

## DISCREPANCY BOUNDS FOR UNIFORMLY ERGODIC MARKOV CHAIN QUASI-MONTE CARLO

BY JOSEF DICK<sup>\*,1</sup>, DANIEL RUDOLF<sup>†,2</sup> AND HOUYING ZHU<sup>\*,3</sup>

*University of New South Wales\* and Universität Jena<sup>†</sup>*

Markov chains can be used to generate samples whose distribution approximates a given target distribution. The quality of the samples of such Markov chains can be measured by the discrepancy between the empirical distribution of the samples and the target distribution. We prove upper bounds on this discrepancy under the assumption that the Markov chain is uniformly ergodic and the driver sequence is deterministic rather than independent  $U(0, 1)$  random variables. In particular, we show the existence of driver sequences for which the discrepancy of the Markov chain from the target distribution with respect to certain test sets converges with (almost) the usual Monte Carlo rate of  $n^{-1/2}$ .

**1. Introduction.** Markov chain Monte Carlo (MCMC) algorithms are used for the approximation of an expected value with respect to the stationary probability measure  $\pi$  of the chain. This is done by simulating a Markov chain  $(X_i)_{i \geq 1}$  and using the sample average  $\frac{1}{n} \sum_{i=1}^n f(X_i)$  to estimate the mean  $\mathbb{E}_\pi(f) := \int_G f(x)\pi(dx)$ , where  $G$  is the state space and  $f$  is a real-valued function defined on  $G$ . This method is a staple tool in the physical sciences and Bayesian statistics.

A single transition from  $X_{i-1}$  to  $X_i$  of a Markov chain is generated by using the current state  $X_{i-1}$  and a random source  $U_i$ , usually taken from an i.i.d.  $U(0, 1)$  sequence  $(U_i)_{i \geq 1}$  of random numbers. In contrast, the Markov chain quasi-Monte Carlo idea is as follows: Substitute the sequence of random numbers by a deterministically constructed finite sequence of numbers  $(u_i)_{1 \leq i \leq n}$  in  $[0, 1]^s$  for all  $n \in \mathbb{N}$ . Numerical experiments suggest that for judiciously chosen deterministic pseudo-random numbers  $(u_i)_{1 \leq i \leq n}$  this can lead to significant improvements. Owen and Tribble [25] and Tribble [35] report an improvement by a factor of up to  $10^3$  and a faster convergence rate for a Gibbs sampling problem. There were also previous attempts which provided evidence that the approach leads to comparable results [16, 18, 33]. Another line of research, dealing with the so-called

---

Received March 2013; revised February 2015.

<sup>1</sup>Recipient of an Australian Research Council Queen Elizabeth II Fellowship (project number DP1097023).

<sup>2</sup>Supported in part by the ARC Discovery Grant DP120101816, by the DFG priority program 1324 and the DFG Research training group 1523.

<sup>3</sup>Supported in part by the ARC Discovery Grant DP120101816.

*MSC2010 subject classifications.* Primary 60J22, 65C40, 62F15; secondary 65C05, 60J05.

*Key words and phrases.* Markov chain Monte Carlo, uniformly ergodic Markov chain, discrepancy theory, probabilistic method.

array-RQMC method, also combines MCMC with quasi-Monte Carlo [15]. For a thorough literature review we refer to [6], Section 1.1 (Literature review).

Recently, in the work of Chen, Dick and Owen [6] and Chen [5], the first theoretical justification of the Markov chain quasi-Monte Carlo approach on continuous state spaces was provided. Therein a consistency result is proven if the random sequence  $(U_i)_{i \geq 1}$  is substituted by a deterministic “completely uniformly distributed” (CUD) sequence  $(u_i)_{i \geq 1}$ , the Markov chain satisfies a contraction assumption and the integrand  $f$  is continuous. For a precise definition of CUD sequences, we refer to [6, 7] and for the construction of weakly CUD sequences we refer to [36]. The consistency result of Markov chain quasi-Monte Carlo corresponds to an ergodic theorem for Markov chain Monte Carlo and can be shown to be equivalent to the statement that the discrepancy between the empirical distribution and target distribution converges to 0 (this follows directly from [6], Theorem 1). However, from the result in [6], it is not clear how fast the sample average converges to the desired expectation. The goal of this paper is to investigate the convergence behavior of such Markov chain quasi-Monte Carlo algorithms. We describe the setting and main results in the following.

Throughout the paper, we deal with uniformly ergodic Markov chains on a state space  $G \subseteq \mathbb{R}^d$  and a probability space  $(G, \mathcal{B}(G), \pi)$ , where  $\mathcal{B}(G)$  is the Borel  $\sigma$ -algebra defined on  $G$  and  $\pi$  is the stationary distribution of the Markov chain; for details see, for example, [21, 27, 30]. We assume that the Markov chain can be generated by an update function  $\varphi : G \times [0, 1]^s \rightarrow G$ , that is,  $X_i = \varphi(X_{i-1}; U_i)$  for all  $i \geq 1$ . We fix a starting point  $x_0 = x$  and replace the random numbers  $(U_i)_{i \geq 1}$  by a deterministic sequence  $(u_i)_{i \geq 1}$  to generate the deterministic points  $x_i = \varphi(x_{i-1}; u_i)$  for  $i \geq 1$ . The convergence behavior of the Markov chain is measured using a generalized Kolmogorov–Smirnov test between the stationary distribution  $\pi$  and the empirical distribution  $\hat{\pi}_n(A) := \frac{1}{n} \sum_{i=1}^n 1_{x_i \in A}$ , where  $1_{x_i \in A}$  is the indicator function of the set  $A \in \mathcal{B}(G)$ . The discrepancy is defined by taking the supremum of  $|\pi(A) - \hat{\pi}_n(A)|$  over all sets in  $\mathcal{A} \subseteq \mathcal{B}(G)$  [since the empirical distribution is based on a finite number of points in  $G$  we generally have  $\mathcal{A} \neq \mathcal{B}(G)$ , see below for a more detailed description]. Under these assumptions, we prove that, for each  $n \in \mathbb{N}$ , there exists a finite sequence of numbers  $(u_i)_{1 \leq i \leq n}$  such that this discrepancy converges with order  $\mathcal{O}(n^{-1/2}(\log n)^{1/2})$  as  $n$  tends to infinity. This is roughly the convergence rate which one would expect from MCMC algorithms based on random inputs.

A drawback of our results is that we are currently not able to give explicit constructions of sequences  $(u_i)_{1 \leq i \leq n}$  for which our discrepancy bounds hold. This is because our proofs make essential use of probabilistic arguments. Namely, we use a Hoeffding inequality by Glynn and Ormoneit [11] and some results by Talagrand [34] on empirical processes and a result by Haussler [13]. Roughly speaking, we use the Hoeffding inequality to show that the probability of all  $(X_i)_{1 \leq i \leq n}$  with small discrepancy is bigger than 0, which implies the existence of a Markov chain with small discrepancy. We do, however, give a criterion (which we call “pull-back discrepancy”) which the numbers  $(u_i)_{1 \leq i \leq n}$  need to satisfy such that the point set

$(x_i)_{1 \leq i \leq n}$  has small discrepancy. This is done by showing that the discrepancy of  $(x_i)_{1 \leq i \leq n}$  is close to the pull-back discrepancy of the driver sequence  $(u_i)_{1 \leq i \leq n}$ . This should eventually lead to explicit constructions of suitable driver sequences. As a corollary to the relation between the discrepancy of the Markov chain and the pull-back discrepancy of the driver sequence, we obtain a Koksma–Hlawka inequality for Markov chains in terms of the discrepancy of the driver sequence. We point out that the pull-back discrepancy generally differs from the CUD property studied in [5] and [6]. Convergence rates beyond the usual Monte Carlo rate of  $n^{-1/2}$  have previously been shown for Array-RQMC [15] and in [5], Chapter 6. In both of these instances, a direct simulation is (at least in principle) possible.

Our results on the discrepancy of the points  $(x_i)_{1 \leq i \leq n}$  can also be understood as an extension of results on point distributions in the unit cube  $[0, 1]^s$  (see [14]) to uniformly ergodic Markov chains.

We give a brief outline of our work. In the next section, we provide background information on uniformly ergodic Markov chains, give a relation between the transition kernel of a Markov chain and their update function and state some examples which satisfy the convergence properties. We also give some background on discrepancy and describe our results in more detail. In Section 3, we provide the notion of discrepancy with respect to the driver sequence and we prove the close relation between the two types of discrepancy for uniformly ergodic Markov chains from which we deduce a Koksma–Hlawka type inequality. In Section 4, we prove the main results. The Appendix contains sections on  $\delta$ -covers, the integration error and some technical proofs.

**2. Background and notation.** In this section, we provide the necessary background on discrepancy and uniformly ergodic Markov chains.

*2.1. Discrepancy.* The convergence behavior of the Markov chain is analyzed with respect to a distance measure between the empirical distribution of the Markov chain and its stationary distribution  $\pi$ . It can be viewed as an extension of the Kolmogorov–Smirnov test and is a well established concept in numerical analysis and number theory [9]. We analyze the empirical distribution of the first  $n$  points of the Markov chain  $X_1, \dots, X_n$  by assigning each point the same weight and defining the empirical measure of a set  $A \in \mathcal{B}(G)$  by

$$\widehat{\pi}_n(A) = \frac{1}{n} \sum_{i=1}^n 1_{X_i \in A},$$

where the indicator function is given by

$$1_{X_i \in A} = \begin{cases} 1, & \text{if } X_i \in A, \\ 0, & \text{otherwise.} \end{cases}$$

The local discrepancy between the empirical distribution and the stationary distribution is then

$$\Delta_{n,A} = \widehat{\pi}_n(A) - \pi(A).$$

To obtain a measure for the discrepancy, we take the supremum of  $|\Delta_{n,A}|$  over certain sets  $A$ . Note that since the empirical measure uses only a finite number of points the local discrepancy  $\Delta_{n,A}$  does not converge to 0 in general if we take the supremum over all sets in  $\mathcal{B}(G)$ . Thus, we restrict the supremum to a set of so-called test sets  $\mathcal{A} \subseteq \mathcal{B}(G)$ . Now we define the discrepancy.

DEFINITION 1 (Discrepancy). The discrepancy of  $P_n = \{X_1, \dots, X_n\} \subseteq G$  is given by

$$D_{\mathcal{A},\pi}^*(P_n) = \sup_{A \in \mathcal{A}} |\Delta_{n,A}|.$$

This is the measure which we use to analyze the convergence behavior of the Markov chain as  $n$  goes to  $\infty$ .

