Discrepancy bounds for deterministic acceptance-rejection samplers[*]

Houying Zhu and Josef Dick

School of Mathematics and Statistics
The University of New South Wales, Sydney, Australia
e-mail: houying.zhu@student.unsw.edu.au*;* josef.dick@unsw.edu.au

Abstract: We consider an acceptance-rejection sampler based on a deterministic driver sequence. The deterministic sequence is chosen such that the discrepancy between the empirical target distribution and the target distribution is small. We use quasi-Monte Carlo (QMC) point sets for this purpose. The empirical evidence shows convergence rates beyond the crude Monte Carlo rate of $N^{-1/2}$. We prove that the discrepancy of samples generated by the QMC acceptance-rejection sampler is bounded from above by $N^{-1/s}$. A lower bound shows that for any given driver sequence, there always exists a target density such that the star discrepancy is at most $N^{-2/(s+1)}$. For a general density, whose domain is the real state space $\mathbb{R}^{s-1}$, the inverse Rosenblatt transformation can be used to convert samples from the $(s-1)$-dimensional cube to $\mathbb{R}^{s-1}$. We show that this transformation is measure preserving. This way, under certain conditions, we obtain the same convergence rate for a general target density defined in $\mathbb{R}^{s-1}$. Moreover, we also consider a deterministic reduced acceptance-rejection algorithm recently introduced by Barekat and Caflisch [F. Barekat and R. Caflisch, Simulation with Fluctuation and Singular Rates. ArXiv:1310.4555 [math.NA], 2013.]

MSC 2010 subject classifications: Primary 62F15; secondary 11K45.
Keywords and phrases: Acceptance-rejection sampler, Star discrepancy, (t, m, s)-nets.

Received July 2013.

1. Introduction

The Monte Carlo (MC) method is one of the widely used numerical methods for simulating probability distributions. However, sometimes it is not possible to sample from a given target distribution. Markov chain Monte Carlo (MCMC) methods have been developed to address this problem. Instead of sampling independent points directly, MCMC samples from a Markov chain whose limiting distribution is the target distribution. MCMC has widened the applications of MC in many different fields [5, 22]. Another deficiency of MC algorithms is its slow convergence rate. Quasi-Monte Carlo (QMC) algorithms on the other hand perform better in improving the convergence rate of Monte Carlo which

partially depends on generating samples with small discrepancy. For a survey of QMC we refer to [10, 11]. Putting the QMC idea into MCMC is a good way to improve the convergence rate and widen practical applications. Recently many results in this direction have been achieved [6, 7, 39, 40]. With this paper we add another result in this direction by using a deterministic driver sequence in an acceptance-rejection algorithm. We prove discrepancy bounds of order $N^{-1/s}$, where the dimension of the state space is $s-1$ and N is the number of samples. The discrepancy here is a generalization of the concept of the Kolmogorov-Smirnov test between the empirical distribution of the samples and the target distribution to higher dimension. A more detailed description of our results will be provided below.

1.1. Previous work on MCMC and QMC

In the following we describe some previous research on QMC and MC. L'Ecuyer studied the convergence behavior of randomized quasi-Monte Carlo for discrete-time Markov chains in [21], known as array-RQMC. The general idea is to obtain a better approximation of the target distribution than with the plain Monte Carlo method by using randomized QMC. In a different direction, Tribble [39] and Tribble and Owen [40] established a condition under which low discrepancy sequences can be used for consistent MCMC estimation for finite state spaces. It has been shown that replacing an IID sequence by a completely uniformly distributed sequence also implies a consistent estimation in finite state spaces. A construction of weakly completely uniformly distributed sequences is also proposed in [40]. As a sequel to the work of Tribble, Chen in his thesis [6] and Chen, Dick and Owen [7] demonstrated that Markov chain quasi-Monte Carlo (MCQMC) algorithms using a completely uniformly distributed sequence as driver sequence gives a consistent result under certain assumptions on the update function and Markov chain. Further, Chen [6] also showed that MCQMC can achieve a convergence rate of $O(N^{-1+\delta})$ for any $\delta > 0$ under certain conditions, but he only showed the existence of a driver sequence.

In our recent work [12], done with Rudolf, we prove upper bounds on the discrepancy under the assumptions that the Markov chain is uniformly ergodic and the driver sequence is deterministic rather than independent uniformly distributed random variables. In particular, we show the existence of driver sequences for which the discrepancy of the Markov chain from the target distribution with respect to certain test sets converges with (almost) the usual Monte Carlo rate of $N^{-1/2}$. A drawback of this result is that we are currently not able to give an explicit construction of a driver sequence for which our discrepancy bounds hold for uniformly ergodic Markov chains. Garber and Choppin in [14] adapted low discrepancy point sets instead of random numbers in sequential Monte Carlo (SMC). They proposed a new algorithm named sequential quasi-Monte Carlo (SQMC). They constructed consistency and stochastic bounds based on randomized QMC point set for this algorithm. It is an open problem to obtain deterministic bounds for SQMC. More literature review about applying QMC to MCMC problems can be found in [7, Section 1].

1.2. Acceptance-rejection algorithms

We now give a description of the algorithms in this paper. Let $\psi : D \to \mathbb{R}_+ = [0, \infty)$ be our target density function, where $D \subseteq \mathbb{R}^{s-1}$ and $\mathbb{R}_+ = [0, \infty)$. We consider the cases where $D = [0, 1]^{s-1}$ or $D = \mathbb{R}^{s-1}$. Assume that it is not possible to sample directly from the target distribution. One possible solution to obtain samples from ψ is to choose a proposal density H from which we can sample and then use an acceptance-rejection algorithm. Assume there exists a constant $L < \infty$ such that $\psi(\boldsymbol{x}) < LH(\boldsymbol{x})$ for all $\boldsymbol{x} \in D$. The following algorithm can be used to obtain samples with distribution ψ.

Algorithm 1 (Random acceptance-rejection (RAR) algorithm)**.** Given a target density $\psi : D \to \mathbb{R}_+$ and a proposal density $H : D \to \mathbb{R}_+$. Assume that there exists a constant $L < \infty$ such that $\psi(\boldsymbol{x}) < LH(\boldsymbol{x})$ for all $\boldsymbol{x}$ in the domain D. We introduce another random variable u having uniform distribution in the unit interval, i.e. $u \sim U([0, 1])$. Then the acceptance-rejection algorithm is given by

1. Draw $X \sim H$ and $u \sim U([0, 1])$.
2. Accept $Y = X$ as a sample of ψ if $u \leq \frac{\psi(\boldsymbol{X})}{LH(\boldsymbol{X})}$, otherwise go back to step 1.

See also [4, 9, 19] for a discussion of related algorithms. For a discussion on how to select proposal densities see for instance [4] and the references therein. The acceptance-rejection sampler works to sample from an unknown density based on a proposal density.

Acceptance-rejection sampling and importance sampling [34, Section 3] are quite similar ideas. Both of them distort a sample from a distribution in order to sample from another one. However, there is a difference in the selection of the constant $L > \psi(\boldsymbol{x})/H(\boldsymbol{x})$ for $\boldsymbol{x}$ in the domain D. The acceptance-rejection method does not work when $\sup_{\boldsymbol{x} \in D} \psi(\boldsymbol{x})/H(\boldsymbol{x}) = \infty$, while importance sampling is still available [33]. In this paper, we only use the acceptance-rejection sampler to get samples of a given target density, since we are interested in obtaining discrepancy bounds for MCQMC. More information on general strategies for generating nonuniform random variables can be found in the monographs [9, 19].

The acceptance-rejection algorithm with deterministic driver sequence is one special class of MCQMC. From the superior distribution properties in terms of the discrepancy of the Sobol sequence [37] one could expect an improvement in the discrepancy of the samples obtained from the acceptance-rejection algorithm based on the Sobol sequence. In one dimension the discrepancy we study is the Kolmogorov-Smirnov test between the target distribution and the empirical distribution of the sample points. For a given point set in the s-dimensional unit cube, the star discrepancy measures the difference between the proportion of points in a subinterval of $[0, 1]^s$ and the Lebesgue measure of this subinterval. We defer the precise definition of discrepancy to Section 3.

In this paper, we replace the IID initial samples with an explicit construction of the driver sequence by using (t, m, s)-nets [11, 27] (obtained from the Sobol sequence). Figure 1 shows a comparison between different driver sequences: de-
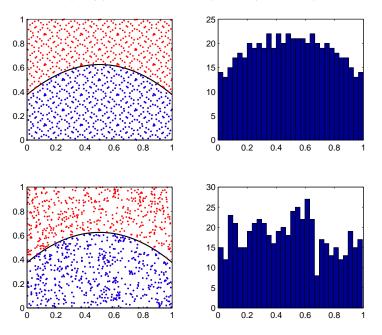
FIG 1. *Different driver sequences: deterministic and pseudo-random points (the total number of points in each case is 2^9) and histograms are w.r.t. samples we accepted.*

terministic points (Sobol points) and pseudo-random uniform points (they both have 2^9 points). The acceptance-rejection sampler works by only accepting those points under the target density curve. The difference of driver sequences will affect the samples we obtain by the acceptance-rejection algorithm, hence the distribution properties of the points which were accepted will be influenced. The right two figures in Fig. 1 show the histograms of the points which we accepted in both cases. Note that the deterministic samples better estimate the density function. Our interest in this paper is in entirely deterministic methods. However, one could also use randomized quasi-Monte Carlo point sets [30, 31, 32] and study a randomized setting.

1.3. Previous work on deterministic acceptance-rejection algorithms

The deterministic acceptance-rejection algorithm has also been discussed by Moskowitz and Caflisch [24] and Wang [41, 42]. Therein a smoothing technique was introduced to improve the numerical performance of the acceptance-rejection algorithm. Wang [42] gave a heuristic argument to indicate a convergence rate of order $N^{-\frac{s+2}{2(s+1)}}$. This argument assumes that the points in elementary intervals are uniformly distributed. Thus this reasoning is not fully deterministic. Our lower bound on the discrepancy (Theorem 2) indicates that this reasoning does not apply in our case. The numerical experiments in [42]

also indicate an improvement using a well chosen deterministic driver sequence (in this case the so-called Halton sequence [17]) compared to a random driver sequence. Recently, Nguyen and Ökten in [25] presented a consistency result of an acceptance-rejection algorithm for low-discrepancy sequences. This algorithm yielded good numerical performances on standard deviation and efficiency. However, proving an explicit convergence rate of the discrepancy for this algorithm is still an open problem. See also [23, 25] for numerical experiments using quasi-Monte Carlo point sets for the related problem of integrating indicator functions.

