DFG-Schwerpunktprogramm 1324

Discrepancy estimates for variance bounding Markov chain quasi-Monte Carlo

J. Dick, D. Rudolf

Preprint 165

Edited by

DFG-Schwerpunktprogramm 1324

Discrepancy estimates for variance bounding Markov chain quasi-Monte Carlo

J. Dick, D. Rudolf

Preprint 165

Discrepancy estimates for variance bounding Markov chain quasi-Monte Carlo

Josef Dick

School of Mathematics and Statistics

The University of New South Wales

Sydney, NSW 2052, Australia

Daniel Rudolf

Mathematical Institute

Friedrich Schiller University Jena

Ernst-Abbe-Platz 2, 07743 Jena, Germany

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 push-back discrepancy of the driver sequence and state a close relation to the star-discrepancy. 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. Finally we show under additional assumption that there even exists a driver sequence such that the discrepancies have a rate of convergence of almost n^{-1}.

1 Introduction

Markov chain Monte Carlo (MCMC) simulations are used in different branches of statistic 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 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 [LS06, Lia98, OT05, Sob74, Tri07]. In particular, Owen and Tribble [OT05] and Tribble [Tri07] 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 [CDO11] and Chen [Che11] 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 [CDO11, CMNO12, TO08]. Thus the sample average converges to the expected value but we do not know how fast this convergence takes place.

Recently, in [DRZ13] 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 [CDO11] and [DRZ13] rather strong conditions, the contraction assumption and uniform ergodicity, are imposed on the Markov chain.

We substantially extend the results of [DRZ13] to Markov chains which satisfy a much weaker convergence condition. Namely, we consider variance bounding Markov chains, introduced by Roberts and Rosenthal in [RR08], and show existence results of good driver sequences. In the following we describe the setting in detail and explain our main contributions.

1.1 Main results

The MCMC sampling can be represented via $X_{i+1} = \varphi(X_i, U_i)$ for $i \geq 2$, with $X_1 = \psi(U_1)$ and the $U_i \sim U(0,1)^s$ are i.i.d. The state X_i is an element in $G \subseteq \mathbb{R}^d$, the function $\varphi \colon G \times [0,1]^s \to G$ is called *update function* and $\psi \colon [0,1]^s \to G$ is called *generator function*. The update function corresponds to a transition kernel, say K. 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 [RR08]. 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

$$\Lambda = \sup\{\lambda \in \sigma(P - \mathbb{E}_\pi \mid L_2)\} \tag{1}$$

where $\sigma(P - \mathbb{E}_\pi \mid L_2)$ denotes the spectrum of $P - \mathbb{E}_\pi$ on L_2. For example let us consider the two state Markov chain which always jumps from one state to the other one. It is periodic and satisfies $\Lambda = -1$, thus it is variance bounding. With this toy example in mind let us point out that the Markov chain does not need to be uniformly or geometrically ergodic, it might even be periodic, and the distribution of X_i, for i arbitrarily large, is not necessarily close to π.

By 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 $x_i = \varphi(x_{i-1}, u_{i-1})$ where $i \geq 2$. 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 $A \subseteq G$. We define the star-discrepancy $D^*_{\mathscr{A},\pi}$ of $S_n = \{x_1, \ldots, x_n\}$ as the supremum of $|\pi(A) - \widehat{\pi}(A)|$ over all $A \in \mathscr{A}$, i.e.

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

3

where $\mathscr{A}$ denotes a certain set of subsets of G. By inverting the iterates of the update function we also define a push-back discrepancy of the driver sequence (the test sets are pushed 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 2) 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 5). 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 $(-\infty, x) = \Pi^d_{i=1}(-\infty, x_i)$. By the Koksma-Hlawka inequality (Theorem 5) we have

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

with $\|f\|_{H_1}$ defined in (24). Thus a bound on the discrepancy leads to an error bound for the approximation of $\mathbb{E}_\pi(f)$.

We show for all $n \ge 16$ 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

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

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 is defined in (1). For the details we refer to Corollary 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 last n_0 states, i.e. of $S_{[n_0,n]} = \{x_{n_0+1}, \ldots, x_{n_0+n}\}$. Under suitable convergence conditions of the Markov chain, for example the existence of an absolute L_2-spectral gap (see Definition 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 [Liu08]. However, it is not clear how to obtain suitable driver sequences which yield such an improvement. 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 deterministic MCMC 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. for all $A, B \in \mathcal{B}(G)$ holds

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

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

$$\int_G K(x,A)\,\pi(\mathrm{d}x) = \pi(A). \tag{3}$$

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

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

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 [Rud12].

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 $\|P - \mathbb{E}_\pi\|_{L_2 \to L_2} = \|P\|_{L_2^0 \to L_2^0}$, for details see [Rud12, Lemma 3.16, p. 44].

Definition 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 [Rud12, Corollary 3.15 and Lemma 3.21].

Proposition 1 *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*

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

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 (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*

$$\Lambda = \sup\{\lambda \in \mathrm{spec}(P \mid L_2^0)\} < 1, \tag{4}$$

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

7

For a motivation of the term variance bounding and a general treatment we refer to [RR08]. In particular by [RR08, 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 1 *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 (4).

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 [Rud91] or [Kre89, 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 1. $\qquad \square$

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 [DRZ13].

Definition 3 (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

$$K(x, A) = \mathbb{P}(\varphi(x; U) \in A) = \lambda_s(B(x, A)), \tag{5}$$

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, see for example [Kal02, 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

$$\varphi_i : G \times [0,1]^{is} \to G,$$
$$\varphi_i(x; u_1, u_2, \ldots, u_i) = \varphi(\varphi_{i-1}(x; u_1, u_2, \ldots, u_{i-1}); u_i).$$

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

Lemma 2 *Let $i, j \in \mathbb{N}$ and $i \geq j$. For any $u_1, \ldots, u_i \in [0,1]^s$ and $x \in G$ we have*

$$\varphi_i(x; u_1, \ldots, u_i) = \varphi_{i-j}(\varphi_j(x; u_1, \ldots, u_j); u_{j+1}, \ldots, u_i). \tag{6}$$

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

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

$$B_i : G \times \mathcal{B}(G) \to \mathcal{B}([0,1]^{is}),$$
$$B_i(x, A) = \{(u_1, u_2, \ldots, u_i) \in [0,1]^{is} : \varphi_i(x; u_1, u_2, \ldots, u_i) \in A\}$$

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

In [DRZ13] we considered the case where the initial state is deterministically chosen. The following assumption is useful to work with general initial distributions.

Assumption 1 *For a probability measure ν on $(G, \mathcal{B}(G))$ we assume that $\psi \colon [0, 1]^s \to G$ is a generator function, i.e. ψ satisfies*

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

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

For a probability measure ν on $(G, \mathcal{B}(G))$ let Assumption 1 be satisfied. Then, for $i \geq 1$ with $i \in \mathbb{N}$ and $A \in \mathcal{B}(G)$, let

$$
\begin{aligned}
C_{i,\psi}(A) &= \{(u_0, u_1, \ldots, u_i) \in [0, 1]^{(i+1)s} \colon \varphi_i(\psi(u_0); u_1, \ldots, u_i) \in A\} \\
&= \{(u_0, u_1, \ldots, u_i) \in [0, 1]^{(i+1)s} \colon (u_1, \ldots, u_i) \in B_i(\psi(u_0), A)\}
\end{aligned}
\tag{7}
$$

where $C_{0,\psi}(A) = \{u_0 \in [0, 1]^s \colon \psi(u_0) \in A\}$. 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 3 *Let K be a transition kernel and ν a distribution on $(G, \mathcal{B}(G))$. Let φ be an update function for the transition kernel K. Let $(X_n)_{n \in \mathbb{N}}$ be a Markov chain with transition kernel K and initial distribution ν. Further, let Assumption 1 for ν be satisfied. Let $i \in \mathbb{N}$ and $F \colon G^i \to \mathbb{R}$. The expectation of F with respect to the joint distribution of $X_1, \ldots, X_i$ is given by*

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

10

Then

$$\mathbb{E}_{\nu,K}(F(X_1,\ldots,X_i))$$
$$= \int_{[0,1]^{is}} F(\psi(u_0), \varphi_1(\psi(u_0), u_1), \ldots, \varphi_{i-1}(\psi(u_0), u_1, \ldots, u_{i-1})) \qquad (8)$$
$$\times \, \mathrm{d}u_0 \, \mathrm{d}u_1 \ldots \mathrm{d}u_{i-1},$$

whenever one of the integrals exist.