In Appendix C, we provide a relationship between the discrepancy  $D_{\mathcal{A},\pi}^*(P_n)$  and the integration error of functions in a certain function space  $H_1$ , where the set of test sets is given by

$$\mathcal{A} = \{(-\infty, x)_G : x \in \bar{\mathbb{R}}^d\},$$

with  $\bar{\mathbb{R}}^d = (\mathbb{R} \cup \{\infty, -\infty\})^d$  and  $(-\infty, x)_G := (-\infty, x) \cap G = \prod_{j=1}^d (-\infty, \xi_j) \cap G$  for  $x = (\xi_1, \dots, \xi_d)$ . In particular, if there is at least one  $i$  with  $\xi_i = -\infty$ , then  $(-\infty, x)_G = \emptyset$ , whereas if all  $\xi_i = \infty$ , then  $(-\infty, x)_G = G \subseteq \mathbb{R}^d$ . For functions  $f \in H_1$ , we have

$$\left| \mathbb{E}_\pi(f) - \frac{1}{n} \sum_{i=1}^n f(X_i) \right| \leq D_{\mathcal{A},\pi}^*(P_n) \|f\|_{H_1}.$$

Inequalities of this form are called Koksma–Hlawka inequalities; see [9], Chapter 2, for more information. See Appendix C for details on the definition of the space  $H_1$  and the proof of the inequality.

2.2. *Markov chains.* The main assumption on the Markov chain in [5] and [6] is the existence of a coupling region, or in a weakened version, a contraction assumption on the update function. Roughly speaking, this means that if one starts two Markov chains at different starting points but uses the same random numbers as updates, then the points of the chain coincide or move closer to each other as the chain progresses. In this paper, we replace this assumption by the assumption that the Markov chain is uniformly ergodic. The concept of uniform ergodicity is much closer to the concept of discrepancy, which allows us to obtain stronger results than previous attempts. We introduce uniformly ergodic Markov chains in the following.

Let  $G \subseteq \mathbb{R}^d$  and let  $\mathcal{B}(G)$  denote the Borel  $\sigma$ -algebra of  $G$ . In the following, we provide the definition of a transition kernel.

DEFINITION 2. The function  $K : G \times \mathcal{B}(G) \rightarrow [0, 1]$  is called a transition kernel if:

- (i) for each  $x \in G$  the mapping  $A \in \mathcal{B}(G) \mapsto K(x, A)$  is a probability measure on  $(G, \mathcal{B}(G))$ , and
- (ii) 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.

Let  $K : G \times \mathcal{B}(G) \rightarrow [0, 1]$  be a transition kernel. We assume that  $\pi$  is the unique stationary distribution of the transition kernel  $K$ , that is,

$$\int_G K(x, A)\pi(dx) = \pi(A) \quad \forall A \in \mathcal{B}(G).$$

The transition kernel  $K$  gives rise to a Markov chain  $X_0, X_1, X_2, \dots \in G$  in the following way. Let  $X_0 = x$  with  $x \in G$  and  $i \in \mathbb{N}$ . Then, for a given  $X_{i-1}$ , we choose  $X_i$  with distribution  $K(X_{i-1}, \cdot)$ , that is, for all  $A \in \mathcal{B}(G)$ , the probability that  $X_i \in A$  is given by  $K(X_{i-1}, A)$ .

DEFINITION 3 (Total variation distance). The total variation distance between the transition kernel  $K(x, \cdot)$  and the stationary distribution  $\pi$  is defined by

$$\|K^j(x, \cdot) - \pi\|_{\text{tv}} = \sup_{A \in \mathcal{B}(G)} |K^j(x, A) - \pi(A)|.$$

Note that with  $K^0(x, A) = 1_{x \in A}$  we have

$$K^j(x, A) = \int_G K(y, A)K^{j-1}(x, dy) = \int_G K^{j-1}(y, A)K(x, dy).$$

DEFINITION 4 (Uniform ergodicity). Let  $\alpha \in [0, 1)$  and  $M \in (0, \infty)$ . The transition kernel  $K$  is uniformly ergodic with  $(\alpha, M)$  iff for any  $x \in G$  and  $j \in \mathbb{N}$  we have

$$\|K^j(x, \cdot) - \pi\|_{\text{tv}} \leq \alpha^j M.$$

A Markov chain with transition kernel  $K$  is called uniformly ergodic if there exists an  $\alpha \in [0, 1)$  and  $M \in (0, \infty)$ , such that the transition kernel is uniformly ergodic with  $(\alpha, M)$ .

REMARK 1. The uniform ergodicity is necessary to apply the Hoeffding inequality [11] and it is also used in the estimates of the discrepancy.

For Markov chains that are geometrically ergodic or have a spectral gap (for definitions, see [28, 30, 31]), one must use other concentration inequalities. The papers [1, 22, 26] might be useful. However, in those cases we are not aware of results which allow us to treat Markov chains which start deterministically, as we consider in this paper.

Let us state a result which provides an equivalent statement to uniform ergodicity. Let  $L_\infty$  be the set of all bounded functions  $f: G \rightarrow \mathbb{R}$ . Then define the operator  $P^j: L_\infty \rightarrow L_\infty$  by

$$P^j f(x) = \int_G f(y) K^j(x, dy),$$

and the expectation with respect to  $\pi$  is denoted by  $\mathbb{E}_\pi(f) = \int_G f(y) \pi(dx)$ . The following result is well known; for a proof of this fact see, for example, [31], Proposition 3.23, page 48.

**PROPOSITION 1.** *Let  $\alpha \in [0, 1)$  and  $M \in (0, \infty)$ . Then the following statements are equivalent:*

- (i) *The transition kernel  $K$  is uniformly ergodic with  $(\alpha, M)$ .*
- (ii) *The operator  $P^j - \mathbb{E}_\pi$  satisfies*

$$\|P^j - \mathbb{E}_\pi\|_{L_\infty \rightarrow L_\infty} \leq 2M\alpha^j, \quad j \in \mathbb{N}.$$

In the following, we introduce update functions  $\varphi$  for a given transition kernel and state some examples.

**DEFINITION 5 (Update function).** Let  $\varphi: G \times [0, 1]^s \rightarrow G$  be a measurable function and

$$B: G \times \mathcal{B}(G) \rightarrow \mathcal{B}([0, 1]^s),$$

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

where  $\mathcal{B}([0, 1]^s)$  is the Borel  $\sigma$ -algebra of  $[0, 1]^s$ . Let  $\lambda_s$  denote the Lebesgue measure on  $\mathbb{R}^s$ . Then the function  $\varphi$  is an update function for the transition kernel  $K$  iff

$$(1) \quad K(x, A) = \mathbb{P}(\varphi(x; U) \in A) = \lambda_s(B(x, A)),$$

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

**EXAMPLE 1 (Direct simulation).** Let us assume that we can sample with respect to  $\pi$ , that is,  $K(x, A) = \pi(A)$  for all  $x \in G$ . For the moment, let  $G = [0, 1]^s$  and let  $\pi$  be the uniform distribution on  $G$ . In this case, we can choose the simple update function  $\varphi(x; u) = u$ , since then

$$\pi(A) = \lambda_s(B(x, A)) \quad \text{for all } x \in G.$$

If  $G$  is a general subset of  $\mathbb{R}^d$  and  $\pi$  is a general probability measure, then we need a generator; see [6]. A generator is a special update function  $\psi: [0, 1]^s \rightarrow G$  such that

$$\pi(A) = \mathbb{P}(\psi(U) \in A), \quad \text{for all } A \in \mathcal{B}(G).$$

Note that the transition kernel  $K(x, A) = \pi(A)$  is uniformly ergodic with  $(\alpha, M)$  for  $\alpha = 0$  and  $M \in (0, \infty)$ .

EXAMPLE 2 (Hit-and-run algorithm). Let  $G \subset \mathbb{R}^d$  be a compact convex body and  $\pi$  be the uniform distribution on  $G$ . Let  $\mathbb{S}^{d-1} = \{x \in \mathbb{R}^d: \|x\|_2 = \langle x, x \rangle^{1/2} = 1\}$  be the  $d - 1$ -dimensional sphere, where  $\langle x, y \rangle$  denotes the standard inner product in  $\mathbb{R}^d$ . Let  $\theta \in \mathbb{S}^{d-1}$  and let  $L(x, \theta)$  be the chord in  $G$  through  $x$  and  $x + \theta$ , that is,

$$L(x, \theta) = \{x + s\theta \in \mathbb{R}^d \mid s \in \mathbb{R}\} \cap G.$$

We assume that we have an oracle which gives us  $a(x, \theta), b(x, \theta) \in G$ , such that

$$[a(x, \theta), b(x, \theta)] = L(x, \theta),$$

where  $[a(x, \theta), b(x, \theta)] = \{\lambda a(x, \theta) + (1 - \lambda)b(x, \theta): \lambda \in [0, 1]\}$ . A transition of the hit and run algorithm works as follows. First, choose a random direction  $\theta$ . Then we sample the next state on  $[a(x, \theta), b(x, \theta)]$  uniformly. Let  $\psi: [0, 1]^{d-1} \rightarrow \mathbb{S}^{d-1}$  be a generator for the uniform distribution on the sphere; see, for instance, [10]. Then we can choose for  $x \in G$  and  $u = (v_1, v_2, \dots, v_d) \in [0, 1]^d$  the update function

$$\varphi(x, u) = v_d a(x, \psi(v_1, \dots, v_{d-1})) + (1 - v_d) b(x, \psi(v_1, \dots, v_{d-1})).$$

In [32], it is shown that there exists an  $\alpha \in [0, 1)$  and an  $M \in (0, \infty)$ , such that the hit-and-run algorithm is uniformly ergodic with  $(\alpha, M)$ .

EXAMPLE 3 (Independence Metropolis sampler). Let  $G = [0, 1]^d$ , assume that a parameter  $\beta > 0$  and a function  $H: [0, 1]^d \rightarrow \mathbb{R}$  are given. Let  $\pi_\beta$  be a probability measure on  $G$ , given by

$$\pi_\beta(A) = \frac{1}{Z_\beta} \int_A \exp(-\beta H(x)) \, dx, \quad A \in \mathcal{B}(G),$$

with normalizing constant  $Z_\beta = \int_G \exp(-\beta H(y)) \, dy$ . Here,  $\pi_\beta$  might be interpreted as Boltzmann distribution with inverse temperature  $\beta$  and Hamiltonian  $H$ . Now let  $\bar{u} \in [0, 1]^d$  and

$$A(x; \bar{u}) = \min\{1, \exp(-\beta(H(\bar{u}) - H(x)))\}$$

be the acceptance probability of the Metropolis transition. Then we can choose for  $x \in [0, 1]^d$  and  $u = (v_1, v_2, \dots, v_{d+1}) \in [0, 1]^{d+1}$  the update function

$$\varphi(x; u) = \begin{cases} (v_1, \dots, v_d), & v_{d+1} \leq A(x; v_1, \dots, v_d), \\ x, & v_{d+1} > A(x; v_1, \dots, v_d). \end{cases}$$

In [20], Theorem 2.1, page 105, it is proven that, if there is a number  $\gamma > 0$  such that  $\exp(\beta \inf_{x \in G} H(x)) \geq \gamma$ , then the independent Metropolis sampler is uniformly ergodic with  $(1 - \gamma, 1)$ . A local proposal Metropolis algorithm can also be uniformly ergodic; see, for example, [19].