It is worth noticing that all results given in previous work are empirical evidence and the discrepancy of samples is not directly investigated. Our work focuses on discrepancy properties of points produced by totally deterministic acceptance-rejection methods. We also prove discrepancy bounds on deterministic acceptance-rejection algorithms, including an upper bound and a lower bound. The combination with the reduced acceptance-rejection sampler provides further evidence of the good performance of the deterministic method. Our algorithm here may also be combined with similar algorithms like the acceptance-complement method, see for instance [9, Section II.5].

Before presenting the theoretical background, we briefly describe deterministic algorithms and some numerical results which show a convergence rate comparison using Monte Carlo and quasi-Monte Carlo methods.

2. Our results

2.1. Construction of driver sequence

In this paper, we use low discrepancy point sets given by (t, s)-sequences (see Definition 3 and Definition 4 below) in base b as driver sequences. The first b^m points of a (t, s)-sequence are a so called (t, m, s)-net in base b. Explicit constructions of (t, s)-sequences in base 2 have been found by Sobol [37], in prime base $b \geq s$ by Faure [13] and in prime-power base b by Niederreiter [26]. In all these constructions t depends only on s but not on m. In practice, since digital nets (based on Sobol points) are included in the statistics toolbox of Matlab, this method is very easy to implement. People seeking more discussion of construction methods can also consult [11, Chapters 4&8].

In the following we describe the algorithm and present some numerical results. Since in general it is computationally too expensive to compute the supremum in the definition of the star-discrepancy exactly, we use a so-called δ-cover to estimate this supremum. An introduction to δ-covers is provided in the appendix. In the numerical discussion, the driver sequence is generated by a (t, m, s)-net in base 2. Specifically, we always use a Sobol sequence [37] to generate (t, m, s)-nets for our experiments.

2.2. Deterministic algorithm for target densities defined on $[0, 1]^s$

We consider now the case where the target density is defined on $[0, 1]^{s-1}$. The following algorithm is a deterministic version of Algorithm 1. For the proofs

later, we need the technical assumption that the target density is pseudo-convex. The definition of pseudo-convexity is discussed in Section 3.

Algorithm 2 (Deterministic acceptance-rejection (DAR) algorithm in $[0,1]^s$)**.** Let the target density $\psi : [0,1]^{s-1} \to \mathbb{R}_+$, where $s \geq 2$, be pseudo-convex. Assume that there exists a constant $L < \infty$ such that $\psi(\boldsymbol{x}) \leq L$ for all $\boldsymbol{x} \in [0,1]^{s-1}$. Let $A = \{\boldsymbol{x} \in [0,1]^s : \psi(x_1, \ldots, x_{s-1}) \geq Lx_s\}$. Suppose we aim to obtain approximately N samples from ψ.

i) Let $M = b^m \geq \lceil N/(\int_{[0,1]^{s-1}} \psi(\boldsymbol{x})/Ld\boldsymbol{x})\rceil$, where $m \in \mathbb{N}$ is the smallest integer satisfying this inequality. Generate a (t,m,s)-net $Q_{m,s} = \{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{b^m-1}\}$ in base b.

ii) Use the acceptance-rejection method for the points $Q_{m,s}$ with respect to the density ψ, i.e. we accept the point $\boldsymbol{x}_n$ if $\boldsymbol{x}_n \in A$, otherwise reject. Let $P_N^{(s)} = A \cap Q_{m,s} = \{\boldsymbol{z}_0, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_{N-1}\}$ be the sample set we accept.

iii) Project the points $P_N^{(s)}$ onto the first $(s-1)$ coordinates. Let $P_N^{(s-1)} = \{\boldsymbol{y}_0, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{N-1}\} \subseteq [0,1]^{s-1}$ be the projections of the points $P_N^{(s)}$.

iv) Return the point set $P_N^{(s-1)}$.

The following example shows a better convergence rate when using a low-discrepancy driver sequence rather than a random point set. In each example the reported discrepancy for the acceptance-rejection algorithm using a random diver sequence is the average of 10 independent runs, which is throughout all the numerical experiments.

Example 1. In this example we consider a non-product target density in $[0,1]^4$. Let the target density ψ be

$$\psi(x_1, x_2, x_3, x_4) = \frac{1}{4}(e^{-x_1} + e^{-x_2} + e^{-x_3} + e^{-x_4}), \quad (x_1, x_2, x_3, x_4) \in [0,1]^4.$$

Figure 2 shows the discrepancy by using deterministic points and pseudo-random points as driver sequence. For the RAR algorithm, we observe a convergence rate of order $N^{-0.482}$, whereas the DAR algorithm shows a convergence rate of the discrepancy of order $N^{-0.659}$.

2.3. Deterministic algorithm for target density defined in real state space

Now we extend the domain of the target density ψ to $\mathbb{R}^{s-1}$ with $s \geq 2$. Assume that there is a proposal density function $H : \mathbb{R}^{s-1} \to \mathbb{R}_+$ such that there exists a constant $L < \infty$ such that $\psi(\boldsymbol{z}) \leq LH(\boldsymbol{z})$ holds for all $\boldsymbol{z} \in \mathbb{R}^{s-1}$.

The inverse Rosenblatt transformation is used to generate samples from the proposal density in the real state space $\mathbb{R}^{s-1}$. Let F be the joint CDF of H and $F_j(z_j|z_1, \ldots, z_{j-1})$ be the conditional CDF of the proposal density for $j = 1, \ldots, s-1$. The transformation T is used to generate points in $\mathbb{R}^{s-1} \times \mathbb{R}_+$ from the unit cube $[0,1]^s$, such that the projection of points onto the first $s-1$
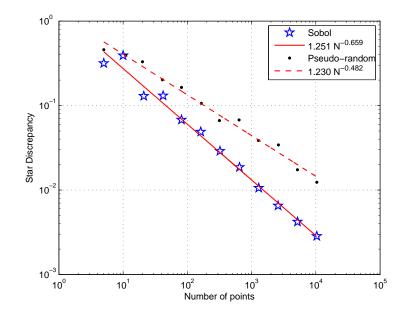
FIG 2. *Convergence order of the star discrepancy of Example 1.*

coordinates has distribution H. More precisely, let $T : [0,1]^s \to \mathbb{R}^s$ be the transformation given by $\boldsymbol{z} = T(\boldsymbol{u})$, where $\boldsymbol{z} = (z_1, \ldots, z_s), \boldsymbol{u} = (u_1, \ldots, u_s)$ and

$$
\begin{aligned}
z_1 &= F_1^{-1}(u_1), \\
z_j &= F_j^{-1}(u_j|u_1, \ldots, u_{j-1}),\ 2 \leq j \leq s-1, \\
z_s &= u_s H(z_1, \ldots, z_{s-1}).
\end{aligned}
\tag{1}
$$

The first $s - 1$ coordinates are produced by the inverse Rosenblatt transformation which converts the points from the unit cube $[0,1]^{s-1}$ to $\mathbb{R}^{s-1}$. The sth coordinate is uniformly distributed on the line

$$\{(1-v)(z_1, \ldots z_{s-1}, 0) + v(z_1, \ldots z_{(s-1)}, H(z_1, \ldots, z_{s-1}), 0 \leq v \leq 1\}$$

if u_s is uniformly distributed in $[0,1]$. More details with respect to the Rosenblatt transformation and extensions can be found in [8, 29, 35].

Algorithm 3 (Deterministic acceptance-rejection algorithm in $\mathbb{R}^s$)**.** Let an unnormalized target density function $\psi : \mathbb{R}^{s-1} \to \mathbb{R}_+$, where $s \geq 2$, be given. Let H be a proposal density $H : \mathbb{R}^{s-1} \to \mathbb{R}_+$, such that there exists a constant $L < \infty$ such that $\psi(\boldsymbol{z}) \leq LH(\boldsymbol{z})$ for all $\boldsymbol{z} \in \mathbb{R}^{s-1}$. Let $A = \{\boldsymbol{z} \in \mathbb{R}^s : \psi(z_1, \ldots, z_{s-1}) \geq LH(z_1, \ldots, z_{s-1})z_s\}$. Suppose we aim to obtain approximately N samples from ψ.

   i) Let $M = b^m \geq \lceil N \int_{[0,1]^{s-1}} H(\boldsymbol{x})d\boldsymbol{x}/(\int_{[0,1]^{s-1}} \psi(\boldsymbol{x})/Ld\boldsymbol{x}) \rceil$, where $m \in \mathbb{N}$

is the smallest integer satisfying this inequality. Generate a (t, m, s)-net $Q_{m,s} = \{\boldsymbol{x}_0, \boldsymbol{x}_1, \dots, \boldsymbol{x}_{M-1}\}$ in base b.

ii) Transform the points into $\mathbb{R}^{s-1} \times \mathbb{R}_+$ from $[0,1]^s$ using the transformation T given in (1) to obtain $\{T(\boldsymbol{x}_0), T(\boldsymbol{x}_1), \dots, T(\boldsymbol{x}_{M-1})\}$.

iii) Take the acceptance-rejection method for the sample $T(\boldsymbol{x}_n)$ with respect to H and ψ in $\mathbb{R}^{s-1} \times \mathbb{R}_+$, i.e. accept the point $T(\boldsymbol{x}_n)$ if $T(\boldsymbol{x}_n) \in A$, otherwise reject. Let $P_N^{(s)} = A \cap T(Q_{m,s}) = \{\boldsymbol{z}_0, \boldsymbol{z}_1, \dots, \boldsymbol{z}_{N-1}\}$.

iv) Project the points $P_N^{(s)}$ we accepted onto the first $(s-1)$-dimensional space. Denote the first $s-1$ coordinates of the points we accept by the acceptance-rejection method by $P_N^{(s-1)} = \{\boldsymbol{y}_0, \boldsymbol{y}_1, \dots, \boldsymbol{y}_{N-1}\} \subseteq \mathbb{R}^{s-1}$.

v) Return the point set $P_N^{(s-1)}$.

We provide an example to demonstrate the performance of Algorithm 3.