Proof. By Assumption 1 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 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 the assertion is proven. $\qquad\square$

Note that the right-hand-side of (8) is the expectation with respect to the uniform distribution in $[0,1]^{is}$.

Corollary 1 *Assume that the conditions of Lemma 3 are satisfied. Then, for $A \in \mathcal{B}(G)$, holds*
$$\nu P^i(A) = \lambda_{(i+1)s}(C_{i,\psi}(A)), \qquad (9)$$
and $\nu P^0(A) = \nu(A) = \lambda_s(C_{0,\psi}(A))$.

Proof. By Lemma 3 we have

$$\nu P^i(A) = \int_G K^i(x, A) \, \nu(\mathrm{d}x)$$

11

$$= \int_G \underbrace{\int_G \cdots \int_G}_{i\text{-times}} 1_{x_{i+1} \in A} \, K(x_i, \mathrm{d}x_{i+1}) \ldots K(x_1, \mathrm{d}x_2) \, \nu(\mathrm{d}x_1)$$

$$= \int_{[0,1]^{(i+1)s}} 1_{\varphi_i(\psi(u_0), u_1, \ldots, u_i) \in A} \, \mathrm{d}u_0 \, \mathrm{d}u_1 \ldots \mathrm{d}u_i$$

$$= \int_{[0,1]^{(i+1)s}} 1_{(u_0, u_1, \ldots, u_i) \in C_{i,\psi}(A)} \, \mathrm{d}u_0 \, \mathrm{d}u_1 \ldots \mathrm{d}u_i = \lambda_{(i+1)s}(C_{i,\psi}(A)),$$

which completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

3 On the push-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 the set $S_n = \{x_1, \ldots, x_n\} \subseteq G$ be given by

$$x_{i+1} = x_{i+1}(x_1) = \varphi(x_i; u_i) = \varphi_i(x_1; u_1, \ldots, u_i), \quad i = 1, \ldots, n-1, \quad (10)$$

where $x_1 = \psi(u_0)$. Note that 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. We call it *push-back discrepancy*. Below we show how this push-back discrepancy is related to the star-discrepancy of S_n.

Definition 4 (Push-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* (7). *Define the local discrepancy function by*

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

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

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

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

The discrepancy of the driver sequence $D^*_{\mathscr{A},\psi,\varphi}(\mathcal{U}_n)$ is a 'push-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 push-back discrepancy of $\mathcal{U}_n$, this is similar to [DRZ13, Theorem 1].

Theorem 1 *Let K be a transition kernel and ν be a distribution on $(G, \mathcal{B}(G))$. Let φ be an update function for K and let us assume that ν satisfies Assumption 1 with generator function ψ. Further, let $\mathcal{U}_n = \{u_0, u_1, \ldots, u_{n-1}\} \subset [0,1]^s$ be the driver sequence, such that S_n is given by (10). Let $\mathscr{A} \subseteq \mathcal{B}(G)$ be a set of test sets. Then*

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

Proof. For any $A \in \mathscr{A}$ we have by (9) 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

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

13

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.   □

Corollary 2 *Assume that the conditions of Theorem 1 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*

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

where $\Lambda_0 = \max\{0,\Lambda\}$ and Λ is defined in (4).

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 1 and Theorem 1.   □

Remark 1 *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*

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

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 $D^*_{\mathscr{A},\pi}(P_n) = D^*_{\mathscr{A},\psi,\varphi}(\mathcal{U}_n)$.

14

4 Monte Carlo rate of convergence

In this section we show 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 (10), 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 $n^{-1/2}$ if the transition kernel or the corresponding Markov operator is variance bounding. 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 1.

4.1 Useful tools: delta-cover and Hoeffding inequality

The concept of a δ-cover will be useful (cf. [Gne08] for a discussion of δ-covers, bracketing numbers and Vapnik-Červonenkis dimension).

Definition 5 *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 [HNWW01, Section 2.1] (see also [DRZ13, Remark 3] for the particular case below).

Proposition 2 *Let Γ_δ be a δ-cover of $\mathscr{A}$ with respect to π. Then, for any $Z_n = \{z_1, \ldots, z_n\}$, holds*

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

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, i.e. the second largest eigenvalue of the transition matrix is less than 1, in [LP04] a Hoeffding inequality is proven. In [Mia12] this is extended to non-reversible Markov chains on general state spaces. The following Hoeffding inequality for reversible, variance bounding Markov chains follows by [Mia12, Theorem 3.3 and the remark after (3.4)].

Proposition 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*

$$\mathbb{P}_{\nu,K}\left[\left|\frac{1}{n}\sum_{i=1}^{n}1_{X_i \in A} - \pi(A)\right| \geq c\right] \leq 2\left\|\frac{d\nu}{d\pi}\right\|_2 \exp\left(-\frac{1-\Lambda_0}{1+\Lambda_0}c^2 n\right), \quad (11)$$

with $\Lambda_0 = \max\{0, \Lambda\}$ and where Λ is defined in (4).

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

$$\Delta_{n,A,\varphi,\psi}(u_0, \ldots, u_{n-1}) = \frac{1}{n}\sum_{i=0}^{n-1}\left[1_{(u_0,\ldots,u_i) \in C_{i,\psi}(A)} - \pi(A)\right]. \quad (12)$$

Lemma 4 *Let K be a transition kernel and ν be a distribution on $(G, \mathcal{B}(G))$. Let φ be an update function of K and let us assume that ν satisfies Assumption 1 with generator function ψ. 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 $c > 0$, holds*

$$\mathbb{P}[|\Delta_{n,A,\varphi,\psi}| \geq c] = \mathbb{P}_{\nu,K}\left[\left|\frac{1}{n}\sum_{i=1}^{n}1_{X_i \in A} - \pi(A)\right| \geq c\right], \quad (13)$$

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 3 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 is follows. $\qquad\square$

4.2 Discrepancy bounds

We show that for any $s \in \mathbb{N}$, for any update function of the transition kernel K, for every initial distribution ν with $\frac{d\nu}{d\pi} \in L_2$ and every n there exists a finite sequence $u_0, u_1, \ldots, u_{n-1} \in [0,1]^s$ such that the star-discrepancy of S_n, given by (10), 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 (12), 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 2 *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 and that ν satisfies Assumption 1 with generator ψ. 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. Further, let φ be an update function for K.*

Then 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

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

satisfies

$$D^*_{\mathscr{A},\pi}(S_n) \le \sqrt{\frac{1+\Lambda_0}{1-\Lambda_0}} \cdot \frac{\sqrt{2\log(|\Gamma_\delta|^2 \left\|\frac{d\nu}{d\pi}\right\|_2)}}{\sqrt{n}} + \delta, \qquad (14)$$

with $\Lambda_0 = \max\{0,\Lambda\}$ *and* Λ *defined in* (4).

Remark 2 *In Lemma 6 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 and Proposition 3 we have for any $c_n \ge 0$ that

$$\mathbb{P}\left[|\Delta_{n,A,\varphi,\psi}| \le c_n\right] \ge 1 - 2\left\|\frac{d\nu}{d\pi}\right\|_2 \exp\left(-\frac{1-\Lambda_0}{1+\Lambda_0}c_n^2 n\right). \qquad (15)$$

Let

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

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

$$\mathbb{P}\left[|\Delta_{n,A,\varphi,\psi}| \le c_n\right] > 1 - \frac{1}{|\widehat{\Gamma}_\delta|}, \qquad (16)$$

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

$$\max_{A \in \widehat{\Gamma}_\delta} |\Delta_{n,A,\varphi,\psi}(u_0, \ldots, u_{n-1})| \le c_n. \qquad (17)$$

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 (15) that (16) holds and that there exists a finite sequence $u_0, \ldots, u_{n-1} \in [0,1]^s$ such that (17) 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 $\pi(D \setminus C) \le \delta$. Hence

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

18

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

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

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

Set

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

and

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

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

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

The last inequality follows by the δ-cover property, (17) 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. $\qquad \square$

By Corollary 2 and Theorem 2 we can also state an upper bound on the push-back discrepancy.