Note that the arguments of Example 3 are also valid for a heat-bath sampler. Let us briefly add some more examples. The slice sampler (for details with respect to the algorithm and update functions, see [24]) is under additional assumptions uniformly ergodic; see [23]. Furthermore, the Gibbs sampler for sampling the uniform distribution is uniformly ergodic if the boundary of  $G$  is smooth enough; see [29].

Above we defined the set  $B(x, A)$ , which is for  $x \in G$  and  $A \in \mathcal{B}(G)$  the set of random numbers  $u$  which takes  $x$  into the set  $A$  using the update function  $\varphi$  with arguments  $x$  and  $u$ . We now define sets of random numbers which take  $x$  to  $A$  in  $i \in \mathbb{N}$  steps. Let  $\varphi_1(x; u) = \varphi(x; u)$  and for  $i > 1$  let

$$\begin{aligned} \varphi_i &: G \times [0, 1]^{is} \rightarrow G, \\ \varphi_i(x; u_1, u_2, \dots, u_i) &= \varphi(\varphi_{i-1}(x; u_1, u_2, \dots, u_{i-1}); u_i), \end{aligned}$$

that is,  $\varphi_i(x; u_1, u_2, \dots, u_i) \in G$  is the point obtained via  $i$  updates using  $u_1, u_2, \dots, u_i \in [0, 1]^s$  where the starting point is  $x \in G$ .

LEMMA 1. *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*

$$(2) \quad \varphi_i(x; u_1, \dots, u_i) = \varphi_{i-j}(\varphi_j(x; u_1, \dots, u_j); u_{j+1}, \dots, u_i).$$

PROOF. The assertion can be proven by induction over  $i$ .  $\square$

For  $i \geq 1$ , let

$$\begin{aligned} B_i &: G \times \mathcal{B}(G) \rightarrow \mathcal{B}([0, 1]^{is}), \\ B_i(x, A) &= \{(u_1, u_2, \dots, u_i) \in [0, 1]^{is} : \varphi_i(x; u_1, u_2, \dots, u_i) \in A\}. \end{aligned}$$

We therefore have  $B_1(x, A) = B(x, A)$ . Note that  $B_i(x, A) \subseteq [0, 1]^{is}$ . The next lemma is important to understand the relation between the update function and the transition kernel.

LEMMA 2. *Let  $\varphi$  be an update function for the transition kernel  $K$ . Let  $n \in \mathbb{N}$  and  $F: G^n \rightarrow \mathbb{R}$ . The expectation with respect to the joint distribution of  $X_1, \dots, X_n$  from the Markov chain starting at  $x_0 \in G$  is given by*

$$\mathbb{E}_{x_0, K}(F(X_1, \dots, X_n)) = \underbrace{\int_G \dots \int_G}_{n\text{-times}} F(x_1, \dots, x_n) K(x_{n-1}, dx_n) \dots K(x_0, dx_1).$$

Then

$$(3) \quad \begin{aligned} \mathbb{E}_{x_0, K}(F(X_1, \dots, X_n)) \\ = \int_{[0, 1]^{ns}} F(\varphi_1(x_0, u_1), \dots, \varphi_n(x_0, u_1, \dots, u_n)) du_1 \dots du_n, \end{aligned}$$

whenever one of the integrals exist.

Note that the right-hand side of (3) is the expectation with respect to the uniform distribution in  $[0, 1]^{ns}$ .

PROOF. First, note that by the definition of the update function we obtain for any  $\pi$ -integrable function  $f: G \rightarrow \mathbb{R}$  that

$$(4) \quad \int_G f(y)K(x, dy) = \int_{[0,1]^s} f(\varphi(x, u)) du.$$

By the application of Lemma 1 and (4), we obtain

$$\begin{aligned} & \int_{[0,1]^{ns}} F(\varphi_1(x_0, u_1), \dots, \varphi_n(x_0, u_1, \dots, u_n)) du_1 \cdots du_n \\ &= \int_{[0,1]^{(n-1)s}} \int_G F(x_1, \varphi_1(x_1, u_2), \dots, \varphi_{n-1}(x_1, u_2, \dots, u_n)) \\ & \quad \times K(x_0, dx_1) du_2 \cdots du_n. \end{aligned}$$

The iteration of this procedure leads to the assertion.  $\square$

COROLLARY 1. *Let  $\varphi$  be an update function for the transition kernel  $K$  and let  $\pi$  be the stationary distribution of  $K$ . For any  $i \in \mathbb{N}$  and  $A \in \mathcal{B}(G)$ , we have  $K^i(x, A) = \lambda_{is}(B_i(x, A))$ . In particular,*

$$\int_G \lambda_{is}(B_i(x, A))\pi(dx) = \pi(A).$$

PROOF. Set for  $n \geq i$

$$F(x_1, \dots, x_n) = 1_{x_i \in A}.$$

Then by Lemma 2, we obtain  $K^i(x, A) = \lambda_{is}(B_i(x, A))$  and by the stationarity of  $\pi$  the proof is complete.  $\square$

**3. On the discrepancies of the Markov chain and driver sequence.** Recall that the star-discrepancy of a point set  $P_n = \{x_1, x_2, \dots, x_n\} \subseteq G$  with respect to the distribution  $\pi$  is given by

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

Let us assume that  $u_1, u_2, \dots, u_n \in [0, 1]^s$  is a finite deterministic sequence. We call this finite sequence driver sequence. Then let the set  $P_n = \{x_1, x_2, \dots, x_n\} \subseteq G$  be given by

$$(5) \quad x_i = \varphi(x_{i-1}; u_i) = \varphi_i(x_0; u_1, \dots, u_i), \quad i = 1, \dots, n.$$

We now define a discrepancy measure on the driver sequence. Below we show how this discrepancy is related to the discrepancy of the Markov chain.

DEFINITION 6 (Pull-back discrepancy). Let  $\mathcal{U}_n = (u_1, u_2, \dots, u_n) \in [0, 1]^{ns}$  and let  $B_i$  be defined as above. Define the local discrepancy function by

$$\Delta_{n,A,\varphi}^{\text{loc}}(x; u_1, \dots, u_n) = \frac{1}{n} \sum_{i=1}^n [1_{(u_1, \dots, u_i) \in B_i(x,A)} - \lambda_{is}(B_i(x,A))].$$

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

$$D_{\mathcal{A},\varphi}^*(\mathcal{U}_n) = \sup_{A \in \mathcal{A}} |\Delta_{n,A,\varphi}^{\text{loc}}(x; u_1, \dots, u_n)|.$$

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

The discrepancy of the driver sequence  $D_{\mathcal{A},\varphi}^*(\mathcal{U}_n)$  is a ‘‘pull-back discrepancy’’ since the test sets  $B_i(x, A)$  are derived from the test sets  $A \in \mathcal{A}$  from the discrepancy of the Markov chain  $D_{\mathcal{A},\pi}^*(P_n)$  via inverting the update function.

The following theorem provides an estimate of the star-discrepancy of  $P_n$  with respect to properties of the driver sequence and the transition kernel.

THEOREM 1. Let  $K$  be a transition kernel defined on  $G \subseteq \mathbb{R}^d$  with stationary distribution  $\pi$ . Let  $\varphi$  be an update function for  $K$ . Let  $x_0 = x$  and let  $u_1, u_2, \dots, u_n \in [0, 1]^s$  be the driver sequence, such that  $P_n$  is given by (5). Let  $\mathcal{A} \subseteq \mathcal{B}(G)$  be a set of test sets. Then

$$|D_{\mathcal{A},\pi}^*(P_n) - D_{\mathcal{A},\varphi}^*(\mathcal{U}_n)| \leq \sup_{A \in \mathcal{A}} \left| \frac{1}{n} \sum_{i=1}^n K^i(x, A) - \pi(A) \right|.$$

PROOF. For any  $A \in \mathcal{A}$ , we have

$$\begin{aligned} & \left| \frac{1}{n} \sum_{i=1}^n 1_{x_i \in A} - \pi(A) \right| \\ &= \left| \frac{1}{n} \sum_{i=1}^n [1_{(u_1, \dots, u_i) \in B_i(x,A)} - K^i(x, A) + K^i(x, A) - \pi(A)] \right| \\ &\leq \left| \frac{1}{n} \sum_{i=1}^n [1_{(u_1, \dots, u_i) \in B_i(x,A)} - \lambda_{is}(B_i(x, A))] \right| + \left| \frac{1}{n} \sum_{i=1}^n K^i(x, A) - \pi(A) \right|. \end{aligned}$$

Note that we used  $\lambda_{is}(B_i(x, A)) = K^i(x, A)$  which follows from Corollary 1. Hence,

$$D_{\mathcal{A},\pi}^*(P_n) \leq D_{\mathcal{A},\varphi}^*(\mathcal{U}_n) + \sup_{A \in \mathcal{A}} \left| \frac{1}{n} \sum_{i=1}^n K^i(x, A) - \pi(A) \right|.$$

The inequality

$$D_{\mathcal{A},\varphi}^*(\mathcal{U}_n) \leq D_{\mathcal{A},\pi}^*(P_n) + \sup_{A \in \mathcal{A}} \left| \frac{1}{n} \sum_{i=1}^n K^i(x, A) - \pi(A) \right|$$

follows by the same arguments.  $\square$

**COROLLARY 2.** *Let us assume that the conditions of Theorem 1 are satisfied. Further, let  $\alpha \in [0, 1)$  and  $M \in (0, \infty)$  and assume that the transition kernel is uniformly ergodic with  $(\alpha, M)$ . Then*

$$|D_{\mathcal{A},\pi}^*(P_n) - D_{\mathcal{A},\varphi}^*(\mathcal{U}_n)| \leq \frac{\alpha M}{n(1 - \alpha)}.$$

**PROOF.** By the uniform ergodicity with  $(\alpha, M)$ , we obtain

$$\begin{aligned} \left| \frac{1}{n} \sum_{i=1}^n K^i(x, A) - \pi(A) \right| &\leq \frac{1}{n} \sum_{i=1}^n |K^i(x, A) - \pi(A)| \\ &\leq \frac{1}{n} \sum_{i=1}^{\infty} \alpha^i M = \frac{\alpha M}{n(1 - \alpha)}. \end{aligned}$$

Then by Theorem 1 the assertion is proven.  $\square$

**REMARK 2.** In the setting of Example 1, where we assumed that  $G = [0, 1]^s$  and  $K(x, A) = \pi(A)$  we obtain that  $\alpha = 0$ . In this case we get the well-studied star-discrepancy for the uniform distribution on  $[0, 1]^s$ ; see, for instance, [9].