Example 2. Let the target density function be given by

$$\psi(x_1, x_2) = \begin{cases} \dfrac{4}{\pi} e^{-(x_1+x_2)}(x_1 x_2)^{1/2}, & x_1, x_2 > 0, \\ 0, & \text{otherwise.} \end{cases}$$

The proposal density function H, which we use to do the acceptance-rejection to generate samples of $\psi(x_1, x_2)$, is chosen as

$$H(x_1, x_2) = \begin{cases} \dfrac{1}{4}, & 0 \le x_1, x_2 \le 1, \\ \dfrac{1}{4x_2^2}, & 0 \le x_1 \le 1, x_2 > 1, \\ \dfrac{1}{4x_1^2}, & x_1 > 1, 0 \le x_2 \le 1, \\ \dfrac{1}{4x_1^2 x_2^2}, & x_1, x_2 > 1, \\ 0, & \text{otherwise.} \end{cases}$$

For this choice of H, we use the transform T defined in Equation (1) to obtain samples from H. The sample $(x_{j,1}, x_{j,2})$ is given by the following transformation

$$x_{j,1} = \begin{cases} 2u_{j,1}, & 0 \le u_{j,1} \le 1/2, \\ 1/2(1-u_{j,1}), & 1/2 < u_{j,1} \le 1, \end{cases}$$

$$x_{j,2} = \begin{cases} 2u_{j,2}, & 0 \le u_{j,2} \le 1/2, \\ 1/2(1-u_{j,2}), & 1/2 < u_{j,2} \le 1. \end{cases}$$

Note that $(u_{j,1}, u_{j,2})$ is the driver sequence given by a $(t, m, 2)$-net in base b.

The order of the star discrepancy is demonstrated in Figure 3 where N is the number of accepted samples. The numerical experiments show that the star discrepancy converges at a rate of $N^{-0.720}$ for this example using quasi-Monte Carlo samples as proposal. The RAR algorithm converges with order $N^{-0.390}$. Again, the DAR sampler outperforms the RAR sampler.
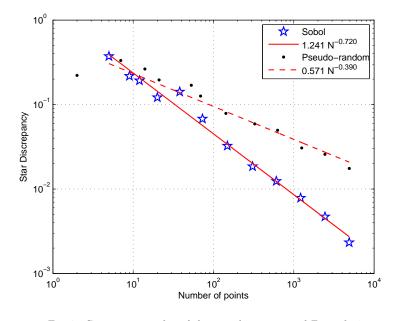
FIG 3. *Convergence order of the star discrepancy of Example 2.*

2.4. A deterministic reduced acceptance-rejection sampler

In this subsection we consider an extension of the DAR sampler. The random version of this reduced method was recently introduced by Barekat and Caflisch in [2]. For a target density function ψ, we carefully select H such that for $\psi - H$ and H the inverse CDF can be computed. For the case $\psi(\boldsymbol{x}) > H(\boldsymbol{x})$, we write $\psi = (\psi - H) + H$ and get samples according to $\psi - H$ and H respectively.

Figure 4 illustrates this method. The sample sets of ψ can be divided into three subsets, $R_{1,1}, R_{1,2}$ and $R_{2,2}$, where $R_{1,2}$ and $R_{2,2}$ can be directly generated by using the inverse CDF of H and $\psi - H$ in a certain range. The acceptance-rejection method is only used to obtain $R_{1,1}$. Compared with the ordinary acceptance-rejection sampler, one obvious merit of this method is that we do not require $\psi(\boldsymbol{x}) \leq H(\boldsymbol{x})$ in the whole domain. Also, this method might give better convergence rates since $R_{1,2}$ and $R_{2,2}$ are obtained via inversion and therefore have low discrepancy. Algorithm 4 gives a simple version of the improved method. More discussion of a general version is available in Section 5.

Algorithm 4 (Deterministic reduced acceptance-rejection (DRAR) algorithm)**.** Let $\psi : [0,1] \to \mathbb{R}_+$ be a target density. Choose a proposal density H such that $\psi - H$ and H can be sampled from directly. Let

$$\mathcal{S} := \{x \in [0,1] : \psi(x) < H(x)\}$$

and

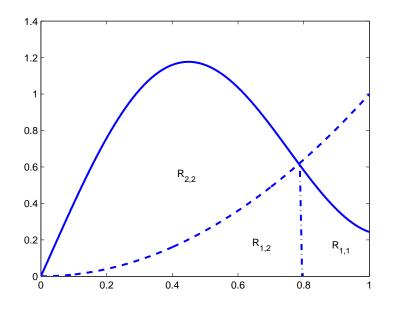$$\mathcal{L} := \{x \in [0,1] : \psi(x) \geq H(x)\}.$$

FIG 4. *The solid line represents the target density ψ and the dashed line is the proposal density H.*

Assume that $\int_{\mathcal{S}} \psi(x)dx / \int_{[0,1]} \psi(x)dx$, $\int_{\mathcal{L}} H(x)dx / \int_{[0,1]} \psi(x)dx$ and $\int_{\mathcal{L}} (\psi - H)(x)dx / \int_{[0,1]} \psi(x)dx$ can be calculated or estimated. Let $F_{H,\mathcal{S}}^{-1}, F_{H,\mathcal{L}}^{-1}$ be the inverse CDF of the proposal density H in the domain $\mathcal{S}$ and $\mathcal{L}$ respectively and $F_{\psi-H,\mathcal{L}}^{-1}$ be the inverse CDF with respect to $\psi - H$ in $\mathcal{L}$. Suppose we aim to generate approximately N samples from ψ. Let

$$N_1 = \left\lceil N \frac{\int_{\mathcal{S}} \psi(x)dx}{\int_{[0,1]} \psi(x)dx} \right\rceil, N_2 = \left\lceil N \frac{\int_{\mathcal{L}} H(x)dx}{\int_{[0,1]} \psi(x)dx} \right\rceil \text{ and } N_3 = \left\lceil N \frac{\int_{\mathcal{L}} (\psi - H)(x)dx}{\int_{[0,1]} \psi(x)dx} \right\rceil.$$

i) Let $\{\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots\} \subset [0,1]^2$ be a $(t,2)$-sequence in base b.

ii) Use the acceptance-rejection method with the target density ψ and the proposal density H on the domain $\mathcal{S}$ using $\{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{M-1}\}$ as driver sequence. Choose M such that N_1 points are accepted by the DAR algorithm. Compute $F_{H,\mathcal{S}}^{-1}(\boldsymbol{x}_n)$ for $n = 0, 1, \ldots, M-1$. Let $\boldsymbol{z}_0, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_{N_1-1}$ be the accepted points. Label the point set as $R_{1,1}$.

iii) Compute the points $F_{H,\mathcal{L}}^{-1}(\boldsymbol{x}_n)$ for $n = 0, 1, \ldots, N_2 - 1$. Let $R_{1,2} = \{F_{H,\mathcal{L}}^{-1}(\boldsymbol{x}_n) : 0 \leq n < N_2\}$.

iv) Compute the points $F_{\psi-H,\mathcal{L}}^{-1}(\boldsymbol{x}_n)$ for $n = 0, 1, \ldots, N_3 - 1$. Let $R_{2,2} = \{F_{\psi-H,\mathcal{L}}^{-1}(\boldsymbol{x}_n) : 0 \leq n < N_3\}$.

v) Project the points in $R_N = R_{1,1} \cup R_{1,2} \cup R_{2,2}$ onto the first coordinate. Return the point set $R_N^{(1)}$.
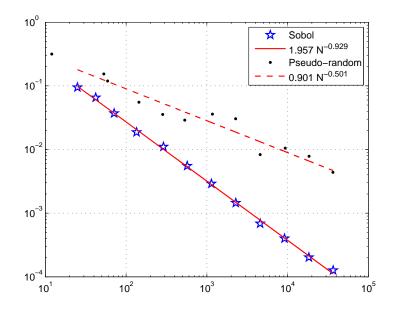
Fig 5. *Convergence order of the star discrepancy of Example 3.*

Since the inverse transform is a measure-preserving transformation, it can preserve the uniformities of the driver sequence. Thus $R_{1,2}$ and $R_{2,2}$ are low discrepancy point sets. The following example verifies the efficiency of the DRAR algorithm. A theoretical result about the discrepancy properties of samples obtained by this class of algorithms is provided in Theorem 5.

Example 3. Let $\psi(x) = \sin(4x) + x^2$ be a density function defined on $[0, 1]$. Instead of seeking a proposal density H such that $\psi(x) \leq H(x)$, we notice that inversion can be implied to $\sin(4x)$ and x^2 independently. However, it can not work for their sum. Choose $H(x) = x^2$. We only do deterministic acceptance-rejection with respect to the target density ψ and proposal density H in the subinterval $\mathcal{S} = (\pi/4, 1]$. In the remaining range $\mathcal{L} = [0, \pi/4]$, we apply the inverse transformation on H and $\psi - H$ to obtain samples based on a deterministic driver sequence.

The discrepancy of the point set generated by Algorithm 4 converges at the rate of $N^{-0.929}$, which is significantly better than the $N^{-0.501}$ convergence rate of a random driver sequence, see Figure 5.

3. Background on discrepancy theory and (t, m, s)-nets

In this section we first establish some notation and some useful definitions and then obtain theoretical results. First we introduce the definition of (t, m, s)-nets and (t, s)-sequence in base b (see [11]) which we use as the driver sequences. The

following fundamental definitions of elementary interval and fair sets are used to define a (t, m, s)-net and (t, s)-sequence in base b.

Definition 1 (b-asdic elementary interval). Let $b \geq 2$ be an integer. An s-dimensional b-asdic elementary interval is an interval of the form

$$\prod_{i=1}^{s} \left[\frac{a_i}{b^{d_i}}, \frac{a_i + 1}{b^{d_i}} \right)$$

with integers $0 \leq a_i < b^{d_i}$ and $d_i \geq 0$ for all $1 \leq i \leq s$. If $d_1, \ldots, d_s$ are such that $d_1 + \cdots + d_s = k$, then we say that the elementary interval is of order k.

Definition 2 (fair sets). For a given set $P_N = \{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{N-1}\}$ consisting of N points in $[0, 1)^s$, we say for a subset J of $[0, 1)^s$ to be fair with respect to P_N, if

$$\frac{1}{N} \sum_{n=0}^{N-1} 1_J(\boldsymbol{x}_n) = \lambda(J),$$

where $1_J(\boldsymbol{x}_n)$ is the indicator function of the set J and λ is the Lebesgue measure.

Definition 3 ((t, m, s)-nets in base b). For a given dimension $s \geq 1$, an integer base $b \geq 2$, a positive integer m and an integer t with $0 \leq t \leq m$, a point set $Q_{m,s}$ of b^m points in $[0, 1)^s$ is called a (t, m, s)-nets in base b if the point set $Q_{m,s}$ is fair with respect to all b-asdic s-dimensional elementary intervals of order at most $m - t$.