Theorem 3 *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 and that ν satisfies Assumption 1 with*

19

generator ψ. 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. Further, let φ be an update function for K.

Then 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 (4).

We refer to Remark 2 and Lemma 6 for a relation between δ and $|\Gamma_\delta|$. Thus, we showed the existence of a driver sequence with small push-back discrepancy. Note that by using Corollary 2 one could also argue the other way around: If one can construct a sequence with small push-back discrepancy then the star-discrepancy of S_n is also small.

Remark 3 *Let us consider a special case of Theorem 2 and Theorem 3. 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*

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

since $\Lambda_0 = \Lambda = 0$. This is essentially the same as Theorem 1 in [HNWW01] in their setting. However, it is not as eloberate as Theorem 4 in [HNWW01], which is based on results by Talagrand [Tal94] and Haussler [Hau95]. We do not know a version of these results which apply to Markov chains (such a result could yield an improvement of Theorems 2 and 3).

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 3.

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 (10), 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 *push-back discrepancy with burn-in*.

Definition 6 (Push-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 (7). 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 $D^*_{n_0,\mathscr{A},\psi,\varphi}(\mathcal{U}_{n_0,n})$ push-back discrepancy with burn-in of $\mathcal{U}_{n_0,n}$.

By adapting Proposition 3 and Lemma 4 to the setting with burn-in we obtain, by the same steps as in the proof of Theorem 2, a bound on the star-discrepancy for $S_{[n_0,n]}$. Further, adapting Theorem 1 and Corollary 2 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 *Let K be a reversible transition kernel with respect to π and let ν be a distribution with $\frac{d\nu}{d\pi} \in L_2$. Assume that P, the Markov operator of K, is variance bounding and that ν satisfies Assumption 1 with generator ψ. 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. Further, let φ be an update function for K.*

Then 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

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

with $\Lambda_0 = \max\{0,\Lambda\}$ and Λ defined in (4). If P has an absolute L_2-spectral gap we have

$$
\begin{aligned}
D^*_{n_0,\mathscr{A},\psi,\varphi}(\mathcal{U}_{n_0,n}) \le & \sqrt{\frac{1+\Lambda_0}{1-\Lambda_0}} \cdot \frac{\sqrt{2\log(|\Gamma_\delta|^2(1+\beta^{n_0}\left\|\frac{d\nu}{d\pi}-1\right\|_2)}}{\sqrt{n}} \\
& + \frac{(1-\Lambda_0^n)\beta^{n_0}}{n\cdot(1-\Lambda_0)}\left\|\frac{d\nu}{d\pi}-1\right\|_2 + \delta,
\end{aligned}
\tag{18}
$$

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

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

$$\tag{19}$$

Equations (18) and (19) 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 parellel 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 2 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 [Bec84], for a proof and further details we refer to [AD, Theorem 1]. We state it as a lemma.

Lemma 5 *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 $\operatorname{supp}\mu$ be the closure of*

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

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

$$D_{\mathscr{A},\mu}^*(Z_r) \le 63\sqrt{d}\,\frac{(2 + \log_2 r)^{(3d+1)/2}}{r}. \tag{20}$$

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

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

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

Corollary 3 *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 $S_r = \{x_1, \ldots, x_r\} \subseteq G$ such that*

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

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 we have that there exists a set $Z_r = \{z_1, \ldots, z_r\} \subseteq \operatorname{supp} \mu$ such that (20) is satisfied. Let $x_i = T^{-1}(z_i)$ for $i = 1, \ldots, r$ 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 $z_1, \ldots, z_r \in \operatorname{supp}\mu \subset T(G)$ 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 (20) the assertion follows. $\qquad\square$

As in [DRZ13, 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 [DRZ13, Lemma 4]. To simplify the bound of Corollary 3, for any $r \in \mathbb{N}$ and $0 < \varepsilon < 1$ we have

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

where

$$C_{\varepsilon, d} = \max_{x \ge 1} \frac{(2 + \log_2 x)^{(3d+1)/2}}{x^\varepsilon} = 4^\varepsilon \left(\frac{3d + 1}{2e\varepsilon \log 2} \right)^{(3d+1)/2}. \tag{21}$$

With this notation we obtain the following result.

Lemma 6 *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 test set $\mathscr{B} = \{(-\infty, x)_G \mid x \in \mathbb{R}^d\}$, any $0 < \delta \le 1$ and $0 < \varepsilon < 1$, there is a δ-cover Γ_δ of $\mathscr{B}$ with respect to π with*

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

where $C_{\varepsilon, d}$ is given by (21).

Proof. The proof of the assertion follows essentially by the same steps as the proof of [DRZ13, Lemma 4]. The only difference is that we use the discrepancy bound of Corollary 3 instead of [HNWW01, Theorem 4]. □

The dependence of the size of the δ-cover on δ is arbitrarily close to order δ^{-d} in Lemma 6, whereas in [DRZ13, Lemma 4] it is of order δ^{-2d}. Furthermore, the constant in Lemma 6 is fully explicit (one can choose $0 < \varepsilon < 1$ to obtain the best bound on the size of the δ-cover).

By Theorem 2 and Lemma 6 we obtain the following result.

Corollary 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 and that ν satisfies Assumption 1 with generator ψ. Let $\mathscr{B} = \{(-\infty, x)_G \mid x \in \mathbb{R}^d\}$ be the set of test sets and φ be an update function of K.*

Then, 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 (10) satisfies

$$D^*_{\mathscr{B},\pi}(S_n) \leq \sqrt{\frac{1 + \Lambda_0}{1 - \Lambda_0}} \cdot \frac{\sqrt{2} \left(\log \left\| \frac{d\nu}{d\pi} \right\|_2 + d \log n + 3d^2 \log(5d)\right)^{1/2}}{\sqrt{n}} + \frac{8}{n^{3/4}}, \quad (22)$$

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

Proof. Let $\varepsilon = 1/4$. Thus $C_{1/4,d} = \sqrt{2}\left(\frac{6d+2}{e \log 2}\right)^{(3d+1)/2}$ and $|\Gamma_\delta| \leq (16\delta^{-4/3}(5d)^{3d})^d$. By $\delta = 8/n^{3/4}$ and Theorem 2 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 [HNWW01, 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 [DRZ13, Appendix B] we define a reproducing kernel Q by

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

25

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 [Aro50]. It is also known that the functions f in H_2 permit the representation

$$f(x) = f_0 + \int_{\mathbb{R}^d} 1_{(-\infty,z)_G}(x)\widetilde{f}(z)\,\rho(\mathrm{d}z), \tag{23}$$

for some $f_0 \in \mathbb{C}$ and function $\widetilde{f} \in L_2(\mathbb{R}^d, \rho)$, see for instance [SC08, Theorem 4.21, p. 121] or follow the same arguments as in [BD14, 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 $1 \le q \le \infty$ we also define the space H_q of functions of the form (23) for which $\widetilde{f} \in L_q(G, \rho)$, with finite norm

$$\|f\|_{H_q} = \left(|f_0|^q + \int_{\mathbb{R}^d} |\widetilde{f}(z)|^q \rho(\mathrm{d}z) \right)^{1/q}. \tag{24}$$

The following result concerning the integration error in H_q is proven in [DRZ13, Theorem 3].