Theorem 1 gives an estimate of the star-discrepancy in terms of the discrepancy of the driver sequence and a quantity which depends on the transition kernel. We have seen in the previous corollary that for uniformly ergodic Markov chains we can further estimate the difference of the discrepancies  $D_{\mathcal{A},\pi}^*(P_n)$  and  $D_{\mathcal{A},\varphi}^*(\mathcal{U}_n)$ . Let us mention here that for geometrically ergodic transition kernel one can prove a similar bound.

In Corollary 5, we state a bound on  $D_{\mathcal{A},\pi}^*(P_n)$  of order  $\mathcal{O}(n^{-1/2}(\log n)^{1/2})$ , and by Corollary 2 we have that the pull-back discrepancy of the driver sequence satisfies the same convergence order.

From Corollary 2 and Theorem 3 in Appendix C, we now obtain the following Koksma–Hlawka inequality (cf. [9], Proposition 2.18) in terms of the pull-back discrepancy.

**COROLLARY 3** (Koksma–Hlawka inequality for uniformly ergodic Markov chains). *Let us assume that the conditions of Theorem 1 are satisfied. Further, let  $\alpha \in [0, 1)$  and  $M \in (0, \infty)$  and assume that the transition kernel is uniformly*

ergodic with  $(\alpha, M)$ . With a measure  $\rho$  on  $\mathbb{R}^d$  let  $H_1$  denote the space of functions  $f : G \rightarrow \mathbb{C}$  permitting the representation

$$f(x) = f_0 + \int_{\mathbb{R}^d} \mathbf{1}_{(-\infty, z)_G}(x) \tilde{f}(z) \rho(dz)$$

for some  $f_0 \in \mathbb{C}$  and  $\tilde{f} \in L_1(\mathbb{R}^d, \rho)$  with finite

$$\|f\|_{H_1} = |f_0| + \int_{\mathbb{R}^d} |\tilde{f}(z)| \rho(dz).$$

Then for all  $f \in H_1$  we have

$$\left| \int_G f(x) \pi(dz) - \frac{1}{n} \sum_{i=1}^n f(x_i) \right| \leq \left( D_{\mathcal{A}, \varphi}^*(\mathcal{U}_n) + \frac{\alpha M}{n(1-\alpha)} \right) \|f\|_{H_1}.$$

Again, in the setting of Example 1 for direct simulation, we have  $\alpha = 0$  and we obtain the Koksma–Hlawka inequality

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

**4. On the existence of good driver sequences.** In this section, we show the existence of finite sequences  $\mathcal{U}_n = (u_1, u_2, \dots, u_n) \in [0, 1]^{ns}$  such that

$$D_{\mathcal{A}, \varphi}^*(\mathcal{U}_n) \quad \text{and} \quad D_{\mathcal{A}, \pi}^*(P_n)$$

converge to 0 if the transition kernel is uniformly ergodic and  $P_n$  is given by (5). The main result is proven for  $D_{\mathcal{A}, \pi}^*(P_n)$ . The result with respect to  $D_{\mathcal{A}, \varphi}^*(\mathcal{U}_n)$  holds by Theorem 1.

The concept of a  $\delta$ -cover will be useful (cf. [12] for a discussion of  $\delta$ -covers, bracketing numbers and Vapnik–Červonenkis dimension).

**DEFINITION 7.** Let  $\mathcal{A} \subseteq \mathcal{B}(G)$  be a set of test sets. A finite subset  $\Gamma_\delta \subseteq \mathcal{A}$  is called a  $\delta$ -cover of  $\mathcal{A}$  with respect to  $\pi$  if for every  $A \in \mathcal{A}$  there are sets  $C, D \in \Gamma_\delta$  such that

$$C \subseteq A \subseteq D$$

and

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

**REMARK 3.** The concept of a  $\delta$ -cover is motivated by the following result. Let us assume that  $\Gamma_\delta$  is a  $\delta$ -cover of  $\mathcal{A}$ . Then, for all  $\{z_1, \dots, z_n\}$ , the following discrepancy inequality holds:

$$\sup_{A \in \mathcal{A}} \left| \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{z_i \in A} - \pi(A) \right| \leq \max_{C \in \Gamma_\delta} \left| \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{z_i \in C} - \pi(C) \right| + \delta.$$

PROOF. Let  $A \in \mathcal{A}$  and  $B \subseteq A \subseteq C$  be such that  $\pi(C \setminus B) \leq \delta$ . Then

$$\frac{1}{n} \sum_{i=1}^n 1_{z_i \in A} - \pi(A) \leq \frac{1}{n} \sum_{i=1}^n 1_{z_i \in C} - \pi(C) + \delta$$

and

$$\frac{1}{n} \sum_{i=1}^n 1_{z_i \in A} - \pi(A) \geq \frac{1}{n} \sum_{i=1}^n 1_{z_i \in B} - \pi(B) - \delta.$$

Thus, the result follows.  $\square$

Let us introduce the notation  $\Delta_{n,A,\varphi,x} = \Delta_{n,A,\varphi}^{\text{loc}}(x; u_1, \dots, u_n)$  and note that

$$\begin{aligned} \Delta_{n,A,\varphi,x} &= \Delta_{n,A,\varphi}^{\text{loc}}(x; u_1, \dots, u_n) \\ (6) \quad &= \frac{1}{n} \sum_{i=1}^n [1_{(u_1, \dots, u_i) \in B_i(x,A)} - \pi(A)] \\ &= \frac{1}{n} \sum_{i=1}^n [1_{\varphi_i(x; u_1, \dots, u_i) \in A} - \pi(A)]. \end{aligned}$$

LEMMA 3. Let  $K$  be a transition kernel with stationary distribution  $\pi$ . Let  $\varphi$  be an update function of  $K$ . Let  $X_1, X_2, \dots, X_n$  be given by a Markov chain with transition kernel  $K$  and  $X_0 = x$ . Then for any  $A \in \mathcal{B}(G)$  and  $c > 0$ , we obtain

$$(7) \quad \mathbb{P}[|\Delta_{n,A,\varphi,x}| \geq c] = \mathbb{P}_{x,K} \left[ \left| \frac{1}{n} \sum_{i=1}^n 1_{X_i \in A} - \pi(A) \right| \geq c \right],$$

where  $\mathbb{P}$  is the probability measure for the uniform distribution in  $[0, 1]^{nS}$  and  $\mathbb{P}_{x,K}$  is the joint probability of  $X_1, \dots, X_n$  with  $X_0 = x$ .

PROOF. Let

$$J(A, c) = \left\{ (z_1, \dots, z_n) \in G^n : \left| \frac{1}{n} \sum_{i=1}^n 1_{z_i \in A} - \pi(A) \right| \geq c \right\}.$$

Set

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

By

$$\mathbb{E}_{x,K}(F(X_1, \dots, X_n)) = \mathbb{P}_{x,K}(J(A, c)),$$

Lemma 2 and (6) the assertion is proven.  $\square$

The next result follows from [11] and gives us a Hoeffding inequality for uniformly ergodic Markov chains. For the convenience of the reader, we provide a proof in Appendix A.

**PROPOSITION 2** (Hoeffding inequality for uniformly ergodic Markov chains). *Assume that the transition kernel  $K$  is uniformly ergodic with  $(\alpha, M)$ . Let  $X_1, X_2, \dots, X_n$  be given by a Markov chain with transition kernel  $K$  and  $X_0 = x$ . Then for any  $A \in \mathcal{B}(G)$  and  $c > 0$ , we obtain*

$$(8) \quad \mathbb{P}_{x,K} \left[ \left| \frac{1}{n} \sum_{i=1}^n 1_{X_i \in A} - \pi(A) \right| \geq c \right] \leq 2 \exp \left( - \frac{(1-\alpha)^2 (nc - 2M/(1-\alpha))^2}{M^2 8n} \right),$$

where  $n \geq \frac{4M}{(1-\alpha)c}$ .

**4.1. Monte Carlo rate of convergence.** We now show that for every starting point  $x_0$  and every  $n$  there exists a finite sequence  $u_1, u_2, \dots, u_n \in [0, 1]^s$  such that the discrepancy of the corresponding Markov chain 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 Markov chains to the local discrepancy function for a fixed test set to show that the probability of point sets with small local discrepancy is large. We then extend this result to the local discrepancy for all sets in the  $\delta$ -cover and finally to all test sets. Using Corollary 2, we are also able to obtain a result for the pull-back discrepancy of the driver sequence. In cases where there are  $\delta$ -covers with  $|\Gamma_\delta| \leq C\delta^{-\kappa}$  for some constants  $C, \kappa > 0$  independent of  $\delta$  (see Appendix B for an example), the proof of the following theorem shows, in particular, that if the finite driver sequence is chosen at random from the uniform distribution, the discrepancy of the induced point set  $P_n$  converges with high probability with almost the Monte Carlo rate.

**THEOREM 2.** *Let  $K$  be a transition kernel with stationary distribution  $\pi$  defined on a set  $G \subseteq \mathbb{R}^d$ . Assume that the transition kernel is uniformly ergodic with  $(\alpha, M)$ . Let  $\mathcal{A} \subseteq \mathcal{B}(G)$  be a set of test sets. Assume that for every  $\delta > 0$  there exists a set  $\Gamma_\delta \subseteq \mathcal{B}(G)$  with  $|\Gamma_\delta| < \infty$  such that  $\Gamma_\delta$  is a  $\delta$ -cover of  $\mathcal{A}$  with respect to  $\pi$ . Let  $\varphi$  be an update function for  $K$ . Then, for any  $x_0 = x$  there exists a driver sequence  $u_1, u_2, \dots, u_n \in [0, 1]^s$  such that  $P_n = \{x_1, \dots, x_n\}$  given by*

$$x_i = \varphi(x_{i-1}; u_i) = \varphi_i(x_0; u_1, \dots, u_i), \quad i = 1, \dots, n$$

satisfies

$$D_{\mathcal{A}, \pi}^*(P_n) \leq \frac{8M}{1-\alpha} \frac{\sqrt{\log |\Gamma_\delta|}}{\sqrt{n}} + \delta.$$

PROOF. Let  $A \in \mathcal{A}$  and  $x_0 = x \in G$ . By Lemma 3 and Proposition 2, we obtain for any  $c_n \geq \frac{4M}{n(1-\alpha)}$  that

$$(9) \quad \mathbb{P}[|\Delta_{x,n,A,\varphi}| \leq c_n] \geq 1 - 2 \exp\left(-\frac{(1-\alpha)^2(nc_n - 2M/(1-\alpha))^2}{M^2 8n}\right).$$