Definition 4 ((t, s)-sequence). For a given dimension $s \geq 1$, an integer base $b \geq 2$ and a positive integer t, a sequence $\{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots\}$ of points in $[0, 1)^s$ is called a (t, s)-sequence in base b if for all integers $m \geq t$ and $k \geq 0$, the point set consisting of the points $\boldsymbol{x}_{kb^m}, \ldots, \boldsymbol{x}_{kb^m + b^m - 1}$ forms a (t, m, s)-net in base b.

The concept of discrepancy is introduced in [16] to measure the deviation of a sequence from the uniform distribution. Now we give the definition of the so-called star discrepancy which enables us to distinguish the quality of point sets with respect to the uniform distribution.

Definition 5 (star discrepancy). Let $P_N = \{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{N-1}\}$ be a point set in $[0, 1)^s$. The star discrepancy D_N^* is defined by

$$D_N^*(P_N) = \sup_{J \subset [0,1)^s} \left| \frac{1}{N} \sum_{n=0}^{N-1} 1_J(\boldsymbol{x}_n) - \lambda(J) \right|,$$

where the supremum is taken over all $J = \prod_{i=1}^{s} [0, \beta_i) \subseteq [0, 1)^s$.

See Figure 6 for an illustration of the concept of discrepancy in the unit square.

If we extend the supremum in Definition 5 over all convex sets in $[0, 1]^s$, we get another interesting discrepancy, the so-called isotropic discrepancy. It is another measure of the distribution properties of point sets with respect to convex sets.
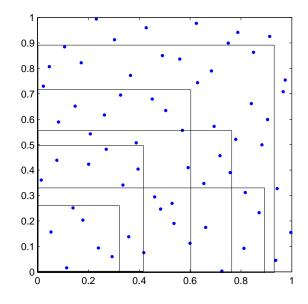
FIG 6. *The discrepancy measures the difference between the proportion of points in each rectangle J and the Lebesgue measure of J. The star discrepancy is defined by the supremum of discrepancies over all rectangles J. If we change J to be convex sets, we obtain the so-called isotropic discrepancy.*

Definition 6 (isotropic discrepancy). Let $P_N = \{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{N-1}\}$ be a point set in $[0,1)^s$. The isotropic discrepancy J_N is defined to be

$$J_N(P_N) = \sup_{J \in \mathcal{C}} \left| \frac{1}{N} \sum_{n=0}^{N-1} 1_J(\boldsymbol{x}_n) - \lambda(J) \right|,$$

where $\mathcal{C}$ is the family of all convex subsets of $[0,1)^s$.

For further reading about the definition and properties of discrepancy, we refer for instance to [11, 16].

For our purposes here we need the definition of pseudo-convex sets which we introduce in the following (see also [1, Definition 2] and Figure 7 for an example).

Definition 7 (pseudo-convex set). Let A be an open subset of $[0,1]^s$ such that there exists a collection of p convex subsets $A_1, \ldots, A_p$ of $[0,1]^s$ satisfying

1. $A_i \cap A_j = \emptyset$ for $i \neq j$;
2. $A \subseteq (A_1 \cup \cdots \cup A_p)$ of $[0,1]^s$;
3. either A_j is a convex part of A ($A_j \subseteq A$ for $j = 1, \ldots, q$) or the complement of A with respect to A_j, $A_j' = A_j \backslash A$ is convex.

Then A is called a pseudo-convex set and $A_1, \ldots, A_p$ is an admissible convex covering for A with p parts and with q convex parts of A.
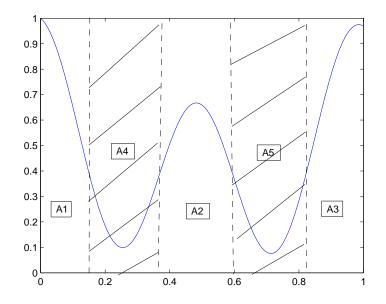
FIG 7. *Shows a pseudo-convex set in the unit square given by the area below graph of the density function and its admissible convex covering. Let A be given by the graph under the curve. Then $A_i, i = 1, \ldots, 5$, is an admissible convex covering of A, where A_1, A_2 and A_3 are convex parts in A, but A_4 and A_5 are the shadowed rectangles covering the remaining part of A. The regions $A_4 \backslash A$ and $A_5 \backslash A$ are convex.*

Remark 1. For convenience, we call a nonnegative function pseudo-convex if and only if the region below its graph is a pseudo-convex set.

Next we present a bound on the isotropic discrepancy of points generated by (t, m, s)-nets. A detailed proof is given in Appendix B.1.

Lemma 1. *Let the point set $Q_{m,s} = \{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{M-1}\} \subseteq [0,1]^s$ be a (t, m, s)-net in base b where $M = b^m$. For the isotropic discrepancy of $Q_{m,s}$ we have*

$$J_M(Q_{m,s}) \leq 2sb^{t/s}M^{-1/s}.$$

A slightly weaker result than Lemma 1 can also be obtained from [28, Korollar 3].

Lemma 2. *For any point set P_N in $[0,1]^s$ we have*

$$J_N(P_N) \leq 2s \left(\frac{4s}{s-1} \right)^{(s-1)/s} (D_N^*(P_N))^{1/s}.$$

Further it is known from [20] that the star discrepancy of a (t, m, s)-net $Q_{m,s}$ in base b, where $M = b^m$, satisfies

$$D_M^*(Q_{m,s}) \leq M^{-1}b^t(\log M)^{s-1}\frac{b^s}{(b+1)2^s(s-1)!(\log b)^{s-1}} + C_sM^{-1}b^t(\log M)^{s-2},$$

for some constant $C_s > 0$. These two inequalities therefore yield a convergence rate of order $M^{-1/s}(\log M)^{1-1/s}$.

The following lemma will be used to get a discrepancy bound for a point set on a pseudo-convex set. It is an extension of [1, Lemma 5] to the unit cube.

Lemma 3. *Let A be a pseudo-convex subset of $[0,1]^s$ with admissible convex covering of p parts with q convex parts of A. Then for any point set $P_N = \{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{N-1}\} \subseteq [0,1]^s$ we have*

$$\left| \frac{1}{N} \sum_{n=0}^{N-1} 1_A(\boldsymbol{x}_n) - \lambda(A) \right| \leq (2p - q) J_N(P_N).$$

4. Discrepancy investigation of the deterministic acceptance-rejection sampler

The first result we get is a discrepancy bound with respect to the target density of samples generated by the acceptance-rejection algorithm with deterministic driver sequences. The star discrepancy of points generated by the acceptance-rejection algorithm with respect to the target density converges at the rate of $N^{-1/s}$, where N is the number of accepted samples. See Theorem 1 for details. The proof uses a bound on the discrepancy of our driver sequence with respect to convex sets (which is called isotropic discrepancy, see Definition 6 for details).

4.1. Upper bound

Let an unnormalized density function $\psi : [0,1]^{s-1} \to \mathbb{R}_+$ be pseudo-convex, and $\int_{[0,1]^{s-1}} \psi(\boldsymbol{z}) d\boldsymbol{z} > 0$, but not necessarily 1. Assume that there exists a constant $L < \infty$ such that $\psi(\boldsymbol{x}) \leq L$ for all $\boldsymbol{x} \in [0,1]^{s-1}$. Let the subset under the graph of ψ/L be defined as

$$A = \{\boldsymbol{x} \in [0,1]^s : \psi(x_1, \ldots, x_{s-1}) \geq L x_s\}, \tag{2}$$

which is pseudo-convex in $[0,1]^s$ as ψ is a pseudo-convex function. Assume that there is an admissible convex covering of A with p parts and with q convex parts of A. Without loss of generality, let $A_1, \ldots, A_q$ be the convex subsets of A and $A_{q+1}, \ldots, A_p$, such that $A'_j = A_j \backslash A$ is convex for $q + 1 \leq j \leq p$.

The definition of the star discrepancy of a point set $\{\boldsymbol{y}_0, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{N-1}\}$ with respect to a density function ψ is given as follows.

Definition 8. *Let $\psi : [0,1]^{s-1} \to \mathbb{R}_+$ be an unnormalized target density. Let $\{\boldsymbol{y}_0, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{N-1}\}$ be a point set in $[0,1]^{s-1}$. The star discrepancy of $\{\boldsymbol{y}_0, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{N-1}\}$ with respect to the density ψ is defined by*

$$D_{N,\psi}^*(P_N^{(s-1)}) = \sup_{\boldsymbol{t} \in [0,1]^{s-1}} \left| \frac{1}{N} \sum_{n=0}^{N-1} 1_{[\boldsymbol{0},\boldsymbol{t})}(\boldsymbol{y}_n) - \frac{1}{C} \int_{[\boldsymbol{0},\boldsymbol{t})} \psi(\boldsymbol{z}) d\boldsymbol{z} \right|,$$

where $C = \int_{[0,1]^{s-1}} \psi(\boldsymbol{z}) d\boldsymbol{z}$ and $[\boldsymbol{0},\boldsymbol{t}) = \prod_{j=1}^{s-1} [0, t_j)$.

Remark 2. Note that $\frac{1}{C}\psi$ is a probability density function on $[0,1]^{s-1}$. Thus the discrepancy in Definition 8 measures the difference between the distribution $\frac{1}{C}\psi$ and the empirical distribution of the sample points with respect to the test sets $[\mathbf{0}, \mathbf{t})$ for $\mathbf{t} \in [0,1]^{s-1}$.

Theorem 1. *Let the unnormalized density function $\psi : [0,1]^{s-1} \to \mathbb{R}_+$, with $s \geq 2$, be pseudo-convex. Assume that there is an admissible convex covering of A given by Equation (2) with p parts and with q convex parts of A. Then the discrepancy of the point set $\{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_{N-1}\} \subseteq [0,1]^{s-1}$ generated by Algorithm 2 using a (t,m,s)-net in base b, for large enough N, satisfies*

$$D^*_{N,\psi}(P_N^{(s-1)}) \leq 8C^{-1}Lsb^{t/s}(2p-q)N^{-1/s},$$

where $C = \int_{[0,1]^{s-1}} \psi(\mathbf{z})d\mathbf{z}$ and $\psi(\mathbf{x}) \leq L$ for all $\mathbf{x} \in [0,1]^{s-1}$.