Theorem 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 $1 \le p, q \le \infty$ 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*

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

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 (Markov chain Koksma-Hlawka inequality) *Assume that the conditions of Corollary 2 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 (24). Let $\mathcal{U}_n = \{u_0, \dots, u_{n-1}\} \subset [0,1]^s$ be the driver sequence, such that $S_n = \{x_1, \dots, x_n\}$ is given by (10). Then for all $f \in H_1$ we have*

$$
\left| \int_G f(x)\pi(\mathrm{d}z) - \frac{1}{n} \sum_{i=1}^n f(x_i) \right|
$$

$$
\leq \left(D^*_{\mathscr{B},\psi,\varphi}(\mathcal{U}_n) + \frac{1 - \Lambda_0^n}{n \cdot (1 - \Lambda_0)} \left\| \frac{\mathrm{d}\nu}{\mathrm{d}\pi} - 1 \right\|_2 \right) \|f\|_{H_1},
$$

with ν induced by ψ and $\Lambda_0 = \max\{0, \Lambda\}$, where Λ is defined in (4).

In the spirit of Remark 1 we obtain for $K(x, A) = \pi(A)$ that $\Lambda = 0$. Further, if $\nu = \pi$ we have the Koksma-Hlawka inequality

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

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 $\rho \colon \mathbb{B}_d \to (0, \infty)$, where ρ is integrable with respect to the Lebegue 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 consider the Metropolis algorithm with ball walk proposal for the approximate sampling of 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

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

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

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

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 [FW94]. Then, $\psi_\gamma\colon [0,1]^d \to D_\gamma(0)$ given by

$$\psi_\gamma(\bar{u}) = (\gamma\, v_d)^{1/d} \widetilde{\psi}(v_1, \ldots, v_{d-1}), \tag{25}$$

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

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

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

$$\varphi_{M,\gamma,\rho}(x,u) = \begin{cases} \varphi_{W,\gamma}(x,v_1,\ldots,v_d) & v_{d+1} \leq A(x,v_1,\ldots,v_d) \\ x & v_{d+1} > A(x,v_1,\ldots,v_d), \end{cases} \qquad (26)$$

where $u = (v_1,\ldots,v_{d+1}) \in [0,1]^{d+1}$ and $x \in \mathbb{B}_d$. Thus the algorithms are given by the update functions above.

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

$$\mathcal{R}_{\alpha,d} = \{\rho\colon \mathbb{B}_d \to (0,\infty) \mid \rho \text{ log-concave}, |\log \rho(x) - \log \rho(y)| \leq \alpha \|x-y\|\}. \qquad (27)$$

Next we provide a lower bound for $\Lambda_{\gamma,\rho}$, defined as in (4) for the transition kernel $M_{\gamma,\rho}$, where the density ρ is log-concave and log-Lipschitz. The result follows by [MN07, Corollary 1, Lemma 13].

Proposition 4 *Let us assume that $\rho \in \mathcal{R}_{\alpha,d}$. Further let*

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

Then

$$1 - \Lambda_{\gamma^*,\rho} \geq \frac{3.125 \cdot 10^{-6}}{d+1} \min\left\{\frac{1}{d+1}, \frac{1}{\alpha}\right\}. \qquad (28)$$

The combination of Proposition 4, Theorem 5, Lemma 6 and Corollary 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 6 *Let ν be the uniform distribution on $(\mathbb{B}_d, \mathcal{B}(\mathbb{B}_d))$ with generator function ψ_1, see (25). 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 (26).

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 4 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 4 and Theorem 5 the assertion follows. $\square$

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 2 reveals that, if the driver sequence is chosen at random from the uniform distribution the discrepancy bound of (14) 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 leads to good discrepancy bounds.

Its proof is based on the 'anywhere-to-anywhere' condition and Corollary 3.

Definition 7 *Let $\varphi : G \times [0,1]^s \to G$ be an update function. 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 3. We obtain a bound on the star-discrepancy for the Markov chain quasi-Monte Carlo construction.

Corollary 6 *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. Let φ be an update function and assume that φ satisfies the* 'anywhere-to-anywhere' *condition. Let $\psi \colon [0,1]^s \to G$ be an arbitrarily surjective measurable function. Then for any $n \in \mathbb{N}$ there exists $u_0, u_1, \ldots u_{n-1} \in [0,1]^s$ such that $S_n = \{x_1, \ldots, x_n\}$ given by $x_1 = \psi(u_0)$ and*
$$x_i = \varphi(x_{i-1}; u_i) \quad for \ i \geq 1,$$

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

The corallary 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 2, then a similar discrepancy bound as in Corollary 6 also holds for the driver sequence $\mathcal{U}_n = \{u_0, u_1, \ldots, u_{n-1}\}$. Namely

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

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.) 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, we do obtain that the push-back discrepancy is the relevant criterion for constructing driver sequences.

Acknowledgements

References

[AD]   Ch. Aistleitner and J. Dick, *Low-discrepancy point sets for non-uniform measures*, Submitted, 2013.

[Aro50]   N. Aronszajn, *Theory of reproducing kernels*, Trans. Amer. Math. Soc. **68** (1950), 337–404.

[BD14]   J. Brauchart and J. Dick, *A characterization of sobolev spaces on the sphere and an extension of stolarsky's invariance principle to arbitrary smoothness*, to appear in Constructive Approx. (2014).

[Bec84]   J. Beck, *Some upper bounds in the theory of irregularities of distribution*, Acta Arith. **43** (1984), 115–130.

[CDO11]   S. Chen, J. Dick, and A. Owen, *Consistency of markov chain quasi-monte carlo on continuous state spaces*, Ann. Statist. **39** (2011), 673–701.

[Che11]   S. Chen, *Consistency and convergence rate of markov chain quasi-monte carlo with examples*, Ph.D. thesis, Stanford University, 2011.

[CMNO12]   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.

[DRZ13]   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).

[FW94]   K. Fang and Y. Wang, *Number-theoretic methods in statistics*, Chapman & Hall, London, 1994.

[Gne08]   M. Gnewuch, *Bracketing numbers for axis-parallel boxes and applications to geometric discrepancy*, J. Complexity **24** (2008), 154–172.

[Hau95]   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.

33

[HNWW01] 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.

[Kal02] O. Kallenberg, *Foundations of modern probability*, Second ed., Probability and its Applications, Springer-Verlag, New York, 2002.

[Kre89] E. Kreyszig, *Introductory functional analysis with applications*, John Wiley & Sons, Inc., 1989.

[Lia98] L. Liao, *Variance reduction in gibbs sampler using quasi random numbers*, J. Comput. Graph. Statist. **7** (1998), 253–266.

[Liu08] J. Liu, *Monte Carlo strategies in scientific computing*, Springer Series in Statistics, Springer, New York, 2008.

[LP04] C. León and F. Perron, *Optimal Hoeffding bounds for discrete reversible Markov chains*, Ann. Appl. Probab. **14** (2004), no. 2, 958–970.

[LS06] C. Lemieux and P. Sidorsky, *Exact sampling with highly uniform point sets*, Math. Comput. Modelling **43** (2006), 339–349.

[Mia12] B. Miasojedow, *Hoeffding's inequalities for geometrically ergodic markov chains on general state space*, Preprint, Available at http://arxiv.org/abs/1201.2265 (2012).

[MN07] P. Mathé and E. Novak, *Simple Monte Carlo and the Metropolis algorithm*, J. Complexity **23** (2007), no. 4-6, 673–696.

[OT05] A. Owen and S. Tribble, *A quasi-monte carlo metropolis algorithm*, Proc. Natl. Acad. Sci. USA, 102 **102** (2005), no. 25, 8844–8849.

[RR08] G. Roberts and J. Rosenthal, *Variance bounding Markov chains*, Ann. Appl. Probab. **18** (2008), no. 3, 1201–1214.

[Rud91] W. Rudin, *Functional analysis*, second ed., International Series in Pure and Applied Mathematics, McGraw-Hill Inc., New York, 1991.

[Rud12]   D. Rudolf, *Explicit error bounds for Markov chain Monte Carlo*, Dissertationes Math. **485** (2012), 93 pp.

[SC08]   I. Steinwart and A. Christmann, *Support vector machines*, Information Science and Statistics, Springer-Verlag New York, 2008.

[Sob74]   I. Sobol, *Pseudo-random numbers for constructing discrete markov chains by the monte carlo method*, USSR Compat. Math. Math. Phys. **14** (1974), 36–45.