Let

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

Set  $m = |\widehat{\Gamma}_\delta|$ . If we have for all  $A \in \widehat{\Gamma}_\delta$  that

$$(10) \quad \mathbb{P}[|\Delta_{x,n,A,\varphi}| \leq c_n] > 1 - \frac{1}{m},$$

then there exists a finite sequence  $u_1, \dots, u_n \in [0, 1]^s$  such that

$$(11) \quad \max_{A \in \widehat{\Gamma}_\delta} |\Delta_{x,n,A,\varphi}| \leq c_n.$$

By (9), we obtain for

$$c_n = \frac{4M}{1-\alpha} \frac{\sqrt{2 \log(2m)}}{\sqrt{n}}$$

that (10) holds and we get the desired result for any  $A \in \widehat{\Gamma}_\delta$ . Now we extend the result from  $\widehat{\Gamma}_\delta$  to  $\mathcal{A}$ . For  $A \in \mathcal{A}$ , there are  $C, D \in \Gamma_\delta$  such that  $C \subseteq A \subseteq D$  and  $\pi(D \setminus C) \leq \delta$ , since  $\Gamma_\delta$  is a  $\delta$ -cover. Hence, we get

$$\begin{aligned} & \left| \frac{1}{n} \sum_{i=1}^n [1_{(u_1, \dots, u_i) \in B_i(x,A)} - \pi(A)] \right| \\ &= \left| \frac{1}{n} \sum_{i=1}^n [1_{(u_1, \dots, u_i) \in B_i(x,D)} - \pi(D)] \right. \\ & \quad \left. - \frac{1}{n} \sum_{i=1}^n [1_{(u_1, \dots, u_i) \in B_i(x,D \setminus A)} - \pi(D \setminus A)] \right| \\ &\leq \left| \frac{1}{n} \sum_{i=1}^n [1_{(u_1, \dots, u_i) \in B_i(x,D)} - \pi(D)] \right| \\ & \quad + \left| \frac{1}{n} \sum_{i=1}^n [1_{(u_1, \dots, u_i) \in B_i(x,D \setminus A)} - \pi(D \setminus A)] \right|. \end{aligned}$$

Let

$$I_1 := \left| \frac{1}{n} \sum_{i=1}^n [1_{(u_1, \dots, u_i) \in B_i(x,D)} - \pi(D)] \right|$$

and

$$I_2 := \left| \frac{1}{n} \sum_{i=1}^n [1_{(u_1, \dots, u_i) \in B_i(x, D \setminus A)} - \pi(D \setminus A)] \right|.$$

By  $D \in \widehat{\Gamma}_\delta$ , we have

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

Furthermore,

$$\begin{aligned} I_2 &= \left| \frac{1}{n} \sum_{i=1}^n 1_{(u_1, \dots, u_i) \in B_i(x, D \setminus A)} - \pi(D \setminus C) + \pi(D \setminus C) - \pi(D \setminus A) \right| \\ &\leq \left| \frac{1}{n} \sum_{i=1}^n [1_{(u_1, \dots, u_i) \in B_i(x, D \setminus C)} - \pi(D \setminus C)] \right| + |\pi(D \setminus C) - \pi(D \setminus A)| \\ &\leq c_n + \delta. \end{aligned}$$

The last inequality follows by the  $\delta$ -cover property, (11) and the fact that  $D \setminus C \in \widehat{\Gamma}_\delta$ . Finally, note that  $m = |\widehat{\Gamma}_\delta| \leq |\Gamma_\delta|^2/2$  which completes the proof.  $\square$

Using Corollary 2, we can also state Theorem 2 in terms of the driver sequence.

**COROLLARY 4.** *Let  $K$  be a transition kernel with stationary distribution  $\pi$  defined on a set  $G \subseteq \mathbb{R}^d$ . Assume that the transition kernel is uniformly ergodic with  $(\alpha, M)$ . Let  $\mathcal{A} \subseteq \mathcal{B}(G)$  be a set of test sets. Assume that for every  $\delta > 0$  there exists a set  $\Gamma_\delta \subseteq \mathcal{B}(G)$  with  $|\Gamma_\delta| < \infty$  such that  $\Gamma_\delta$  is a  $\delta$ -cover of  $\mathcal{A}$  with respect to  $\pi$ . Let  $\varphi$  be an update function for  $K$ . Then for any  $x_0 = x$  there exists a driver sequence  $u_1, u_2, \dots, u_n \in [0, 1]^s$  such that*

$$D_{\mathcal{A},\varphi}^*(\mathcal{U}_n) \leq \frac{8M}{1-\alpha} \frac{\sqrt{\log |\Gamma_\delta|}}{\sqrt{n}} + \delta + \frac{\alpha M}{n(1-\alpha)}.$$

Let  $P_n = \{x_1, \dots, x_n\}$  given by

$$x_i = \varphi(x_{i-1}; u_i) = \varphi_i(x_0; u_1, \dots, u_i), \quad i = 1, \dots, n.$$

Then  $P_n$  satisfies

$$D_{\mathcal{A},\pi}^*(P_n) \leq \frac{8M}{1-\alpha} \frac{\sqrt{\log |\Gamma_\delta|}}{\sqrt{n}} + \delta + \frac{2\alpha M}{n(1-\alpha)}.$$

This corollary has two consequences. One is the existence of a driver sequence with small pull-back discrepancy. The second is that if one can construct such a sequence with small pull-back discrepancy, then the Markov chain which one obtains using this driver sequence also has small discrepancy. Thus, the pull-back discrepancy is a sufficient criterion for the construction of good driver sequences.

Theorem 2 and Corollary 4 depend on  $\delta$  and the size of the  $\delta$ -cover  $\Gamma_\delta$ . For a certain set of test sets, we have the following result.

COROLLARY 5. Let  $K$  be a transition kernel with stationary distribution  $\pi$  defined on a set  $G \subseteq \mathbb{R}^d$ . Assume that the transition kernel is uniformly ergodic with  $(\alpha, M)$ . Let the set of test sets  $\mathcal{A} \subseteq \mathcal{B}(G)$  be given by

$$\mathcal{A} = \{(-\infty, x) \cap G \mid x \in \bar{\mathbb{R}}^d\},$$

where  $\bar{\mathbb{R}}^d = (\mathbb{R} \cup \{\infty, -\infty\})^d$ . Let  $\varphi$  be an update function for  $K$ . Then for any  $x_0 = x$  there exists a driver sequence  $u_1, u_2, \dots, u_n \in [0, 1]^s$  and an absolute constant  $c > 0$  such that

$$D_{\mathcal{A}, \varphi}^*(\mathcal{U}_n) \leq \frac{8M}{1 - \alpha} \frac{\sqrt{d \log(3 + 4c^2n)}}{\sqrt{n}} + \frac{\sqrt{d}}{\sqrt{n}} + \frac{\alpha M}{n(1 - \alpha)}.$$

Let  $P_n = \{x_1, \dots, x_n\}$  given by

$$x_i = \varphi(x_{i-1}; u_i) = \varphi_i(x_0; u_1, \dots, u_i), \quad i = 1, \dots, n.$$

Then  $P_n$  satisfies

$$D_{\mathcal{A}, \pi}^*(P_n) \leq \frac{8M}{1 - \alpha} \frac{\sqrt{d \log(3 + 4c^2n)}}{\sqrt{n}} + \frac{\sqrt{d}}{\sqrt{n}} + \frac{2\alpha M}{n(1 - \alpha)}.$$

PROOF. The result follows by Lemma 4 in Appendix B, which shows the existence of  $\delta$ -covers with

$$|\Gamma_\delta| \leq (4 + 3c^2d\delta^{-2})^d,$$

and Corollary 4.  $\square$

4.2. *Optimality of the Monte Carlo rate.* We now show that the exponent  $-1/2$  of  $n$  in Theorem 2 cannot be improved in general. We do so by specializing Theorem 2 to the sphere  $\mathbb{S}^d$ . Recall that  $\mathbb{S}^d = \{x \in \mathbb{R}^{d+1} : \|x\|_2 = \langle x, x \rangle^{1/2} = 1\}$ , where  $\langle x, y \rangle$  denotes the standard inner product in  $\mathbb{R}^{d+1}$ . A spherical cap  $C(x, t) \subseteq \mathbb{S}^d$  with center  $x \in \mathbb{S}^d$  and  $-1 \leq t \leq 1$  is given by

$$C(x, t) = \{y \in \mathbb{S}^d : \langle x, y \rangle > t\}.$$

Let  $\mathcal{C} = \{C(x, t) : x \in \mathbb{S}^d, -1 \leq t \leq 1\}$  be the test set of spherical caps of  $\mathbb{S}^d$ . The normalized area of a spherical cap  $C(x, t)$  for  $0 \leq t \leq 1$  is given by

$$\pi(C(x, t)) = \frac{1}{2} \frac{B(1 - t^2; d/2, 1/2)}{B(1; d/2, 1/2)},$$

where  $B$  is the incomplete beta function

$$B(1 - t^2; d/2, 1/2) = \int_0^{1-t^2} z^{d/2-1} (1 - z)^{-1/2} dz.$$

Then the spherical cap discrepancy of a point set  $P_n = \{x_1, x_2, \dots, x_n\} \subseteq \mathbb{S}^d$  is given by

$$D_{\mathbb{S}^d, \mathcal{C}}^*(P_n) = \sup_{C \in \mathcal{C}} \left| \frac{1}{n} \sum_{i=1}^n 1_{x_i \in C} - \pi(C) \right|.$$

The following result is an application of Theorem 2. For a proof of the corollary, we refer to Appendix D.

**COROLLARY 6.** *There exists an absolute constant  $c > 0$  independent of  $n$  and  $d$  such that for each  $n$  and  $d$  there exists a set of points  $P_n = \{x_1, x_2, \dots, x_n\} \subseteq \mathbb{S}^d$  such that the spherical cap discrepancy satisfies*

$$D_{\mathbb{S}^d, \mathcal{C}}^*(P_n) \leq c \frac{\sqrt{d} + \sqrt{(d+1) \log n}}{\sqrt{n}}.$$

We have shown the existence of points on  $\mathbb{S}^d$  for which the spherical cap discrepancy is of order  $\sqrt{\frac{\log n}{n}}$ . Since this result follows by specializing Theorem 2 to the sphere, any improvement of the exponent  $-1/2$  of  $n$  in Theorem 2 would yield an improvement of the exponent of  $n$  in Corollary 6. However, it is known that the spherical cap discrepancy of any point set is at least  $n^{-1/2-1/(2d)}$ ; see [3]. Thus, an exponent smaller than  $-1/2$  in Corollary 6 would yield a contradiction to the lower bound on the spherical cap discrepancy for large enough  $d$ . Thus, at this level of generality, the exponent of  $n$  in Theorem 2 cannot be improved.