We postpone the proof of this theorem to Appendix B.1.

4.2. Lower bound

In this section, we provide a lower bound on the star discrepancy with respect to a convex density function. The general idea is to find, for a given driver point set, a density function satisfying a certain convergence rate.

Theorem 2. *Let P_M be an arbitrary point set in $[0,1]^s$. Then there exists a concave density function ψ defined in $[0,1]^{s-1}$ such that, for N samples generated by the acceptance-rejection algorithm with respect to P_M and ψ, we have*

$$D^*_{N,\psi}(P_N) \geq c_s N^{-\frac{2}{s+1}},$$

where $c_s > 0$ is independent of N and P_M but depends on s.

A detailed proof is provided in Appendix B.2. We would like to point out that the lower bound also limits the convergence rates which we can obtain in our current approach via convex sets. Note that a concave function is also pseudo-convex as defined in Remark 1.

Additionally, note that [3] (in dimension $s = 2$) and [38] (for dimension $s > 2$) showed the existence of points with discrepancy with respect to convex sets bounded from above by $N^{-2/(s+1)}(\log N)^{c(s)}$ (where $c(s)$ is a function of only s). This would yield an improvement of our results from $N^{-1/s}$ to $N^{-2/(s+1)}(\log N)^{c(s)}$, however, those constructions are not explicit and can therefore not be used in computation.

4.3. Generalization to real state space

We consider now the case where the target density is defined on $\mathbb{R}^{s-1}$ with $s \geq 2$. The aim is to show a discrepancy bound on samples generated by the deterministic acceptance-rejection method. The discrepancy with respect to a given density function $\psi : \mathbb{R}^{s-1} \to \mathbb{R}_+$ is defined as follows.

Definition 9. Let $P_N = \{z_0, z_1, \ldots, z_{N-1}\}$ be a point set in $\mathbb{R}^{s-1}$. Let $\psi : \mathbb{R}^{s-1} \to \mathbb{R}_+$ be an unnormalized probability density function. Then the star discrepancy $D^*_{N,\psi}(P_N)$ is defined by

$$D^*_{N,\psi}(P_N) = \sup_{\boldsymbol{t} \in \mathbb{R}^{s-1}} \left| \frac{1}{N} \sum_{n=0}^{N-1} 1_{(-\boldsymbol{\infty}, \boldsymbol{t}]}(\boldsymbol{z}_n) - \frac{1}{C} \int_{(-\boldsymbol{\infty}, \boldsymbol{t}]} \psi(\boldsymbol{z}) d\boldsymbol{z} \right|,$$

where $C = \int_{\mathbb{R}^{s-1}} \psi(\boldsymbol{z}) d\boldsymbol{z}$ and $(-\boldsymbol{\infty}, \boldsymbol{t}] = \prod_{j=1}^{s-1}(-\infty, t_j]$ for $\boldsymbol{t} = (t_1, \ldots, t_{s-1})$.

We use the transformation T given in Equation (1) to generate samples of H. For the sake of investigating discrepancy, the following result is helpful. The lemma shows that the transformations T and its inversion T^{-1} are both measure-preserving. For the proofs later, we assume that the proposal density H is a product measure, i.e. $H = \prod_{j=1}^{s-1} H_j$, where H_j is the marginal density with respect to z_j.

Lemma 4. *The transformation T from the s-dimensional unit cube to $\mathbb{R}^{s-1} \times \mathbb{R}_+$ given in (1) is measure-preserving, i.e. $Volume(T(D)) = Volume(D)$ holds for any measurable set $D \subseteq [0,1]^s$. This is true for T^{-1} as well.*

To prove a bound on the discrepancy of the samples generated by Algorithm 3, the following assumption is needed.

Assumption 1. Let ψ be the target density and H be a product measure proposal density function, which is chosen such that its inverse CDF can be computed. Let $A = \{\boldsymbol{z} \in \mathbb{R}^s : \psi(z_1, \ldots, z_{s-1}) \geq Lz_s H(z_1, \ldots, z_{s-1})\}$ and the transformation T^{-1} is defined as the inversion of transform T. Then we assume that $T^{-1}(A)$ is pseudo-convex.

As the mappings T and T^{-1} are measure preserving, and since there are the same number of samples in an arbitrary subset $D \subseteq [0,1]^s$ and the corresponding subset $T(D) \subseteq \mathbb{R}^{s-1} \times \mathbb{R}_+$, we can consider the discrepancy in the unit cube instead of that in $\mathbb{R}^{s-1} \times \mathbb{R}_+$. Following by similar proof arguments as for Theorem 1 and Theorem 2, we obtain the same discrepancy bounds including an upper bound and a lower bound for the general density ψ defined in the real state space $\mathbb{R}^{s-1}$.

Theorem 3. *Let the unnormalized target density $\psi : \mathbb{R}^{s-1} \to \mathbb{R}_+$ and the proposal density $H : \mathbb{R}^{s-1} \to \mathbb{R}_+$ satisfy Assumption 1. Then the discrepancy of the point set $P_N^{(s-1)} = \{\boldsymbol{y}_0, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{N-1}\} \subseteq \mathbb{R}^{s-1}$ generated by Algorithm 3 satisfies*

$$D^*_{N,\psi}(P_N^{(s-1)}) \leq 8LC^{-1} s b^{t/s} (2p-q) N^{-1/s},$$

for N large enough, where $C = \int_{\mathbb{R}^{s-1}} \psi(\boldsymbol{z}) d\boldsymbol{z}$ and L is such that $\psi(\boldsymbol{x}) \leq LH(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^{s-1}$.

Theorem 4. *Let H be a product density function defined on $\mathbb{R}^{s-1}$. Let T be the transformation given in Equation (1) associated to H. Let P_M be an arbitrary point set in $[0,1]^s$, then $T(P_M)$ is a point set in $\mathbb{R}^{s-1}$. Then there ex-*

ists an unnormalized density function ψ defined in $\mathbb{R}^{s-1}$ satisfying the assumption in Theorem 3 such that the star discrepancy of the points generated by the acceptance-rejection sampler with respect to ψ and H satisfies

$$D_{N,\psi}^*(P_N^{(s-1)}) \geq c_s N^{-\frac{2}{s+1}},$$

where c_s is independent of N and P_M, but only dependent on s.

5. Discrepancy properties of the deterministic reduced acceptance-rejection sampler

Algorithm 4 can be extended to a more general case. Consider the target density $\psi(\boldsymbol{x}) = \sum_{\ell=1}^k H_\ell(\boldsymbol{x})$, where $\boldsymbol{x} \in D \subset \mathbb{R}^s$. If it is possible to sample from $H_\ell(\boldsymbol{x})$ individually and the expectations of H_ℓ can be calculated or estimated with low cost, then we can use an embedded deterministic reduced acceptance-rejection sampler in each step. Let

$$\mathcal{S}_\ell = \{\boldsymbol{x} \in D : \psi_{k-\ell+1}(\boldsymbol{x}) < H_\ell(\boldsymbol{x})\},$$
and $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ (3)
$$\mathcal{L}_\ell = \{\boldsymbol{x} \in D : \psi_{k-\ell+1}(\boldsymbol{x}) \geq H_\ell(\boldsymbol{x})\},$$

where $\psi_{k-\ell+1}(\boldsymbol{x}) = \sum_{i=\ell}^k H_i(\boldsymbol{x})$ for $\ell = 1, \ldots, k-1$, and, in particular, ψ_k is the target density.

Suppose we aim to sample N points from the target density ψ. The sample set can be divided into two types, namely, points generated from the sets $\mathcal{S}_\ell$'s and $\mathcal{L}_\ell$'s respectively. We apply a deterministic acceptance-rejection method given in Algorithm 3 in each $\mathcal{S}_\ell$ with respect to $\psi_{k-\ell+1}$ and H_ℓ. Note that we get $\lceil N \int_{\mathcal{S}_\ell} \psi_{k-\ell+1}(\boldsymbol{x})d\boldsymbol{x} / \int_D \psi_k(\boldsymbol{x})d\boldsymbol{x} \rceil$ points from $\mathcal{S}_\ell$ for $\ell = 1, \ldots, k-1$. For sampling from $\mathcal{L}_\ell$, the remaining samples come from applying the inverse transformation of H_ℓ in $\mathcal{L}_\ell$. Then we obtain additional $\lceil N \int_{\mathcal{L}_\ell} H_\ell(\boldsymbol{x})d\boldsymbol{x} / \int_D \psi_k(\boldsymbol{x})d\boldsymbol{x} \rceil$ points from $\mathcal{L}_\ell$ for $\ell = 1, \ldots, k$. We conduct the procedure inductively until we get samples from $H_k(\boldsymbol{x})$. We assume that $\int_{\mathcal{S}_\ell} \psi_{k-\ell+1}(\boldsymbol{x})d\boldsymbol{x} / \int_D \psi_k(\boldsymbol{x})d\boldsymbol{x}$ and $\int_{\mathcal{L}_\ell} H_\ell(\boldsymbol{x})d\boldsymbol{x} / \int_D \psi_k(\boldsymbol{x})d\boldsymbol{x}$ can be calculated or estimated.

The following algorithm is an extension of the DRAR algorithm, which summarizes the embedding idea.