[Tal94]   M. Talagrand, *Sharper bounds for gaussian and empirical processes*, Ann. Probab. **22** (1994), 28–76.

[TO08]   S. Tribble and A. Owen, *Construction of weakly cud sequences for mcmc sampling*, Electron. J. Stat. **2** (2008), 634–660.

[Tri07]   S. Tribble, *Markov chain monte carlo algorithms using completely uniformly distributed driving sequences*, Ph.D. thesis, Stanford University, 2007.

Preprint Series DFG-SPP 1324

http://www.dfg-spp1324.de

Reports

[1] R. Ramlau, G. Teschke, and M. Zhariy. A Compressive Landweber Iteration for Solving Ill-Posed Inverse Problems. Preprint 1, DFG-SPP 1324, September 2008.

[2] G. Plonka. The Easy Path Wavelet Transform: A New Adaptive Wavelet Transform for Sparse Representation of Two-dimensional Data. Preprint 2, DFG-SPP 1324, September 2008.

[3] E. Novak and H. Woźniakowski. Optimal Order of Convergence and (In-) Tractability of Multivariate Approximation of Smooth Functions. Preprint 3, DFG-SPP 1324, October 2008.

[4] M. Espig, L. Grasedyck, and W. Hackbusch. Black Box Low Tensor Rank Approximation Using Fibre-Crosses. Preprint 4, DFG-SPP 1324, October 2008.

[5] T. Bonesky, S. Dahlke, P. Maass, and T. Raasch. Adaptive Wavelet Methods and Sparsity Reconstruction for Inverse Heat Conduction Problems. Preprint 5, DFG-SPP 1324, January 2009.

[6] E. Novak and H. Woźniakowski. Approximation of Infinitely Differentiable Multivariate Functions Is Intractable. Preprint 6, DFG-SPP 1324, January 2009.

[7] J. Ma and G. Plonka. A Review of Curvelets and Recent Applications. Preprint 7, DFG-SPP 1324, February 2009.

[8] L. Denis, D. A. Lorenz, and D. Trede. Greedy Solution of Ill-Posed Problems: Error Bounds and Exact Inversion. Preprint 8, DFG-SPP 1324, April 2009.

[9] U. Friedrich. A Two Parameter Generalization of Lions' Nonoverlapping Domain Decomposition Method for Linear Elliptic PDEs. Preprint 9, DFG-SPP 1324, April 2009.

[10] K. Bredies and D. A. Lorenz. Minimization of Non-smooth, Non-convex Functionals by Iterative Thresholding. Preprint 10, DFG-SPP 1324, April 2009.

[11] K. Bredies and D. A. Lorenz. Regularization with Non-convex Separable Constraints. Preprint 11, DFG-SPP 1324, April 2009.

[12] M. Döhler, S. Kunis, and D. Potts. Nonequispaced Hyperbolic Cross Fast Fourier Transform. Preprint 12, DFG-SPP 1324, April 2009.

[13] C. Bender. Dual Pricing of Multi-Exercise Options under Volume Constraints. Preprint 13, DFG-SPP 1324, April 2009.

[14] T. Müller-Gronbach and K. Ritter. Variable Subspace Sampling and Multi-level Algorithms. Preprint 14, DFG-SPP 1324, May 2009.

[15] G. Plonka, S. Tenorth, and A. Iske. Optimally Sparse Image Representation by the Easy Path Wavelet Transform. Preprint 15, DFG-SPP 1324, May 2009.

[16] S. Dahlke, E. Novak, and W. Sickel. Optimal Approximation of Elliptic Problems by Linear and Nonlinear Mappings IV: Errors in L_2 and Other Norms. Preprint 16, DFG-SPP 1324, June 2009.

[17] B. Jin, T. Khan, P. Maass, and M. Pidcock. Function Spaces and Optimal Currents in Impedance Tomography. Preprint 17, DFG-SPP 1324, June 2009.

[18] G. Plonka and J. Ma. Curvelet-Wavelet Regularized Split Bregman Iteration for Compressed Sensing. Preprint 18, DFG-SPP 1324, June 2009.

[19] G. Teschke and C. Borries. Accelerated Projected Steepest Descent Method for Nonlinear Inverse Problems with Sparsity Constraints. Preprint 19, DFG-SPP 1324, July 2009.

[20] L. Grasedyck. Hierarchical Singular Value Decomposition of Tensors. Preprint 20, DFG-SPP 1324, July 2009.

[21] D. Rudolf. Error Bounds for Computing the Expectation by Markov Chain Monte Carlo. Preprint 21, DFG-SPP 1324, July 2009.

[22] M. Hansen and W. Sickel. Best m-term Approximation and Lizorkin-Triebel Spaces. Preprint 22, DFG-SPP 1324, August 2009.

[23] F.J. Hickernell, T. Müller-Gronbach, B. Niu, and K. Ritter. Multi-level Monte Carlo Algorithms for Infinite-dimensional Integration on $\mathbb{R}^{\mathbb{N}}$. Preprint 23, DFG-SPP 1324, August 2009.

[24] S. Dereich and F. Heidenreich. A Multilevel Monte Carlo Algorithm for Lévy Driven Stochastic Differential Equations. Preprint 24, DFG-SPP 1324, August 2009.

[25] S. Dahlke, M. Fornasier, and T. Raasch. Multilevel Preconditioning for Adaptive Sparse Optimization. Preprint 25, DFG-SPP 1324, August 2009.

[26] S. Dereich. Multilevel Monte Carlo Algorithms for Lévy-driven SDEs with Gaussian Correction. Preprint 26, DFG-SPP 1324, August 2009.

[27] G. Plonka, S. Tenorth, and D. Roşca. A New Hybrid Method for Image Approximation using the Easy Path Wavelet Transform. Preprint 27, DFG-SPP 1324, October 2009.

[28] O. Koch and C. Lubich. Dynamical Low-rank Approximation of Tensors. Preprint 28, DFG-SPP 1324, November 2009.

[29] E. Faou, V. Gradinaru, and C. Lubich. Computing Semi-classical Quantum Dynamics with Hagedorn Wavepackets. Preprint 29, DFG-SPP 1324, November 2009.

[30] D. Conte and C. Lubich. An Error Analysis of the Multi-configuration Time-dependent Hartree Method of Quantum Dynamics. Preprint 30, DFG-SPP 1324, November 2009.

[31] C. E. Powell and E. Ullmann. Preconditioning Stochastic Galerkin Saddle Point Problems. Preprint 31, DFG-SPP 1324, November 2009.

[32] O. G. Ernst and E. Ullmann. Stochastic Galerkin Matrices. Preprint 32, DFG-SPP 1324, November 2009.

[33] F. Lindner and R. L. Schilling. Weak Order for the Discretization of the Stochastic Heat Equation Driven by Impulsive Noise. Preprint 33, DFG-SPP 1324, November 2009.

[34] L. Kämmerer and S. Kunis. On the Stability of the Hyperbolic Cross Discrete Fourier Transform. Preprint 34, DFG-SPP 1324, December 2009.

[35] P. Cerejeiras, M. Ferreira, U. Kähler, and G. Teschke. Inversion of the noisy Radon transform on $SO(3)$ by Gabor frames and sparse recovery principles. Preprint 35, DFG-SPP 1324, January 2010.

[36] T. Jahnke and T. Udrescu. Solving Chemical Master Equations by Adaptive Wavelet Compression. Preprint 36, DFG-SPP 1324, January 2010.

[37] P. Kittipoom, G. Kutyniok, and W.-Q Lim. Irregular Shearlet Frames: Geometry and Approximation Properties. Preprint 37, DFG-SPP 1324, February 2010.

[38] G. Kutyniok and W.-Q Lim. Compactly Supported Shearlets are Optimally Sparse. Preprint 38, DFG-SPP 1324, February 2010.

[39] M. Hansen and W. Sickel. Best m-Term Approximation and Tensor Products of Sobolev and Besov Spaces – the Case of Non-compact Embeddings. Preprint 39, DFG-SPP 1324, March 2010.