We point out that a bound on the spherical cap discrepancy can also be deduced from [14], Theorem 4, by using a bound on the Vapnik–Červonenkis dimension for  $\mathcal{C}$ .

### APPENDIX A: PROOF OF PROPOSITION 2

Since the statement and the assumptions in Proposition 2 are slightly different from those in [11], Theorem 2, we prove the desired Hoeffding inequality by following the arguments in [11]. We set  $f(x) = 1_{x \in A} - \pi(A)$  and obtain by the uniform ergodicity

$$P^n f(x) = K^n(x, A) - \pi(A) \leq \min\{1, \alpha^n M\}.$$

Let  $g(x) = \sum_{n=0}^\infty P^n f(x)$  and note that  $\|g\|_\infty \leq M/(1 - \alpha)$ . Here, let us mention that even

$$\|g\|_\infty \leq 1 + (1 - \alpha)^{-1} + \frac{\log M}{\log \alpha^{-1}},$$

which improves upon the dependence of  $M$ . We have  $g(x) - Pg(x) = f(x)$  for all  $x \in G$ . Let  $D(X_{j-1}, X_j) = g(X_j) - Pg(X_{j-1})$  for  $j = 2, \dots, n + 1$ , then

$$\begin{aligned} \sum_{j=1}^n 1_{X_j \in A} - n \cdot \pi(A) &= \sum_{j=1}^n f(X_j) = \sum_{j=1}^n (g(X_j) - Pg(X_j)) \\ &= g(X_1) - g(X_{n+1}) + \sum_{j=2}^{n+1} D(X_{j-1}, X_j) \\ &\leq 2\|g\|_\infty + \sum_{j=2}^{n+1} D(X_{j-1}, X_j). \end{aligned}$$

Thus, for  $\theta > 0$  we obtain

$$\begin{aligned} \mathbb{E}_{x,K} \left[ \exp \left( \theta \sum_{j=1}^n 1_{X_j \in A} - \theta n \cdot \pi(A) \right) \right] \\ \leq \exp(2\theta \|g\|_\infty) \mathbb{E}_{x,K} \left[ \exp \left( \theta \sum_{j=2}^{n+1} D(X_{j-1}, X_j) \right) \right] \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}_{x,K} \left[ \exp \left( \theta \sum_{j=2}^{n+1} D(X_{j-1}, X_j) \right) \right] \\ = \mathbb{E}_{x,K} \left[ \exp \left( \theta \sum_{j=2}^n D(X_{j-1}, X_j) \right) \right. \\ \left. \times \mathbb{E}_{x,K} (\exp(\theta D(X_n, X_{n+1})) \mid X_1, \dots, X_n) \right]. \end{aligned}$$

By virtue of [8], Lemma 8.1 (see also [11], equation (5)), we obtain

$$(12) \quad \mathbb{E}_{x,K} (\exp(\theta D(X_n, X_{n+1})) \mid X_1, \dots, X_n) \leq \exp(\theta^2 \|g\|_\infty^2 / 2).$$

This can be repeated iteratively so that

$$\mathbb{E}_{x,K} \left[ \exp \left( \theta \sum_{j=1}^n 1_{X_j \in A} - \theta n \cdot \pi(A) \right) \right] \leq \exp(2\theta \|g\|_\infty + n\theta^2 \|g\|_\infty^2 / 2).$$

Markov's inequality leads to

$$\mathbb{P}_{x,K} \left[ \frac{1}{n} \sum_{j=1}^n 1_{X_j \in A} - \pi(A) \geq c \right] \leq \exp(-\theta n c + 2\theta \|g\|_\infty + n\theta^2 \|g\|_\infty^2 / 2).$$

The bound is best possible for

$$\theta = \frac{nc - 2\|g\|_\infty}{n\|g\|_\infty^2}.$$

Finally, by  $\|g\|_\infty \leq M/(1 - \alpha)$  and by repeating this analysis with  $f = \pi(A) - 1_{x \in A}$  we obtain the assertion.

APPENDIX B: DELTA-COVERS

We need a deep result of the theory of empirical processes which follows from Talagrand [34], Theorem 6.6 and Haussler [13], Corollary 1. For a more general version, see also [14], Theorem 4.

PROPOSITION 3. *There exists an absolute constant  $c > 0$  such that for each cumulative distribution function  $L$  on  $(G, \mathcal{B}(G))$  the following holds: For all  $r \in \mathbb{N}$  there exist  $y_1, \dots, y_r \in G$  with*

$$\sup_{x \in \mathbb{Q}^d} \left| L(x) - \frac{1}{r} \sum_{i=1}^r 1_{y_i \in (-\infty, x)_G} \right| \leq c\sqrt{dr}^{-1/2}.$$

In this subsection, we study  $\delta$ -covers in  $G$  with respect to the probability measure  $\pi$ .

LEMMA 4. *Let  $G \subseteq \mathbb{R}^d$  and let  $(G, \mathcal{B}(G), \pi)$  be a probability space where  $\mathcal{B}(G)$  is the Borel  $\sigma$ -algebra of  $G$ . Assume that  $\pi$  is absolutely continuous with respect to the Lebesgue measure. Define the set  $\mathcal{A} \subseteq \mathcal{B}(G)$  of test sets by*

$$\mathcal{A} = \{(-\infty, x)_G : x \in \bar{\mathbb{R}}^d\}.$$

Then for any  $\delta > 0$  there exists a  $\delta$ -cover  $\Gamma_\delta$  of  $\mathcal{A}$  with

$$|\Gamma_\delta| \leq (3 + 4c^2d\delta^{-2})^d,$$

where  $c > 0$  is an absolute constant.

PROOF. Let  $\delta > 0$  be given and let  $r \in \mathbb{N}$  be the smallest integer such that  $2c\sqrt{dr}^{-1/2} \leq \delta$ . By Proposition 3, there are points  $y_1, \dots, y_r \in G$  such that

$$(13) \quad \sup_{x \in \bar{\mathbb{R}}^d} \left| \pi((-\infty, x)_G) - \frac{1}{r} \sum_{i=1}^r 1_{y_i \in (-\infty, x)_G} \right| \leq c\sqrt{dr}^{-1/2} \leq \frac{\delta}{2}.$$

Let  $y_i = (\eta_{i,1}, \dots, \eta_{i,d})$ . We now define the set

$$\Gamma_\delta = \left\{ \prod_{j=1}^d (-\infty, z_j) \cap G : z_j \in \{-\infty, \infty, \eta_{1,j}, \eta_{2,j}, \dots, \eta_{r,j}\} \text{ for } 1 \leq j \leq d \right\}.$$

The cardinality of  $\Gamma_\delta$  satisfies

$$|\Gamma_\delta| = (2 + r)^d \leq (3 + 4c^2 d \delta^{-2})^d.$$

It remains to show that  $\Gamma_\delta$  is a  $\delta$ -cover of  $\mathcal{A}$ .

Let  $z \in \mathbb{R}^d$  be arbitrary. Then there exist  $(-\infty, x)_G, (-\infty, y)_G \in \Gamma_\delta$  such that

$$(-\infty, x]_G \subseteq (-\infty, z)_G \subseteq (-\infty, y)_G$$

and

$$\begin{aligned} (-\infty, x]_G \cap \{y_1, \dots, y_r\} &= (-\infty, z)_G \cap \{y_1, \dots, y_r\} \\ &= (-\infty, y)_G \cap \{y_1, \dots, y_r\}. \end{aligned}$$

Using (13), we obtain

$$\begin{aligned} &\pi((-\infty, y)_G \setminus (-\infty, x)_G) \\ &\leq \left| \pi((-\infty, y)_G) - \frac{1}{r} \sum_{i=1}^r 1_{y_i \in (-\infty, y)_G} \right| \\ &\quad + \left| \pi((-\infty, x]_G) - \frac{1}{r} \sum_{i=1}^r 1_{y_i \in (-\infty, x]_G} \right| \\ &\leq \delta. \end{aligned}$$

Thus,  $\Gamma_\delta$  is a  $\delta$ -cover.  $\square$

### APPENDIX C: INTEGRATION ERROR

In Appendix B, we considered test sets which are intersections of boxes with the state space  $G$ . We define a reproducing kernel  $Q$  by

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

where  $\rho$  is a measure on  $\mathbb{R}^d$  with  $\int_{\mathbb{R}^d} \rho(dz) < \infty$ . The function  $Q$  is symmetric  $Q(x, y) = Q(y, x)$  and positive semi-definite, that is, for any  $x_1, \dots, x_n \in G$  and complex numbers  $b_1, \dots, b_n \in \mathbb{C}$  we have

$$\sum_{k, \ell=1}^n b_k \bar{b}_\ell Q(x_k, x_\ell) = \left| \sum_{k=1}^n b_k \right|^2 + \int_{\mathbb{R}^d} \left| \sum_{k=1}^n b_k 1_{(-\infty, z)_G}(x_k) \right|^2 \rho(dz) \geq 0,$$

where  $\bar{b}_\ell$  denotes the complex conjugate of  $b_\ell$ . Thus,  $Q$  uniquely defines a reproducing kernel Hilbert space  $H_2 = H_2(Q)$  of functions defined on  $G$ . See [2] for more information on reproducing kernels and reproducing kernel Hilbert spaces. In fact, the functions  $f$  in  $H_2$  permit the representation

$$(14) \quad f(x) = f_0 + \int_{\mathbb{R}^d} 1_{(-\infty, z)_G}(x) \tilde{f}(z) \rho(dz),$$

for some  $f_0 \in \mathbb{C}$  and function  $\tilde{f} \in L_2(\mathbb{R}^d, \rho)$ , which can for instance be shown using the same arguments as in [4], Appendix A. The inner product in  $H_2$  is given by

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

With these definitions, we have the reproducing property

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

For  $1 \leq q \leq \infty$ , we also define the space  $H_q$  of functions of the form (14) for which  $\tilde{f} \in L_q(\mathbb{R}^d, \rho)$ , with norm

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

We provide a simple example.

EXAMPLE 4. Let  $G = [0, 1]$  and let  $\rho$  be the Lebesgue measure, then

$$Q(x, y) = 1 + \int_0^1 1_{[0, z)}(x) 1_{[0, z)}(y) dz = 1 + \min\{1 - x, 1 - y\}.$$

The function  $\tilde{f} = -f'$ , where  $f'$  is the usual derivative of  $f$ , and (14) is then  $f(x) = f_0 + \int_0^1 1_{[0, z)}(x) \tilde{f}(z) dz = f_0 - \int_x^1 f'(z) dz = f_0 + f(x) - f(1)$ . Thus,  $f(1) = f_0$  and  $H_q$  is the space of all absolutely continuous functions  $f$  for which  $f' \in L_q([0, 1], \rho)$ .

We have the following result concerning the integration error in  $H_q$ .