Algorithm 5. Let $\psi(\boldsymbol{x}) = \sum_{\ell=1}^k H_\ell(\boldsymbol{x}), \boldsymbol{x} \in D \subset \mathbb{R}^{s-1}$, be a target density we aim to sample from. Define $\psi_{k-\ell+1}(\boldsymbol{x}) = \sum_{i=\ell}^k H_i(\boldsymbol{x})$ for $\ell = 1, \ldots, k$. Denote $\mathcal{S}_\ell$ and $\mathcal{L}_\ell$ like in Equation (3) and assume that $\int_{\mathcal{S}_\ell} \psi_{k-\ell+1}(\boldsymbol{x})d\boldsymbol{x} / \int_D \psi(\boldsymbol{x})d\boldsymbol{x}$ and $\int_{\mathcal{L}_\ell} H_\ell(\boldsymbol{x})d\boldsymbol{x} / \int_D \psi(\boldsymbol{x})d\boldsymbol{x}$ can be calculated or estimated. Further assume that we can sample from H_ℓ individually by applying the transformation $T_{H_\ell,\mathcal{S}_\ell}$ and $T_{H_\ell,\mathcal{L}_\ell}$ given in Equation (1) in $\mathcal{S}_\ell$ and $\mathcal{L}_\ell$ respectively. Suppose we aim to generate N samples from ψ. Let

$$N_{1,\ell} = \left\lceil N \frac{\int_{\mathcal{S}_\ell} \psi_{k-\ell+1}(\boldsymbol{x})d\boldsymbol{x}}{\int_D \psi(\boldsymbol{x})d\boldsymbol{x}} \right\rceil \text{ and } N_{2,\ell} = \left\lceil N \frac{\int_{\mathcal{L}_\ell} H_\ell(\boldsymbol{x})d\boldsymbol{x}}{\int_D \psi(\boldsymbol{x})d\boldsymbol{x}} \right\rceil.$$

For ℓ from 1 to k do:

i)   Let $\{\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots\} \subset [0,1]^s$ be a (t,s)-sequence in base b.

ii) Compute $T_{H_\ell, \mathcal{S}_\ell}(\boldsymbol{x}_n)$ for $n = 0, 1, 2, \ldots$. Use the acceptance-rejection method with respect to $\psi_{k-\ell+1}$ and H_ℓ on the domain $\mathcal{S}_\ell$ using $\{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{M-1}\}$ as driver sequence. Choose M such that $N_{1,\ell}$ points are accepted by the DAR algorithm. Let $R_{1,\ell} = \{\boldsymbol{z}_0, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_{N_{1,\ell}-1}\}$ be the accepted points.

iii) Compute $T_{H_\ell, \mathcal{L}_\ell}(\boldsymbol{x}_n)$ for $n = 0, 1, \ldots, N_{2,\ell} - 1$. Let $R_{2,\ell} = \{T_{H_\ell, \mathcal{L}_\ell}(\boldsymbol{x}_n) : 0 \leq n < N_{2,\ell}\}$.

Let $R_N^{(s)} = \bigcup_{\ell=1}^{k} (R_{1,\ell} \cup R_{2,\ell})$ and let $R_N^{(s-1)}$ denote the projection of $R_N^{(s)}$ onto the first $s-1$ coordinates. Return the set $R_N^{(s-1)}$.

Now we consider the discrepancy properties of sample points produced by this algorithm. Note that the sample set of $\psi = \sum_{\ell=1}^{k} H_\ell$ can be decomposed into several subsets with different star discrepancy.

Theorem 5. *For a given target density $\psi(\boldsymbol{x}) = \sum_{\ell=1}^{k} H_\ell(\boldsymbol{x})$, for $\boldsymbol{x} \in D \subset \mathbb{R}^{s-1}$, let $\psi_{k-\ell+1}(\boldsymbol{x}) = \sum_{i=\ell}^{k} H_i(\boldsymbol{x})$, where ψ and H_ℓ, for $1 \leq \ell \leq k$, are piecewise continuous. Let $\mathcal{S}_\ell$ and $\mathcal{L}_\ell$ be given by (3). Let $R_N^{(s-1)}$ be the sample set generated by Algorithm 5, where*

$$N_{1,\ell} = \left\lceil N \int_{\mathcal{S}_\ell} \psi_{k-\ell+1}(\boldsymbol{x}) d\boldsymbol{x} / \int_D \psi_k(\boldsymbol{x}) d\boldsymbol{x} \right\rceil,$$

which is the number of points generated from $\mathcal{S}_\ell$, and

$$N_{2,\ell} = \left\lceil N \int_{\mathcal{L}_\ell} H_\ell(\boldsymbol{x}) d\boldsymbol{x} / \int_D \psi_k(\boldsymbol{x}) d\boldsymbol{x} \right\rceil,$$

which is the number of points generated from $\mathcal{L}_\ell$ for $\ell = 1, \ldots, k$. Assume that $N_{1,\ell}$ and $N_{2,\ell}$ can be calculated or estimated for the given target density ψ and N. Then we have

$$D_{N,\psi}^*(R_N^{(s-1)}) \leq \sum_{\ell=1}^{k-1} \frac{N_{1,\ell}}{N} D_{\mathcal{S}_\ell, \psi_{k-\ell+1}}^* + \sum_{\ell=1}^{k} \frac{N_{2,\ell}}{N} D_{\mathcal{L}_\ell, H_\ell}^* + \frac{1}{N},$$

where $D_{\mathcal{S}_\ell, \psi_{k-\ell+1}}^*$ and $D_{\mathcal{L}_\ell, H_\ell}^*$ is the discrepancy of the samples in $\mathcal{S}_\ell$ and $\mathcal{L}_\ell$ respectively.

The proof of Theorem 5 is given in Appendix B.3. Note that this method achieves an improved acceptance rate of points since we are only rejecting points in a certain range. For the remaining domain, we get samples by applying the inverse transform. To be more exact, all point sets from $\mathcal{L}_\ell$ have low discrepancy since the inverse transformation is directly applied with respect to H_ℓ for $\ell = 1, \ldots, k$. Now we consider the star discrepancy of points generated from $\mathcal{S}_\ell$.

The following result from [20] gives an improved upper bound on the star discrepancy on the first M terms of a (t,s)-sequence in base b with $s \geq 2$.

Lemma 5. *The star discrepancy of the first M terms of a (t, s)-sequence P_M in base b with $s \geq 2$ satisfies*

$$D_M^*(P_M) \leq M^{-1}b^t(\log M)^s \frac{b^s(b-1)}{(b+1)2^{s+1}(s)!(\log b)^s} + C_s M^{-1}b^t(\log M)^{s-1},$$

for some constant $C_s > 0$ depending only on s.

With the help of Lemma 2, we obtain a bound on the isotropic discrepancy of the first M points of a (t, s)-sequence.

Lemma 6. *Let the point set $P_M = \{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{M-1}\} \subset [0, 1]^s$ be the first M terms of a (t, s)-sequence. For the isotropic discrepancy of P_M we have*

$$J_M(P_M) \leq 2s \left(\frac{4s}{s-1}\right)^{(s-1)/s} \left(\frac{b^{t/s}\left(\frac{b^s(b-1)}{(b+1)2^{s+1}(s)!(\log b)^s}\right)^{1/s} \log M}{M^{1/s}}\right.$$
$$\left. + \frac{C_s' b^{t/s}(\log M)^{(s-1)/s}}{M^{1/s}}\right),$$

for some constant $C_s' > 0$ depending only on s.

Hence, for the star discrepancy of $R_{1,\ell}$ for $1 \leq \ell < k$, using a (t, s)-sequence as a diver sequence in the DAR algorithm we have a convergence rate of order $N_{1,\ell}^{-1/s} \log N_{1,\ell}$. We omit a detailed proof since similar arguments as for proving Theorem 1 can be used. The following corollary holds by substituting the proper upper bounds and $N_{1,\ell}, N_{2,\ell}$ in terms of N.

Corollary 1. *Suppose that the target density $\psi(\boldsymbol{x}) = \sum_{\ell=1}^{k} H_\ell(\boldsymbol{x})$, for $\boldsymbol{x} \in D \subset \mathbb{R}^{s-1}$ satisfies all assumptions stated in Theorem 5 and ψ and H_ℓ for $1 \leq \ell \leq k$ satisfy the conditions in Theorem 1 if $D = [0, 1]^{s-1}$ or Theorem 3 if $D = \mathbb{R}^{s-1}$. Assume that H_ℓ is a product density on $\mathcal{S}_\ell$ and $\mathcal{L}_\ell$ for $1 \leq \ell \leq k$. Let $R_N^{(s-1)}$ be the sample set generated by Algorithm 5. Then we have*

$$D_{N,\psi}^*(R_N^{(s-1)}) \leq \sum_{j=1}^{k-1} \frac{C_{\mathcal{S}_\ell, \psi_{k-\ell+1}} \alpha_j^{1-1/s} \log(\alpha_\ell N)}{N^{1/s}} + \sum_{\ell=1}^{k} \frac{C_{\mathcal{L}_\ell, H_\ell}(\log \beta_\ell N)^{s-1}}{N} + \frac{1}{N},$$

where

$$\alpha_\ell = \frac{\int_{\mathcal{S}_\ell} \psi_{k-\ell+1}(\boldsymbol{x})d\boldsymbol{x}}{\int_D \psi(\boldsymbol{x})d\boldsymbol{x}} \quad and \quad \beta_\ell = \frac{\int_{\mathcal{L}_\ell} H_\ell(\boldsymbol{x})d\boldsymbol{x}}{\int_D \psi(\boldsymbol{x})d\boldsymbol{x}},$$

and $C_{\mathcal{S}_\ell, \psi_{k-\ell+1}}$ and $C_{\mathcal{L}_\ell, H_\ell}$ are constants associated with $\mathcal{S}_\ell, \psi_{k-\ell+1}$ and $\mathcal{L}_\ell, H_\ell$ respectively.

6. Conclusion and outlook

As is well known, the integration error using a Monte Carlo method converges at the rate of $N^{-1/2}$. The acceptance-rejection sampler with a deterministic driver sequence, which is a simple class of MCQMC methods, performs much better in our numerical experiments than the theoretical result $N^{-1/s}$ of Theorem 1 or Theorem 3, for a density defined on $[0, 1]^{s-1}$ or $\mathbb{R}^{s-1}$, would imply. The three

examples even demonstrate that it is possible to achieve a better convergence rate than with standard Monte Carlo using a well-chosen deterministic driver sequence.

The main drawback of the acceptance-rejection sampler is that it might reject many points in high dimension. Some methods to improve the acceptance rate of points are included in [19]. For a special class of density functions given by a finite sum, we propose an embedding deterministic reduced acceptance-rejection algorithm. This algorithm produces a better star discrepancy convergence rate in our numerical example.

Appendix A: δ-cover to approximate star discrepancy

Since it is computationally too expensive to compute the supremum in the definition of the star-discrepancy exactly for dimensions larger than one, we use a so-called δ-cover to estimate this supremum.