[40] B. Niu, F.J. Hickernell, T. Müller-Gronbach, and K. Ritter. Deterministic Multilevel Algorithms for Infinite-dimensional Integration on $\mathbb{R}^{\mathbb{N}}$. Preprint 40, DFG-SPP 1324, March 2010.

[41] P. Kittipoom, G. Kutyniok, and W.-Q Lim. Construction of Compactly Supported Shearlet Frames. Preprint 41, DFG-SPP 1324, March 2010.

[42] C. Bender and J. Steiner. Error Criteria for Numerical Solutions of Backward SDEs. Preprint 42, DFG-SPP 1324, April 2010.

[43] L. Grasedyck. Polynomial Approximation in Hierarchical Tucker Format by Vector-Tensorization. Preprint 43, DFG-SPP 1324, April 2010.

[44] M. Hansen und W. Sickel. Best m-Term Approximation and Sobolev-Besov Spaces of Dominating Mixed Smoothness - the Case of Compact Embeddings. Preprint 44, DFG-SPP 1324, April 2010.

[45] P. Binev, W. Dahmen, and P. Lamby. Fast High-Dimensional Approximation with Sparse Occupancy Trees. Preprint 45, DFG-SPP 1324, May 2010.

[46] J. Ballani and L. Grasedyck. A Projection Method to Solve Linear Systems in Tensor Format. Preprint 46, DFG-SPP 1324, May 2010.

[47] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk. Convergence Rates for Greedy Algorithms in Reduced Basis Methods. Preprint 47, DFG-SPP 1324, May 2010.

[48] S. Kestler and K. Urban. Adaptive Wavelet Methods on Unbounded Domains. Preprint 48, DFG-SPP 1324, June 2010.

[49] H. Yserentant. The Mixed Regularity of Electronic Wave Functions Multiplied by Explicit Correlation Factors. Preprint 49, DFG-SPP 1324, June 2010.

[50] H. Yserentant. On the Complexity of the Electronic Schrödinger Equation. Preprint 50, DFG-SPP 1324, June 2010.

[51] M. Guillemard and A. Iske. Curvature Analysis of Frequency Modulated Manifolds in Dimensionality Reduction. Preprint 51, DFG-SPP 1324, June 2010.

[52] E. Herrholz and G. Teschke. Compressive Sensing Principles and Iterative Sparse Recovery for Inverse and Ill-Posed Problems. Preprint 52, DFG-SPP 1324, July 2010.

[53] L. Kämmerer, S. Kunis, and D. Potts. Interpolation Lattices for Hyperbolic Cross Trigonometric Polynomials. Preprint 53, DFG-SPP 1324, July 2010.

[54] G. Kutyniok and W.-Q Lim. Shearlets on Bounded Domains. Preprint 54, DFG-SPP 1324, July 2010.

[55] A. Zeiser. Wavelet Approximation in Weighted Sobolev Spaces of Mixed Order with Applications to the Electronic Schrödinger Equation. Preprint 55, DFG-SPP 1324, July 2010.

[56] G. Kutyniok, J. Lemvig, and W.-Q Lim. Compactly Supported Shearlets. Preprint 56, DFG-SPP 1324, July 2010.

[57] A. Zeiser. On the Optimality of the Inexact Inverse Iteration Coupled with Adaptive Finite Element Methods. Preprint 57, DFG-SPP 1324, July 2010.

[58] S. Jokar. Sparse Recovery and Kronecker Products. Preprint 58, DFG-SPP 1324, August 2010.

[59] T. Aboiyar, E. H. Georgoulis, and A. Iske. Adaptive ADER Methods Using Kernel-Based Polyharmonic Spline WENO Reconstruction. Preprint 59, DFG-SPP 1324, August 2010.

[60] O. G. Ernst, A. Mugler, H.-J. Starkloff, and E. Ullmann. On the Convergence of Generalized Polynomial Chaos Expansions. Preprint 60, DFG-SPP 1324, August 2010.

[61] S. Holtz, T. Rohwedder, and R. Schneider. On Manifolds of Tensors of Fixed TT-Rank. Preprint 61, DFG-SPP 1324, September 2010.

[62] J. Ballani, L. Grasedyck, and M. Kluge. Black Box Approximation of Tensors in Hierarchical Tucker Format. Preprint 62, DFG-SPP 1324, October 2010.

[63] M. Hansen. On Tensor Products of Quasi-Banach Spaces. Preprint 63, DFG-SPP 1324, October 2010.

[64] S. Dahlke, G. Steidl, and G. Teschke. Shearlet Coorbit Spaces: Compactly Supported Analyzing Shearlets, Traces and Embeddings. Preprint 64, DFG-SPP 1324, October 2010.

[65] W. Hackbusch. Tensorisation of Vectors and their Efficient Convolution. Preprint 65, DFG-SPP 1324, November 2010.

[66] P. A. Cioica, S. Dahlke, S. Kinzel, F. Lindner, T. Raasch, K. Ritter, and R. L. Schilling. Spatial Besov Regularity for Stochastic Partial Differential Equations on Lipschitz Domains. Preprint 66, DFG-SPP 1324, November 2010.

[67] E. Novak and H. Woźniakowski. On the Power of Function Values for the Approximation Problem in Various Settings. Preprint 67, DFG-SPP 1324, November 2010.

[68] A. Hinrichs, E. Novak, and H. Woźniakowski. The Curse of Dimensionality for Monotone and Convex Functions of Many Variables. Preprint 68, DFG-SPP 1324, November 2010.

[69] G. Kutyniok and W.-Q Lim. Image Separation Using Shearlets. Preprint 69, DFG-SPP 1324, November 2010.

[70] B. Jin and P. Maass. An Analysis of Electrical Impedance Tomography with Applications to Tikhonov Regularization. Preprint 70, DFG-SPP 1324, December 2010.

[71] S. Holtz, T. Rohwedder, and R. Schneider. The Alternating Linear Scheme for Tensor Optimisation in the TT Format. Preprint 71, DFG-SPP 1324, December 2010.

[72] T. Müller-Gronbach and K. Ritter. A Local Refinement Strategy for Constructive Quantization of Scalar SDEs. Preprint 72, DFG-SPP 1324, December 2010.

[73] T. Rohwedder and R. Schneider. An Analysis for the DIIS Acceleration Method used in Quantum Chemistry Calculations. Preprint 73, DFG-SPP 1324, December 2010.

[74] C. Bender and J. Steiner. Least-Squares Monte Carlo for Backward SDEs. Preprint 74, DFG-SPP 1324, December 2010.

[75] C. Bender. Primal and Dual Pricing of Multiple Exercise Options in Continuous Time. Preprint 75, DFG-SPP 1324, December 2010.

[76] H. Harbrecht, M. Peters, and R. Schneider. On the Low-rank Approximation by the Pivoted Cholesky Decomposition. Preprint 76, DFG-SPP 1324, December 2010.

[77] P. A. Cioica, S. Dahlke, N. Döhring, S. Kinzel, F. Lindner, T. Raasch, K. Ritter, and R. L. Schilling. Adaptive Wavelet Methods for Elliptic Stochastic Partial Differential Equations. Preprint 77, DFG-SPP 1324, January 2011.

[78] G. Plonka, S. Tenorth, and A. Iske. Optimal Representation of Piecewise Hölder Smooth Bivariate Functions by the Easy Path Wavelet Transform. Preprint 78, DFG-SPP 1324, January 2011.

[79] A. Mugler and H.-J. Starkloff. On Elliptic Partial Differential Equations with Random Coefficients. Preprint 79, DFG-SPP 1324, January 2011.

[80] T. Müller-Gronbach, K. Ritter, and L. Yaroslavtseva. A Derandomization of the Euler Scheme for Scalar Stochastic Differential Equations. Preprint 80, DFG-SPP 1324, January 2011.

[81] W. Dahmen, C. Huang, C. Schwab, and G. Welper. Adaptive Petrov-Galerkin methods for first order transport equations. Preprint 81, DFG-SPP 1324, January 2011.

[82] K. Grella and C. Schwab. Sparse Tensor Spherical Harmonics Approximation in Radiative Transfer. Preprint 82, DFG-SPP 1324, January 2011.