THEOREM 3. Let  $G \subseteq \mathbb{R}^d$  and  $\pi$  be a probability measure on  $G$ . Further, let  $\mathcal{A} = \{(-\infty, x)_G : x \in G\}$ . We assume that  $1 \leq p, q \leq \infty$  with  $1/p + 1/q = 1$ . Then for  $P_n = \{x_1, x_2, \dots, x_n\} \subseteq G$  and for all  $f \in H_q$  we have

$$\left| \int_G f(z) \pi(dz) - \frac{1}{n} \sum_{i=1}^n f(x_i) \right| \leq \|f\|_{H_q} D_{p, \mathcal{A}, \pi}^*(P_n),$$

where

$$D_{p, \mathcal{A}, \pi}^*(P_n) = \left( \int_{\mathbb{R}^d} \left| \int_G 1_{(-\infty, z)_G}(y) \pi(dy) - \frac{1}{n} \sum_{i=1}^n 1_{(-\infty, z)_G}(x_i) \right|^p \rho(dz) \right)^{1/p},$$

and for  $p = \infty$  let

$$D_{\mathcal{A}, \pi}^*(P_n) := D_{\infty, \mathcal{A}, \pi}^*(P_n) = \sup_{z \in G} \left| \int_G 1_{(-\infty, z)_G}(y) \pi(dy) - \frac{1}{n} \sum_{i=1}^n 1_{(-\infty, z)_G}(x_i) \right|.$$

PROOF. Let

$$e(f, P_n) = \int_G f(z)\pi(\mathrm{d}z) - \frac{1}{n} \sum_{i=1}^n f(x_i)$$

denote the quadrature error when approximating the integral  $\int_G f(z)\pi(\mathrm{d}z)$  by  $\frac{1}{n} \sum_{i=1}^n f(x_i)$  where  $P_n = \{x_1, x_2, \dots, x_n\}$ .

Let  $h(x) = \int_G Q(x, y)\pi(\mathrm{d}y) - \frac{1}{n} \sum_{i=1}^n Q(x, x_i)$ , then we have

$$h(x) = \int_{\mathbb{R}^d} 1_{(-\infty, z)_G}(x) \left( \int_G 1_{(-\infty, z)_G}(y)\pi(\mathrm{d}y) - \frac{1}{n} \sum_{i=1}^n 1_{(-\infty, z)_G}(x_i) \right) \rho(\mathrm{d}z)$$

and, therefore,  $h \in H_p$  for any  $1 \leq p \leq \infty$ . Let

$$\tilde{h}(z) = \int_G 1_{(-\infty, z)_G}(y)\pi(\mathrm{d}y) - \frac{1}{n} \sum_{i=1}^n 1_{(-\infty, z)_G}(x_i).$$

Further, for  $f \in H_q$  we have  $\tilde{f} \in L_q(\mathbb{R}^d, \rho)$ , and thus

$$e(f, P_n) = \int_{\mathbb{R}^d} \tilde{f}(z)\tilde{h}(z)\rho(\mathrm{d}z).$$

Using Hölder's inequality, we have

$$\begin{aligned} |e(f, P_n)| &\leq \int_{\mathbb{R}^d} |\tilde{f}(z)| |\tilde{h}(z)| \rho(\mathrm{d}z) \\ &\leq \left( \int_{\mathbb{R}^d} |\tilde{f}(z)|^q \rho(\mathrm{d}z) \right)^{1/q} \left( \int_{\mathbb{R}^d} |\tilde{h}(z)|^p \rho(\mathrm{d}z) \right)^{1/p}, \end{aligned}$$

where  $1 \leq p, q \leq \infty$  are Hölder conjugates  $1/p + 1/q = 1$ , with the obvious modifications for  $p, q = \infty$ . Thus, the result follows.  $\square$

Thus, we can use the bounds from the theorems above to obtain a bound on the integration error  $|e(f, P_n)|$ , where  $P_n$  is the set of points from the Markov chain, for functions  $f$  with representation (14) and  $\|f\|_{H_1} < \infty$ .

#### APPENDIX D: DELTA-COVERS FOR THE SPHERE

We use Theorem 2 where  $\pi$  is the normalized Lebesgue surface measure on the sphere  $\mathbb{S}^d$ . Let  $\psi : [0, 1]^d \rightarrow \mathbb{S}^d$  be an area-preserving mapping from  $[0, 1]^d$  to  $\mathbb{S}^d$  (i.e., a generator function) (see [10]), and let the update function  $\varphi : \mathbb{S}^d \times [0, 1]^d \rightarrow \mathbb{S}^d$  be given by

$$\varphi(x, u) = \psi(u).$$

The transition kernel is given by  $K(x, A) = \pi(A)$  which is uniformly ergodic with  $(\alpha, M)$  for  $\alpha = 0$  and  $M = 1$ .

In order to obtain a bound on the spherical cap discrepancy using Theorem 2, it remains to construct a  $\delta$ -cover on  $\mathbb{S}^d$  of suitable size. We construct a  $\delta$ -cover  $\Gamma_\delta$  by specifying a set of centers and heights in the following.

LEMMA 5. Let  $\mathbb{S}^d \subseteq \mathbb{R}^{d+1}$  denote the  $d$ -dimensional sphere. Let  $\mathcal{C} = \{C(x, t) : x \in \mathbb{S}^d, -1 \leq t \leq 1\}$  denote the set of spherical caps of  $\mathbb{S}^d$ . Then for any  $\delta > 0$  there exists a  $\delta$ -cover  $\Gamma_\delta$  of  $\mathcal{C}$  with respect to the normalized surface Lebesgue measure on  $\mathbb{S}^d$  with  $|\Gamma_\delta| \leq c d^{d+1} \delta^{-2(d+1)}$ , where  $c > 0$  is a constant independent of  $d$  and  $\delta$ .

The result of Corollary 6 follows now from Theorem 2 and Lemma 5 by setting  $\delta = d^{1/2} n^{-1/2}$ . The remainder of this subsection is concerned with the proof of Lemma 5.

Let  $y_1, y_2, \dots, y_N \in \mathbb{S}^d$  be given such that

$$(15) \quad \sup_{x \in \mathbb{S}^d} \min_{1 \leq i \leq N} \|x - y_i\|_2 \leq c_d N^{-1/d},$$

where  $c_d > 0$  is a constant depending only on  $d$ . The existence of such point sets follows, for instance, from [17]. Therein an equal area partition of  $\mathbb{S}^d$  into  $N$  parts was shown with diameter bounded by  $c_d N^{-1/d}$ . Thus, by taking one point in each partition we obtain (15). Indeed, from the proof of [17], Theorem 2.6, we obtain that the constant  $c_d$  can be chosen as

$$c_d = 8 \left( \frac{d \sqrt{\pi} \Gamma(d/2)}{\Gamma((d+1)/2)} \right)^{1/d} \leq 8 d^{1/d} \pi^{1/(2d)} \leq 8 \cdot 3^{1/3} \sqrt{\pi} < 21,$$

where  $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$  denotes the Gamma function. For  $x, y \in \mathbb{S}^d$ , we have  $\|x - y\|_2^2 = 2(1 - \langle x, y \rangle)$ .

Let  $v = \langle x, y \rangle$ . Then we obtain the following result.

LEMMA 6. We have  $C(x, t) \subseteq C(y, u)$  if and only if  $v = \langle x, y \rangle > u$  and

$$(16) \quad t^2 + u^2 + v^2 - 2tuv > 1.$$

PROOF. The condition  $v > u$  ensures that  $x \in C(y, u)$ . Let  $z \in C(x, t)$ , that is,  $\langle z, x \rangle > t$ . Then  $z \in C(y, u)$  if and only if  $\langle z, y \rangle > u$ . The point  $z$  is furthest from  $y$  (as measured by the Euclidean distance) if it lies on the great circle containing  $x$  and  $y$ . Assuming that  $x, y, z$  all lie on the same great circle such that  $x$  is between  $y$  and  $z$ , we have

$$\begin{aligned} \|y - z\| &= 2 \sin \left( \arcsin \frac{\|x - y\|}{2} + \arcsin \frac{\|x - z\|}{2} \right) \\ &= \|x - y\| \sqrt{1 - \|x - z\|^2/4} + \|x - z\| \sqrt{1 - \|x - y\|^2/4}. \end{aligned}$$

The result now follows by using  $\|x - y\|^2 = 2(1 - v)$  and  $\|x - z\|^2 < 2(1 - t)$ . □

The next lemma gives us a  $\delta$ -cover of  $\mathcal{C}$  with respect to  $\pi$ .

LEMMA 7. *Let*

$$N = \left\lceil \frac{35^d c_d^d}{B^{2d}(1; d/2, 1/2)\delta^{2d}} \right\rceil.$$

Let  $M = \lfloor N^{1/d}/c_d \rfloor$  and  $T = \{-1 + k/M : k = 0, 1, \dots, 2M\}$ . Then the set

$$\Gamma_\delta = \{C(y_i, t) : 1 \leq i \leq N, t \in T\}$$

is a  $\delta$ -cover of  $\mathcal{C}$  with respect to  $\pi$ .

PROOF. Let  $C(x, t) \in \mathcal{C}$  be an arbitrary spherical cap. Let  $y_i$  be such that  $\|x - y_i\|_2 \leq C_d N^{-1/d} \leq 1/M$ , thus

$$\langle x, y_i \rangle > 1 - \frac{1}{2M^2}.$$

Let  $u, w \in T$  be such that  $u + 2/M \leq t \leq w - 2/M$  and  $w - u \leq 5/M$ .

We now show that for this choice we have  $C(x, t) \subseteq C(y_i, u)$ . First, assume that  $u \geq 0$ . Then using (16) with  $v > 1 - 1/(2M^2)$  and  $t - u \geq 2/M$  we obtain

$$\begin{aligned} t^2 + u^2 + v^2 - 2tuv &\geq t^2 - 2tu + u^2 + v^2 \\ &\geq (t - u)^2 + (1 - 1/(2M^2))^2 \\ &\geq 4/M^2 + 1 - 1/M^2 + 1/(4M^4) > 1. \end{aligned}$$

Now assume that  $u < 0$ . Then

$$\begin{aligned} t^2 + u^2 + v^2 - 2tuv &\geq t^2 - 2tu + u^2 + v^2 - tu/M^2 \\ &\geq (t - u)^2 + (1 - 1/(2M^2))^2 - 1/M^2 \\ &\geq 4/M^2 + 1 - 1/M^2 + 1/(4M^4) - 1/M^2 > 1. \end{aligned}$$

Thus,  $C(x, t) \subseteq C(y_i, u)$ . We now show that  $C(y_i, w) \subseteq C(x, t)$ . If  $w = 1$ , we have  $C(y_i, w) = \emptyset$  in which case the result holds trivially. Thus, we can assume that  $t < 1 - 2/M$ , which implies that  $y_i \in C(x, t)$ . We use (16) again with  $v > 1 - 1/(2M^2)$  and  $|u - t| \geq 1/M$ . Thus, the result follows by the same arguments as in the previous case.

