n^{-1/2}$.

On the other hand, the choice of the driver sequence depends on the initial distribution ν and the transition kernel. It would be interesting to prove the existence of a universal driver sequence, which yields Monte Carlo type behavior for a class of initial distributions and transition kernels. (For a finite set of initial distributions and transition kernels such a result can be obtained from our results since for any given initial distribution and transition kernel we can show the existence of good driver sequences with high probability, see Remark 4.8.) Further, the proven bounds on the discrepancy are based on a covering argument with the Vapnik-Červonenkis dimension. It is natural to ask whether one can get better estimates with other covering arguments, for example Dudley's entropy [9] or its variants.

Another open problem is the explicit construction of suitable driver sequences. The results in this paper do not give any indication how such a construction could be obtained. However, as a step towards explicit constructions, we do obtain that the pull-back discrepancy is the relevant criterion for constructing driver sequences.

References

[1] Ch. Aistleitner and J. Dick, *Low-discrepancy point sets for non-uniform measures*, Acta Arith. **163** (2014), 345–369. MR-3217671

[2] N. Aronszajn, *Theory of reproducing kernels*, Trans. Amer. Math. Soc. **68** (1950), 337–404. MR-0051437

[3] J. Beck, *Some upper bounds in the theory of irregularities of distribution*, Acta Arith. **43** (1984), 115–130. MR-0736726

[4] J. Brauchart and J. Dick, *A characterization of sobolev spaces on the sphere and an extension of stolarsky's invariance principle to arbitrary smoothness*, Constr. Approx. **38** (2013), 397–445. MR-3122277

[5] S. Chen, *Consistency and convergence rate of Markov chain quasi-Monte Carlo with examples*, Ph.D. thesis, Stanford University, 2011.

[6] S. Chen, J. Dick, and A. Owen, *Consistency of Markov chain quasi-Monte Carlo on continuous state spaces*, Ann. Statist. **39** (2011), 673–701. MR-2816335

[7] S. Chen, M. Matsumoto, T. Nishimura, and A. Owen, *New inputs and methods for Markov chain quasi-Monte Carlo*, In Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing, (L. Plaskota and H. Woźniakowski, eds.) (2012), 313–327. MR-3173841

[8] J. Dick, D. Rudolf, and H. Zhu, *Discrepancy bounds for uniformly ergodic Markov chain quasi-Monte Carlo*, Preprint, Available at http://arxiv.org/abs/1303.2423 (2013).

[9] R. Dudley, *The sizes of compact subsets of Hilbert space and continuity of Gaussian processes*, J. Functional Analysis **1** (1967), 290–330. MR-0220340

[10] K. Fang and Y. Wang, *Number-theoretic methods in statistics*, Chapman & Hall, London, 1994. MR-1284470

[11] M. Gnewuch, *Bracketing numbers for axis-parallel boxes and applications to geometric discrepancy*, J. Complexity **24** (2008), 154–172. MR-2400314

[12] D. Haussler, *Sphere packing numbers for subsets of the boolean n-cube with bounded Vapnik-Chervonenkis dimension*, J. Combin. Theory Ser. A **69** (1995), 217–232. MR-1313896

[13] S. Heinrich, E. Novak, G. Wasilkowski, and H. Woźniakowski, *The inverse of the star-discrepancy depends linearly on the dimension*, Acta Arith. **96** (2001), 279–302. MR-1814282

[14] O. Kallenberg, *Foundations of modern probability*, second ed., Probability and its Applications, Springer-Verlag, New York, 2002. MR-1876169

[15] E. Kreyszig, *Introductory functional analysis with applications*, John Wiley & Sons, Inc., 1989. MR-0992618

[16] L. Kuipers and H. Niederreiter, *Uniform distribution of sequences*, Dover Publications, New York, 2006.

[17] C. Lemieux and P. Sidorsky, *Exact sampling with highly uniform point sets*, Math. Comput. Modelling **43** (2006), 339–349. MR-2214643

[18] C. León and F. Perron, *Optimal Hoeffding bounds for discrete reversible Markov chains*, Ann. Appl. Probab. **14** (2004), no. 2, 958–970. MR-2052909

[19] L. Liao, *Variance reduction in gibbs sampler using quasi random numbers*, J. Comput. Graph. Statist. **7** (1998), 253–266.

[20] J. Liu, *Monte Carlo strategies in scientific computing*, Springer Series in Statistics, Springer, New York, 2008. MR-2401592

[21] P. Mathé and E. Novak, *Simple Monte Carlo and the Metropolis algorithm*, J. Complexity **23** (2007), no. 4-6, 673–696. MR-2372022

[22] B. Miasojedow, *Hoeffding's inequalities for geometrically ergodic Markov chains on general state space*, Statist. Probab. Lett. **87** (2014), 115–120 . MR-3168944

[23] A. Owen and S. Tribble, *A quasi-Monte Carlo Metropolis algorithm*, Proc. Natl. Acad. Sci. USA, 102 **102** (2005), no. 25, 8844–8849. MR-2168266

[24] G. Roberts and J. Rosenthal, *Variance bounding Markov chains*, Ann. Appl. Probab. **18** (2008), no. 3, 1201–1214. MR-2418242

[25] W. Rudin, *Functional analysis*, second ed., International Series in Pure and Applied Mathematics, McGraw-Hill Inc., New York, 1991. MR-1157815

[26] D. Rudolf, *Explicit error bounds for Markov chain Monte Carlo*, Dissertationes Math. **485** (2012), 93 pp. MR-2977521

[27] I. Sobol, *Pseudo-random numbers for constructing discrete Markov chains by the Monte Carlo method*, USSR Compat. Math. Math. Phys. **14** (1974), 36–45. MR-0339444

[28] I. Steinwart and A. Christmann, *Support vector machines*, Information Science and Statistics, Springer-Verlag New York, 2008. MR-2450103

[29] M. Talagrand, *Sharper bounds for gaussian and empirical processes*, Ann. Probab. **22** (1994), 28–76. MR-1258865

[30] S. Tribble, *Markov chain Monte Carlo algorithms using completely uniformly distributed driving sequences*, Ph.D. thesis, Stanford University, 2007. MR-2710331

[31] S. Tribble and A. Owen, *Construction of weakly cud sequences for MCMC sampling*, Electron. J. Stat. **2** (2008), 634–660. MR-2426105

We are grateful to Art Owen for helpful discussions. He coined the phrase 'anywhere-to-anywhere'.