[83] D.A. Lorenz, S. Schiffler, and D. Trede. Beyond Convergence Rates: Exact Inversion With Tikhonov Regularization With Sparsity Constraints. Preprint 83, DFG-SPP 1324, January 2011.

[84] S. Dereich, M. Scheutzow, and R. Schottstedt. Constructive quantization: Approximation by empirical measures. Preprint 84, DFG-SPP 1324, January 2011.

[85] S. Dahlke and W. Sickel. On Besov Regularity of Solutions to Nonlinear Elliptic Partial Differential Equations. Preprint 85, DFG-SPP 1324, January 2011.

[86] S. Dahlke, U. Friedrich, P. Maass, T. Raasch, and R.A. Ressel. An adaptive wavelet method for parameter identification problems in parabolic partial differential equations. Preprint 86, DFG-SPP 1324, January 2011.

[87] A. Cohen, W. Dahmen, and G. Welper. Adaptivity and Variational Stabilization for Convection-Diffusion Equations. Preprint 87, DFG-SPP 1324, January 2011.

[88] T. Jahnke. On Reduced Models for the Chemical Master Equation. Preprint 88, DFG-SPP 1324, January 2011.

[89] P. Binev, W. Dahmen, R. DeVore, P. Lamby, D. Savu, and R. Sharpley. Compressed Sensing and Electron Microscopy. Preprint 89, DFG-SPP 1324, March 2011.

[90] P. Binev, F. Blanco-Silva, D. Blom, W. Dahmen, P. Lamby, R. Sharpley, and T. Vogt. High Quality Image Formation by Nonlocal Means Applied to High-Angle Annular Dark Field Scanning Transmission Electron Microscopy (HAADF-STEM). Preprint 90, DFG-SPP 1324, March 2011.

[91] R. A. Ressel. A Parameter Identification Problem for a Nonlinear Parabolic Differential Equation. Preprint 91, DFG-SPP 1324, May 2011.

[92] G. Kutyniok. Data Separation by Sparse Representations. Preprint 92, DFG-SPP 1324, May 2011.

[93] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok. Introduction to Compressed Sensing. Preprint 93, DFG-SPP 1324, May 2011.

[94] H.-C. Kreusler and H. Yserentant. The Mixed Regularity of Electronic Wave Functions in Fractional Order and Weighted Sobolev Spaces. Preprint 94, DFG-SPP 1324, June 2011.

[95] E. Ullmann, H. C. Elman, and O. G. Ernst. Efficient Iterative Solvers for Stochastic Galerkin Discretizations of Log-Transformed Random Diffusion Problems. Preprint 95, DFG-SPP 1324, June 2011.

[96] S. Kunis and I. Melzer. On the Butterfly Sparse Fourier Transform. Preprint 96, DFG-SPP 1324, June 2011.

[97] T. Rohwedder. The Continuous Coupled Cluster Formulation for the Electronic Schrödinger Equation. Preprint 97, DFG-SPP 1324, June 2011.

[98] T. Rohwedder and R. Schneider. Error Estimates for the Coupled Cluster Method. Preprint 98, DFG-SPP 1324, June 2011.

[99] P. A. Cioica and S. Dahlke. Spatial Besov Regularity for Semilinear Stochastic Partial Differential Equations on Bounded Lipschitz Domains. Preprint 99, DFG-SPP 1324, July 2011.

[100] L. Grasedyck and W. Hackbusch. An Introduction to Hierarchical (H-) Rank and TT-Rank of Tensors with Examples. Preprint 100, DFG-SPP 1324, August 2011.

[101] N. Chegini, S. Dahlke, U. Friedrich, and R. Stevenson. Piecewise Tensor Product Wavelet Bases by Extensions and Approximation Rates. Preprint 101, DFG-SPP 1324, September 2011.

[102] S. Dahlke, P. Oswald, and T. Raasch. A Note on Quarkonial Systems and Multilevel Partition of Unity Methods. Preprint 102, DFG-SPP 1324, September 2011.

[103] A. Uschmajew. Local Convergence of the Alternating Least Squares Algorithm For Canonical Tensor Approximation. Preprint 103, DFG-SPP 1324, September 2011.

[104] S. Kvaal. Multiconfigurational time-dependent Hartree method for describing particle loss due to absorbing boundary conditions. Preprint 104, DFG-SPP 1324, September 2011.

[105] M. Guillemard and A. Iske. On Groupoid C*-Algebras, Persistent Homology and Time-Frequency Analysis. Preprint 105, DFG-SPP 1324, September 2011.

[106] A. Hinrichs, E. Novak, and H. Woźniakowski. Discontinuous information in the worst case and randomized settings. Preprint 106, DFG-SPP 1324, September 2011.

[107] M. Espig, W. Hackbusch, A. Litvinenko, H. Matthies, and E. Zander. Efficient Analysis of High Dimensional Data in Tensor Formats. Preprint 107, DFG-SPP 1324, September 2011.

[108] M. Espig, W. Hackbusch, S. Handschuh, and R. Schneider. Optimization Problems in Contracted Tensor Networks. Preprint 108, DFG-SPP 1324, October 2011.

[109] S. Dereich, T. Müller-Gronbach, and K. Ritter. On the Complexity of Computing Quadrature Formulas for SDEs. Preprint 109, DFG-SPP 1324, October 2011.

[110] D. Belomestny. Solving optimal stopping problems by empirical dual optimization and penalization. Preprint 110, DFG-SPP 1324, November 2011.

[111] D. Belomestny and J. Schoenmakers. Multilevel dual approach for pricing American style derivatives. Preprint 111, DFG-SPP 1324, November 2011.

[112] T. Rohwedder and A. Uschmajew. Local convergence of alternating schemes for optimization of convex problems in the TT format. Preprint 112, DFG-SPP 1324, December 2011.

[113] T. Görner, R. Hielscher, and S. Kunis. Efficient and accurate computation of spherical mean values at scattered center points. Preprint 113, DFG-SPP 1324, December 2011.

[114] Y. Dong, T. Görner, and S. Kunis. An iterative reconstruction scheme for photoacoustic imaging. Preprint 114, DFG-SPP 1324, December 2011.

[115] L. Kämmerer. Reconstructing hyperbolic cross trigonometric polynomials by sampling along generated sets. Preprint 115, DFG-SPP 1324, February 2012.

[116] H. Chen and R. Schneider. Numerical analysis of augmented plane waves methods for full-potential electronic structure calculations. Preprint 116, DFG-SPP 1324, February 2012.

[117] J. Ma, G. Plonka, and M.Y. Hussaini. Compressive Video Sampling with Approximate Message Passing Decoding. Preprint 117, DFG-SPP 1324, February 2012.

[118] D. Heinen and G. Plonka. Wavelet shrinkage on paths for scattered data denoising. Preprint 118, DFG-SPP 1324, February 2012.

[119] T. Jahnke and M. Kreim. Error bound for piecewise deterministic processes modeling stochastic reaction systems. Preprint 119, DFG-SPP 1324, March 2012.

[120] C. Bender and J. Steiner. A-posteriori estimates for backward SDEs. Preprint 120, DFG-SPP 1324, April 2012.

[121] M. Espig, W. Hackbusch, A. Litvinenkoy, H.G. Matthiesy, and P. Wähnert. Efficient low-rank approximation of the stochastic Galerkin matrix in tensor formats. Preprint 121, DFG-SPP 1324, May 2012.

[122] O. Bokanowski, J. Garcke, M. Griebel, and I. Klompmaker. An adaptive sparse grid semi-Lagrangian scheme for first order Hamilton-Jacobi Bellman equations. Preprint 122, DFG-SPP 1324, June 2012.

[123] A. Mugler and H.-J. Starkloff. On the convergence of the stochastic Galerkin method for random elliptic partial differential equations. Preprint 123, DFG-SPP 1324, June 2012.

[124] P.A. Cioica, S. Dahlke, N. Döhring, U. Friedrich, S. Kinzel, F. Lindner, T. Raasch, K. Ritter, and R.L. Schilling. On the convergence analysis of Rothe's method. Preprint 124, DFG-SPP 1324, July 2012.