Thus, we have  $C(y_i, w) \subseteq C(x, t) \subseteq C(y_i, u)$  with  $w - u \leq 5/M$ . For  $w, u \geq 0$ , we have

$$\begin{aligned} &2B(1; d/2, 1/2)\pi(C(y_i, u) \setminus C(y_i, w)) \\ &= \int_{1-w^2}^{1-u^2} z^{d/2-1}(1-z)^{-1/2} dz \\ &\leq \int_{\max\{0, 1-u^2-10u/M-25/M^2\}}^{1-u^2} z^{d/2-1}(1-z)^{-1/2} dz \end{aligned}$$

$$\begin{aligned} &\leq \sup_{10/M+25/M^2 \leq r \leq 1} \int_{r-10/M-25/M^2}^r z^{d/2-1} (1-z)^{-1/2} dz \\ &\leq \int_{1-10/M-25/M^2}^1 (1-z)^{-1/2} dz \\ &\leq \sqrt{10/M + 25/M^2} \leq \sqrt{35/M}. \end{aligned}$$

Thus, in general we have

$$\pi(C(y_i, u) \setminus C(y_i, w)) \leq \frac{\sqrt{35/M}}{B(1; d/2, 1/2)} \leq \frac{\sqrt{35}}{B(1; d/2, 1/2)} \frac{\sqrt{c_d}}{N^{1/(2d)}}.$$

The last expression is bounded by  $\delta$  for

$$N = \left\lceil \frac{35^d c_d^d}{B^{2d}(1; d/2, 1/2) \delta^{2d}} \right\rceil. \quad \square$$

LEMMA 8. Assume that  $N^{1/d}/c_d > 1/2$  [otherwise (15) is trivial]. Then we have for  $\delta = d^{1/2}n^{-1/2}$  that there exists an absolute constant  $c > 0$ , such that

$$|\Gamma_\delta| \leq cn^{d+1}.$$

PROOF. We have

$$\begin{aligned} |\Gamma_\delta| &\leq N(2M + 1) \\ &\leq N(2N^{1/d}/c_d + 1) \\ &\leq 4N^{1+1/d}/c_d \\ &\leq \frac{8 \cdot 35^{d+1} c_d^d}{B^{2(d+1)}(1; d/2, 1/2) \delta^{2(d+1)}}. \end{aligned}$$

Thus, for  $\delta = d^{1/2}n^{-1/2}$  we obtain

$$|\Gamma_\delta| \leq \frac{n^{d+1}}{d^{d+1}} \frac{8 \cdot 35^{d+1} c_d^d}{B^{2(d+1)}(1; d/2, 1/2)} \leq \frac{n^{d+1}}{d^{d+1}} \frac{280 \cdot 735^d (\Gamma((d+1)/2))^{2(d+1)}}{(\Gamma(d/2)\Gamma(1/2))^{2(d+1)}},$$

where  $\Gamma$  is the Gamma function. Using Stirling’s formula for the Gamma function,

$$\Gamma(z) = \sqrt{\frac{2\pi}{z}} \left(\frac{z}{e}\right)^z (1 + \mathcal{O}(z^{-1}))$$

we obtain that there is an absolute constant  $c > 0$  such that  $|\Gamma_\delta| \leq cn^{d+1}$ .  $\square$

## REFERENCES

- [1] ADAMCZAK, R. (2008). A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.* **13** 1000–1034. [MR2424985](#)
- [2] ARONSZAJN, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc.* **68** 337–404. [MR0051437](#)
- [3] BECK, J. (1984). Sums of distances between points on a sphere—an application of the theory of irregularities of distribution to discrete geometry. *Mathematika* **31** 33–41. [MR0762175](#)
- [4] BRAUCHART, J. S. and DICK, J. (2013). A characterization of Sobolev spaces on the sphere and an extension of Stolarsky’s invariance principle to arbitrary smoothness. *Constr. Approx.* **38** 397–445. [MR3122277](#)
- [5] CHEN, S. (2011). Consistency and convergence rate of Markov chain quasi-Monte Carlo with examples. Ph.D. thesis, Stanford Univ., Stanford, CA.
- [6] CHEN, S., DICK, J. and OWEN, A. B. (2011). Consistency of Markov chain quasi-Monte Carlo on continuous state spaces. *Ann. Statist.* **39** 673–701. [MR2816335](#)
- [7] CHEN, S., MATSUMOTO, M., NISHIMURA, T. and OWEN, A. B. (2012). New inputs and methods for Markov chain quasi-Monte Carlo. In *Monte Carlo and Quasi-Monte Carlo Methods 2010. Springer Proc. Math. Stat.* **23** 313–327. Springer, Heidelberg. [MR3173841](#)
- [8] DEVROYE, L., GYÖRFI, L. and LUGOSI, G. (1996). *A Probabilistic Theory of Pattern Recognition. Applications of Mathematics (New York)* **31**. Springer, New York. [MR1383093](#)
- [9] DICK, J. and PILLICHSHAMMER, F. (2010). *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*. Cambridge Univ. Press, Cambridge. [MR2683394](#)
- [10] FANG, K.-T. and WANG, Y. (1994). *Number-Theoretic Methods in Statistics*. Chapman & Hall, London. [MR1284470](#)
- [11] GLYNN, P. W. and ORMONEIT, D. (2002). Hoeffding’s inequality for uniformly ergodic Markov chains. *Statist. Probab. Lett.* **56** 143–146. [MR1881167](#)
- [12] GNEWUCH, M. (2008). Bracketing numbers for axis-parallel boxes and applications to geometric discrepancy. *J. Complexity* **24** 154–172. [MR2400314](#)
- [13] HAUSSLER, D. (1995). Sphere packing numbers for subsets of the Boolean  $n$ -cube with bounded Vapnik–Chervonenkis dimension. *J. Combin. Theory Ser. A* **69** 217–232. [MR1313896](#)
- [14] HEINRICH, S., NOVAK, E., WASILKOWSKI, G. W. and WOŹNIAKOWSKI, H. (2001). The inverse of the star-discrepancy depends linearly on the dimension. *Acta Arith.* **96** 279–302. [MR1814282](#)
- [15] L’ECUYER, P., LECOT, C. and TUFFIN, B. (2008). A randomized quasi-Monte Carlo simulation method for Markov chains. *Oper. Res.* **56** 958–975.
- [16] LEMIEUX, C. and SIDORSKY, P. (2006). Exact sampling with highly uniform point sets. *Math. Comput. Modelling* **43** 339–349. [MR2214643](#)
- [17] LEOPARDI, P. (2009). Diameter bounds for equal area partitions of the unit sphere. *Electron. Trans. Numer. Anal.* **35** 1–16. [MR2582801](#)
- [18] LIAO, L. (1998). Variance reduction in Gibbs sampler using quasi random numbers. *J. Comput. Graph. Statist.* **7** 253–266.
- [19] MATHÉ, P. and NOVAK, E. (2007). Simple Monte Carlo and the Metropolis algorithm. *J. Complexity* **23** 673–696. [MR2372022](#)
- [20] MENGERSEN, K. L. and TWEEDIE, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* **24** 101–121. [MR1389882](#)
- [21] MEYN, S. and TWEEDIE, R. L. (2009). *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge Univ. Press, Cambridge. [MR2509253](#)
- [22] MIASOJEDOW, B. (2014). Hoeffding’s inequalities for geometrically ergodic Markov chains on general state space. *Statist. Probab. Lett.* **87** 115–120. [MR3168944](#)

- [23] MIRA, A. and TIERNEY, L. (2002). Efficiency and convergence properties of slice samplers. *Scand. J. Stat.* **29** 1–12. [MR1894377](#)
- [24] NEAL, R. M. (2003). Slice sampling. *Ann. Statist.* **31** 705–767. [MR1994729](#)
- [25] OWEN, A. B. and TRIBBLE, S. D. (2005). A quasi-Monte Carlo Metropolis algorithm. *Proc. Natl. Acad. Sci. USA* **102** 8844–8849 (electronic). [MR2168266](#)
- [26] PAULIN, D. (2015). Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electron. J. Probab.* **20** no. 79, 32. [MR3383563](#)
- [27] ROBERT, C. P. and CASELLA, G. (2004). *Monte Carlo Statistical Methods*, 2nd ed. Springer, New York. [MR2080278](#)
- [28] ROBERTS, G. O. and ROSENTHAL, J. S. (1997). Geometric ergodicity and hybrid Markov chains. *Electron. Commun. Probab.* **2** 13–25 (electronic). [MR1448322](#)
- [29] ROBERTS, G. O. and ROSENTHAL, J. S. (1998). On convergence rates of Gibbs samplers for uniform distributions. *Ann. Appl. Probab.* **8** 1291–1302. [MR1661176](#)
- [30] ROBERTS, G. O. and ROSENTHAL, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probab. Surv.* **1** 20–71. [MR2095565](#)
- [31] RUDOLF, D. (2012). Explicit error bounds for Markov chain Monte Carlo. *Dissertationes Math. (Rozprawy Mat.)* **485** 1–93. [MR2977521](#)
- [32] SMITH, R. L. (1984). Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions. *Oper. Res.* **32** 1296–1308. [MR0775260](#)
- [33] SOBOL, I. M. (1974). Pseudorandom numbers for the construction of discrete Markov chains by a Monte Carlo method. *USSR Compat. Math. Math. Phys.* **14** 36–45.
- [34] TALAGRAND, M. (1994). Sharper bounds for Gaussian and empirical processes. *Ann. Probab.* **22** 28–76. [MR1258865](#)
- [35] TRIBBLE, S. D. (2007). Markov chain Monte Carlo algorithms using completely uniformly distributed driving sequences. Ph.D. thesis, Stanford Univ. [MR2710331](#)
- [36] TRIBBLE, S. D. and OWEN, A. B. (2008). Construction of weakly CUD sequences for MCMC sampling. *Electron. J. Stat.* **2** 634–660. [MR2426105](#)

J. DICK  
H. ZHU  
SCHOOL OF MATHEMATICS AND STATISTICS  
UNIVERSITY OF NEW SOUTH WALES  
SYDNEY, NSW 2052  
AUSTRALIA  
E-MAIL: [josef.dick@unsw.edu.au](mailto:josef.dick@unsw.edu.au)  
[houying.zhu@unsw.edu.au](mailto:houying.zhu@unsw.edu.au)

D. RUDOLF  
INSTITUT FÜR MATHEMATIK  
UNIVERSITÄT JENA  
ERNST-ABBE-PLATZ 2  
07743 JENA  
GERMANY  
E-MAIL: [daniel.rudolf@uni-jena.de](mailto:daniel.rudolf@uni-jena.de)