Definition 10. Let $(G, \mathcal{B}(G), \psi)$ be a probability space where $G \subseteq \mathbb{R}^{s-1}$ and $\mathcal{B}(G)$ is the Borrel σ-algebra defined on G. Let $\mathscr{A} \subseteq \mathcal{B}(G)$ be a set of test sets. A finite subset $\Gamma_\delta \subseteq \mathscr{A}$ is called a δ-cover of $\mathscr{A}$ with respect to ψ if for every $A \in \mathscr{A}$ there are sets $U, V \in \Gamma_\delta$ such that

$$U \subseteq A \subseteq V$$

and

$$\psi(V \setminus U) \leq \delta.$$

The concept of δ-cover is motivated by the following result [15]. Assume that Γ_δ is a δ-cover of $\mathscr{A}$ with respect to the distribution ψ. For all $\{z_0, z_1, \ldots, z_{N-1}\}$, the following discrepancy inequality holds

$$\sup_{A \in \mathscr{A}} \left| \frac{1}{N} \sum_{n=0}^{N-1} 1_{z_n \in A} - \psi(A) \right| \leq \max_{U \in \Gamma_\delta} \left| \frac{1}{N} \sum_{n=0}^{N-1} 1_{z_n \in U} - \psi(U) \right| + \delta.$$

In the experiments we choose $\mathscr{A}$ to be the set of intervals $[\mathbf{0}, \mathbf{t})$, where $\mathbf{t}$ runs through all points in the domain. For densities defined in $[0,1]^{s-1}$, we set $\Gamma_\delta = \{\prod_{j=1}^{s-1}[0, a_j 2^{-m}) : a_j \in \mathbb{Z}, 0 \leq a_j \leq 2^m\}$, which means that the δ-cover becomes finer as the number of samples increases, thus it can yield a more accurate approximation of the star discrepancy. For densities defined in $\mathbb{R}^{s-1}$, we choose δ-covers with respect to m as $\Gamma_\delta = \{\prod_{j=1}^{s-1}[0, F_j^{-1}(a_j 2^{-m})) : a_j \in \mathbb{Z}, 0 \leq a_j \leq 2^m\}$, where F_j^{-1} is the inverse marginal CDF with respect to the proposal density H. Note that the approximation of the star-discrepancy is computationally expensive, thus our experiments only go up to several thousand sample points. However, the generation of samples using a (t, m, s)-net is fast.

Appendix B: Proofs

Before giving the proofs, we need some preparation.

Consider the following elementary intervals

$$W_k = \prod_{j=1}^{s} \left[\frac{c_j}{b^k}, \frac{c_j + 1}{b^k} \right), \tag{4}$$

with $0 \leq c_j < b^k$ (where c_j is an integer) for $j = 1, \ldots, s$. The diagonal of W_k has length $\sqrt{s}/b^k$ and the volume is b^{-sk}. Let J be an arbitrary convex set in $[0,1]^s$. Let W_k^o denote the union of cubes W_k fully contained in J,

$$W_k^o = \bigcup_{W_k \subseteq J} W_k. \tag{5}$$

Let $\overline{W}_k$ denote the union of cubes W_k having non-empty intersection with J or its boundary $\partial(J)$,

$$\overline{W}_k = \bigcup_{W_k \cap (J \cup \partial(J)) \neq \emptyset} W_k. \tag{6}$$

Lemma 7. *Let $k \in \mathbb{N}$. Let J be an arbitrary convex set in $[0,1]^s$. For the W_k^o and $\overline{W}_k$ constructed by (5) and (6), we have*

$$\lambda(\overline{W}_k \setminus J) \leq 2sb^{-k} \text{ and } \lambda(J \setminus W_k^o) \leq 2sb^{-k}.$$

To illustrate the result we provide the following simple argument which yields a slightly weaker result. Based on the construction of W_k, its diagonal length is $\sqrt{s}/b^k$. Then

$$\overline{W}_k \setminus J \subseteq B := \{\boldsymbol{x} \in [0,1]^s \setminus J : \| \boldsymbol{x} - \boldsymbol{y} \| \leq \sqrt{s} b^{-k} \text{ for some } \boldsymbol{y} \in J\},$$

where $\| \cdot \|$ is the Euclidean norm. Therefore

$$\lambda(\overline{W}_k \setminus J) \leq \lambda(B).$$

Note that the outer surface area of a convex set in $[0,1]^s$ is bounded by the surface area of the unit cube $[0,1]^s$, which is $2s$. Thus the Lebesgue measure of the set B is bounded by the outer surface area times the diameter. Therefore

$$\lambda(\overline{W}_k \setminus J) \leq \lambda(B) \leq 2s\sqrt{s} b^{-k}.$$

The result for $\lambda(J \setminus W_k^o)$ follows by a similar discussion as the proof above.

Remark 3. Note that in [28] it was also shown that the constant $2s$ is best possible.

Now we extend the result in Lemma 7 to pseudo-convex sets.

Corollary 2. *Let J be an arbitrary pseudo-convex set in $[0,1]^s$ with admissible convex covering of p parts with q convex parts of J. For W_k^o and $\overline{W}_k$ given by (5) and (6) we have*

$$\lambda(\overline{W}_k \setminus J) \leq 2psb^{-k} \text{ and } \lambda(J \setminus W_k^o) \leq 2psb^{-k}.$$

Proof. Let $A_1, \ldots, A_p$ be an admissible convex covering of J with p parts. Without loss of generality, let $A_1, \ldots, A_q$ be the convex subsets of J and $A_{q+1}, \ldots, A_p$ be such that $A'_j = A_j \backslash J$ is convex for $q+1 \leq j \leq p$. It can be shown that

$$J = \bigcup_{j=1}^{q} A_j \cup \bigcup_{j=q+1}^{p} (A_j \backslash A'_j). \tag{7}$$

Therefore

$$\overline{W}_k \backslash J \subseteq \left(\bigcup_{j=1}^{q} B_j \cup \bigcup_{j=q+1}^{p} B'_j \right),$$

where

$$B_j = \left\{ \boldsymbol{y} \in [0,1]^s \backslash A_j : \| \boldsymbol{x} - \boldsymbol{y} \| \leq \sqrt{s}b^{-k} \text{ for some } \boldsymbol{x} \in A_j \right\}, \; j = 1, \ldots, q,$$

and

$$B'_j = \{ \boldsymbol{y} \in A'_j : \| \boldsymbol{x} - \boldsymbol{y} \| \leq \sqrt{s}b^{-k} \text{ for some } \boldsymbol{x} \in [0,1]^s \backslash A'_j \}, \; j = q+1, \ldots, p.$$

Since $B_j \cup A_j$ for $j = 1, \ldots, q$, and $B'_j \cup A'_j$ for $j = q+1, \ldots, p$ are convex, using Lemma 7, we obtain

$$\lambda(\overline{W}_k \backslash J) \leq \lambda \left(\bigcup_{j=1}^{q} B_j \right) + \lambda \left(\bigcup_{j=q+1}^{p} B'_j \right) \leq 2psb^{-k}.$$

The result for $\lambda(J \backslash W_k^o)$ follows by a similar discussion. $\qquad\square$

B.1. Proof of upper bound

Proof of Lemma 1. For the point set $Q_{m,s} = \{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{M-1}\} \subseteq [0,1]^s$ generated by a (t, m, s)-net in base b with $M = b^m$, let $k = \lfloor \frac{m-t}{s} \rfloor$. Let J be an arbitrary convex set in $[0,1]^s$. Consider the elementary interval W_k given by Equation (4). For W_k^o and $\overline{W}_k$ given by (5) and (6), obviously, $W_k^o \subseteq J \subseteq \overline{W}_k$. The sets W_k^o and $\overline{W}_k$ are fair with respect to the net, that is

$$\frac{1}{M} \sum_{n=0}^{M-1} 1_{\overline{W}_k}(\boldsymbol{x}_n) = \lambda(\overline{W}_k) \quad \text{and} \quad \frac{1}{M} \sum_{n=0}^{M-1} 1_{W_k^o}(\boldsymbol{x}_n) = \lambda(W_k^o).$$

Then

$$\frac{1}{M} \sum_{n=0}^{M-1} 1_J(\boldsymbol{x}_n) - \lambda(J) \leq \frac{1}{M} \sum_{n=0}^{M-1} 1_{\overline{W}_k}(\boldsymbol{x}_n) - \lambda(\overline{W}_k) + \lambda(\overline{W}_k \backslash J)$$

$$= \lambda(\overline{W}_k \backslash J),$$

and

$$\frac{1}{M} \sum_{n=0}^{M-1} 1_J(\boldsymbol{x}_n) - \lambda(J) \geq \frac{1}{M} \sum_{n=0}^{M-1} 1_{W_k^o}(\boldsymbol{x}_n) - \lambda(W_k^o) - \lambda(J \setminus W_k^o)$$
$$= -\lambda(J \setminus W_k^o).$$

By Lemma 7, we have

$$\lambda(\overline{W}_k \setminus J) \leq 2sb^{-k} \text{ and } \lambda(J \setminus W_k^o) \leq 2sb^{-k}.$$

Thus we obtain

$$\left| \frac{1}{M} \sum_{n=0}^{M-1} 1_J(\boldsymbol{x}_n) - \lambda(J) \right| \leq 2sb^{-k} \leq 2sb^{t/s} M^{-1/s}.$$

Since the bound holds for arbitrary convex sets, the proof is complete. $\qquad\square$

Proof of Theorem 1. Let $J_{\boldsymbol{t}}^* = ([\boldsymbol{0}, \boldsymbol{t}) \times [0,1]) \bigcap A$, where $\boldsymbol{t} = (t_1, \ldots, t_{s-1})$ and $A = \{\boldsymbol{x} \in [0,1]^s : \psi(x_1, \ldots, x_{s-1}) \geq Lx_s\}$. Since $\boldsymbol{y}_n$ are the first $s-1$ coordinates of $\boldsymbol{z}_n \in A$ for $n = 0, \ldots, N-1$, we have

$$\sum_{n=0}^{M-1} 1_{J_{\boldsymbol{t}}^*}(\boldsymbol{x}_n) = \sum_{n=0}^{N-1} 1_{J_{\boldsymbol{t}}^*}(\boldsymbol{z}_n) = \sum_{n=0}^{N-1} 1_{[\boldsymbol{0}, \boldsymbol{t})}(\boldsymbol{y}_n).$$

Therefore

$$\left| \frac{1}{N} \sum_{n=0}^{N-1} 1_{[\boldsymbol{0}, \boldsymbol{t})}(\boldsymbol{y}_n) - \frac{1}{C} \int_{[\boldsymbol{0}, \boldsymbol{t})} \psi(\boldsymbol{z}) d\boldsymbol{z} \right| = \left| \frac{1}{N} \sum_{n=0}^{M-1} 1_{J_{\boldsymbol{t}}^*}(\boldsymbol{x}_n) - \frac{1}{\lambda(A)} \lambda(J_{\boldsymbol{t}}^*) \right|.$$