[125] P. Binev, A. Cohen, W. Dahmen, and R. DeVore. Classification Algorithms using Adaptive Partitioning. Preprint 125, DFG-SPP 1324, July 2012.

[126] C. Lubich, T. Rohwedder, R. Schneider, and B. Vandereycken. Dynamical approximation of hierarchical Tucker and Tensor-Train tensors. Preprint 126, DFG-SPP 1324, July 2012.

[127] M. Kovács, S. Larsson, and K. Urban. On Wavelet-Galerkin methods for semilinear parabolic equations with additive noise. Preprint 127, DFG-SPP 1324, August 2012.

[128] M. Bachmayr, H. Chen, and R. Schneider. Numerical analysis of Gaussian approximations in quantum chemistry. Preprint 128, DFG-SPP 1324, August 2012.

[129] D. Rudolf. Explicit error bounds for Markov chain Monte Carlo. Preprint 129, DFG-SPP 1324, August 2012.

[130] P.A. Cioica, K.-H. Kim, K. Lee, and F. Lindner. On the $L_q(L_p)$-regularity and Besov smoothness of stochastic parabolic equations on bounded Lipschitz domains. Preprint 130, DFG-SPP 1324, December 2012.

[131] M. Hansen. $n-$term Approximation Rates and Besov Regularity for Elliptic PDEs on Polyhedral Domains. Preprint 131, DFG-SPP 1324, December 2012.

[132] R. E. Bank and H. Yserentant. On the H^1 -stability of the L_2 -projection onto finite element spaces. Preprint 132, DFG-SPP 1324, December 2012.

[133] M. Gnewuch, S. Mayer, and K. Ritter. On Weighted Hilbert Spaces and Integration of Functions of Infinitely Many Variables. Preprint 133, DFG-SPP 1324, December 2012.

[134] D. Crisan, J. Diehl, P.K. Friz, and H. Oberhauser. Robust Filtering: Correlated Noise and Multidimensional Observation. Preprint 134, DFG-SPP 1324, January 2013.

[135] Wolfgang Dahmen, Christian Plesken, and Gerrit Welper. Double Greedy Algorithms: Reduced Basis Methods for Transport Dominated Problems. Preprint 135, DFG-SPP 1324, February 2013.

[136] Aicke Hinrichs, Erich Novak, Mario Ullrich, and Henryk Wozniakowski. The Curse of Dimensionality for Numerical Integration of Smooth Functions. Preprint 136, DFG-SPP 1324, February 2013.

[137] Markus Bachmayr, Wolfgang Dahmen, Ronald DeVore, and Lars Grasedyck. Approximation of High-Dimensional Rank One Tensors. Preprint 137, DFG-SPP 1324, March 2013.

[138] Markus Bachmayr and Wolfgang Dahmen. Adaptive Near-Optimal Rank Tensor Approximation for High-Dimensional Operator Equations. Preprint 138, DFG-SPP 1324, April 2013.

[139] Felix Lindner. Singular Behavior of the Solution to the Stochastic Heat Equation on a Polygonal Domain. Preprint 139, DFG-SPP 1324, May 2013.

[140] Stephan Dahlke, Dominik Lellek, Shiu Hong Lui, and Rob Stevenson. Adaptive Wavelet Schwarz Methods for the Navier-Stokes Equation. Preprint 140, DFG-SPP 1324, May 2013.

[141] Jonas Ballani and Lars Grasedyck. Tree Adaptive Approximation in the Hierarchical Tensor Format. Preprint 141, DFG-SPP 1324, June 2013.

[142] Harry Yserentant. A short theory of the Rayleigh-Ritz method. Preprint 142, DFG-SPP 1324, July 2013.

[143] M. Hefter and K. Ritter. On Embeddings of Weighted Tensor Product Hilbert Spaces. Preprint 143, DFG-SPP 1324, August 2013.

[144] M. Altmayer and A. Neuenkirch. Multilevel Monte Carlo Quadrature of Discontinuous Payoffs in the Generalized Heston Model using Malliavin Integration by Parts. Preprint 144, DFG-SPP 1324, August 2013.

[145] L. Kämmerer, D. Potts, and T. Volkmer. Approximation of multivariate functions by trigonometric polynomials based on rank-1 lattice sampling. Preprint 145, DFG-SPP 1324, September 2013.

[146] C. Bender, N. Schweizer, and J. Zhuo. A primal-dual algorithm for BSDEs. Preprint 146, DFG-SPP 1324, October 2013.

[147] D. Rudolf. Hit-and-run for numerical integration. Preprint 147, DFG-SPP 1324, October 2013.

[148] D. Rudolf and M. Ullrich. Positivity of hit-and-run and related algorithms. Preprint 148, DFG-SPP 1324, October 2013.

[149] L. Grasedyck, M. Kluge, and S. Krämer. Alternating Directions Fitting (ADF) of Hierarchical Low Rank Tensors. Preprint 149, DFG-SPP 1324, October 2013.

[150] F. Filbir, S. Kunis, and R. Seyfried. Effective discretization of direct reconstruction schemes for photoacoustic imaging in spherical geometries. Preprint 150, DFG-SPP 1324, November 2013.

[151] E. Novak, M. Ullrich, and H. Woźniakowski. Complexity of Oscillatory Integration for Univariate Sobolev Spaces. Preprint 151, DFG-SPP 1324, November 2013.

[152] A. Hinrichs, E. Novak, and M. Ullrich. A Tractability Result for the Clenshaw Curtis Smolyak Algorithm. Preprint 152, DFG-SPP 1324, November 2013.

[153] M. Hein, S. Setzer, L. Jost, and S. Rangapuram. The Total Variation on Hypergraphs - Learning on Hypergraphs Revisited. Preprint 153, DFG-SPP 1324, November 2013.

[154] M. Kovács, S. Larsson, and F. Lindgren. On the Backward Euler Approximation of the Stochastic Allen-Chan Equation. Preprint 154, DFG-SPP 1324, November 2013.

[155] S. Dahlke, M. Fornasier, U. Friedrich, and T. Raasch. Multilevel preconditioning for sparse optimization of functionals with nonconvex fidelity terms. Preprint 155, DFG-SPP 1324, December 2013.

[156] T. Müller-Gronbach, K. Ritter, and L. Yaroslavtseva. On the complexity of computing quadrature formulas for marginal distributions of SDEs. Preprint 156, DFG-SPP 1324, January 2014.

[157] M. Giles, T. Nagapetyan, and K. Ritter. Multi-Level Monte Carlo Approximation of Distribution Functions and Densities. Preprint 157, DFG-SPP 1324, February 2014.

[158] F. Dickmann and N. Schweizer. Faster Comparison of Stopping Times by Nested Conditional Monte Carlo . Preprint 158, DFG-SPP 1324, February 2014.

[159] L. Kämmerer, D. Potts, and T. Volkmer. Approximation of multivariate periodic functions by trigonometric polynomials based on sampling along rank-1 lattice with generating vector of Korobov form . Preprint 159, DFG-SPP 1324, February 2014.

[160] S. Dereich and T. Müller-Gronbach. Quadrature for Self-Affine Distributions on $\mathbb{R}^d$. Preprint 160, DFG-SPP 1324, March 2014.

[161] S. Dereich and S. Li. Multilevel Monte Carlo for Lévy-driven SDEs: central limit theorems for adaptive Euler schemes. Preprint 161, DFG-SPP 1324, March 2014.

[162] M. Kluge. Sampling Rules for Tensor Reconstruction in Hierarchical Tucker Format. Preprint 162, DFG-SPP 1324, April 2014.

[163] D. Rudolf and N. Schweizer. Error Bounds of MCMC for Functions with Unbounded Stationary Variance. Preprint 163, DFG-SPP 1324, April 2014.

[164] E. Novak and D. Rudolf. Tractability of the approximation of high-dimensional rank one tensors. Preprint 164, DFG-SPP 1324, April 2014.

[165] J. Dick and D. Rudolf. Discrepancy estimates for variance bounding Markov chain quasi-Monte Carlo. Preprint 165, DFG-SPP 1324, April 2014.