The right-hand side above is now bounded by

$$\frac{M}{N} \left| \frac{1}{M} \sum_{n=0}^{M-1} 1_{J_{\boldsymbol{t}}^*}(\boldsymbol{x}_n) - \lambda(J_{\boldsymbol{t}}^*) \right| + \left| \lambda(J_{\boldsymbol{t}}^*) \left(\frac{M}{N} - \frac{1}{\lambda(A)} \right) \right|$$
$$\leq \frac{M}{N} \left(\left| \frac{1}{M} \sum_{n=0}^{M-1} 1_{J_{\boldsymbol{t}}^*}(\boldsymbol{x}_n) - \lambda(J_{\boldsymbol{t}}^*) \right| + \left| \lambda(A) - \frac{1}{M} \sum_{n=0}^{M-1} 1_A(\boldsymbol{x}_n) \right| \right),$$

where we used the estimation $\lambda(J_{\boldsymbol{t}}^*) \leq \lambda(A)$ and the fact that $N = \sum_{n=0}^{M-1} 1_A(\boldsymbol{x}_n)$. Since $J_{\boldsymbol{t}}^*$ is also pseudo-convex, it follows from Lemma 3 that we can bound the above expression by

$$\frac{M}{N} 2(2p - q) J_M(P_M^{(s)}).$$

In addition, $\lim_{M \to \infty} \frac{N}{M} = \lambda(A)$, which means $\lim_{M \to \infty} \frac{N}{M} = \int_{[0,1]^{s-1}} \psi(\boldsymbol{z}) d\boldsymbol{z} / L = C/L$. Hence there is an M_0 such that $\frac{N}{M} \geq C/(2L)$ for all $M \geq M_0$. Thus $\frac{M}{N} \leq \frac{2L}{C}$ for all $M \geq M_0$. Further we have $N \leq M$. Using Lemma 1 we obtain the bound

$$\frac{M}{N} 2(2p - q) J_M(P_M^{(s)}) \leq 8LC^{-1} sb^{t/s} (2p - q) N^{-1/s}. \qquad\square$$

B.2. Proof of lower bound

Let
$$\mathbb{S}^{s-1} = \left\{ \boldsymbol{x} \in \mathbb{R}^s : \left\| \boldsymbol{x} - (\tfrac{1}{2}, \ldots, \tfrac{1}{2}, 1 - s) \right\| = s \right\},$$

where $\| \cdot \|$ denotes the Euclidian distance. Let
$$\Upsilon^{s-1} = [0, 1]^s \cap \mathbb{S}^{s-1}.$$

This set is the part of a sphere of radius s and centre $(1/2, \ldots, 1/2, 1 - s)$ which lies in the cube $[0, 1]^s$. The radius has been chosen such that Υ^{s-1} is the graph of a concave function.

The (closed) spherical cap $C(\boldsymbol{y}, \rho) \subseteq \mathbb{S}^{s-1}$ with center $\boldsymbol{y} \in \mathbb{S}^{s-1}$ and angular radius $\rho \in [0, \pi]$ is defined by

$$C(\boldsymbol{y}, \rho) = \left\{ \boldsymbol{y} \in \mathbb{S}^{s-1} | \boldsymbol{x} \cdot \boldsymbol{y} \geq s^2 \cos \rho \right\}.$$

The packing of Υ^{s-1} considered here is constructed by identical spherical caps which are non-overlapping, that is, $C(\boldsymbol{y}_i, \rho), C(\boldsymbol{y}_j, \rho) \subseteq \mathbb{S}^{s-1}$ with $i \neq j$ touch at most at their boundaries, and such that the $C(\boldsymbol{y}_i, \rho_n)$ are fully contained in Υ^{s-1}.

Lemma 8. *Let $s \geq 1$. For any $n \in \mathbb{N}$ there exist M_n points $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{M_n}$ on $\Upsilon^{s-1} \subset [0, 1]^s$ and an angular radius ρ_n, with*

$$\rho_n = c_1 (2n)^{-1/(s-1)},$$
$$n \leq M_n \leq c_2 n,$$

such that the caps $C(\boldsymbol{y}_i, \rho_n)$, $i = 1, \ldots, M_n$, form a packing of Υ^{s-1}. The positive constants c_1 and c_2 depend only on the dimension s.

The lemma is essentially well-known for spheres. The explicit proof is due to Winer [43] and Hesse gives a summary in [18, Lemma 1]. A similar argument can be used for our case.

Now we give the proof of Theorem 2 whose proof follows the argument from the proof of [36, Theorem 1].

Proof. We may suppose $s \geq 2$. Let B be the intersection of the unit cube $[0, 1]^s$ with the ball of radius s and centre $(1/2, \ldots, 1/2, 1 - s)$, that is,

$$B = [0, 1]^s \cap \left\{ \boldsymbol{x} \in \mathbb{R}^s : \left\| \boldsymbol{x} - (\tfrac{1}{2}, \ldots, \tfrac{1}{2}, 1 - s) \right\| \leq s \right\}.$$

Let C be a closed spherical cap on $\mathbb{S}^{s-1}$ with spherical radius ρ. The convex hull $\overline{C}$ of C is a solid spherical cap. For $0 < \rho < \pi/2$, the normalized Lebesgue surface measure $\lambda(\overline{C})$ is a continuous function of ρ with

$$c_1 \rho^{s+1} < \lambda(\overline{C}) < c_2 \rho^{s+1}. \tag{8}$$

If N is sufficiently large, there is a positive real number ρ_0 such that a cap C of spherical radius ρ_0 has measure

$$\lambda(\overline{C}) = \frac{1}{2N}.$$

In view of (8), we have $0 < \rho_0 < c_3 N^{-1/(s+1)}$. We now pick as many pairwise disjoint caps, which are fully contained in Υ^{s-1} and which have radius ρ_0, as possible, say $C_1, \ldots, C_M$. By Lemma 8, for large N and hence small ρ_0 we have $M \geq c_4 \rho_0^{-(s-1)}$, hence

$$M \geq c_5 N^{(s-1)/(s+1)}. \tag{9}$$

Given a sequence of numbers $\sigma_1, \ldots, \sigma_M$, with each σ_i either 1 or -1, let $B(\sigma_1, \ldots, \sigma_M)$ consist of all $\boldsymbol{x} \in B$ which do not lie in a cap $\overline{C_i}$ with $\sigma_i = -1$. In other words, $B(\sigma_1, \ldots, \sigma_M)$ is obtained from B by removing the solid caps $\overline{C_i}$ for which $\sigma_i = -1$.

Now the local discrepancy function $\Delta_{P_N}(H)$ defined by $\Delta_{P_N}(H) = \sum_{i=1}^{N} 1_H(P_N) - N\lambda(H)$ is additive, i.e. it satisfies

$$\Delta_{P_N}(H \cup H') = \Delta_{P_N}(H) + \Delta_{P_N}(H')$$

if $H \cap H' = \emptyset$. It follows easily that

$$\Delta_{P_N}(B(\sigma_1, \ldots, \sigma_M)) - \Delta_{P_N}(B(-\sigma_1, \ldots, -\sigma_M)) = \sum_{i=1}^{M} \sigma_i \Delta_{P_N}(\overline{C_i}).$$

We have

$$\Delta_{P_N}(\overline{C_i}) = \sum_{i=1}^{N} 1_{\overline{C_i}}(P_N) - N\lambda(\overline{C_i}) = \sum_{i=1}^{N} 1_{\overline{C_i}}(P_N) - \frac{1}{2}.$$

Hence for every i, either $\Delta_{P_N}(\overline{C_i}) \geq \frac{1}{2}$ or $\Delta_{P_N}(\overline{C_i}) = -\frac{1}{2}$. Choose σ_i such that $\sigma_i \Delta_{P_N}(\overline{C_i}) \geq 1/2$ for $1 \leq i \leq M$. Then

$$\Delta_{P_N}(B(\sigma_1, \ldots, \sigma_M)) - \Delta_{P_N}(B(-\sigma_1, \ldots, -\sigma_M)) \geq M/2,$$

and either $J = B(\sigma_1, \ldots, \sigma_M)$ or $J = B(-\sigma_1, \ldots, -\sigma_M)$ has $|\Delta_{P_N}(J)| \geq M/4$. In addition, J is a convex set due to its construction.

Thus by (9),

$$D_N^*(\lambda, J) \geq \frac{1}{4} \frac{M}{N} \geq c_6 N^{-2/(s+1)}.$$

We take ψ as the boundary of J excluding the boundary of $[0,1]^s$, which completes the proof. $\qquad\square$

B.3. Proof of Theorem 5

Proof. In what follows we restrict our investigations to the case $k = 2$ for simplicity, the general case can be proved by similar arguments. Let $\psi = H_1 + H_2$ be the target density function. Assume that we can apply the inverse CDF on H_1 and H_2 to generate samples. Let

$$\mathcal{S} := \{\boldsymbol{x} \in D : \psi(\boldsymbol{x}) < H_1(\boldsymbol{x})\},$$

and

$$\mathcal{L} := \{\boldsymbol{x} \in D : \psi(\boldsymbol{x}) \geq H_1(\boldsymbol{x})\}.$$

The final sample set $R_N^{(1)}$ is a superposition of the three subsets, $R_{1,1}, R_{1,2}$ and $R_{2,2}$, see Figure 4. Define $R_{i,\ell} = \{x_0^{(i,\ell)}, x_1^{(i,\ell)}, \ldots, x_{N_{i,\ell}-1}^{(i,\ell)}\}$ for $i, \ell = 1, 2$. The number $N_{i,\ell}$ of the points in each subset is given by

$$N_{1,1} = \left\lceil N \frac{\int_{\mathcal{S}} \psi(\boldsymbol{x})d\boldsymbol{x}}{\int_D \psi(\boldsymbol{x})d\boldsymbol{x}} \right\rceil, N_{1,2} = \left\lceil N \frac{\int_{\mathcal{L}} H_1(\boldsymbol{x})d\boldsymbol{x}}{\int_D \psi(\boldsymbol{x})d\boldsymbol{x}} \right\rceil \text{ and } N_{2,2} = \left\lceil N \frac{\int_{\mathcal{L}} H_2(\boldsymbol{x})d\boldsymbol{x}}{\int_D \psi(\boldsymbol{x})d\boldsymbol{x}} \right\rceil.$$

Then there exists $\delta_i \in [0, 1)$ for $i = 1, 2, 3$ such that

$$N_{1,1} = N \frac{\int_{\mathcal{S}} \psi(\boldsymbol{x})d\boldsymbol{x}}{\int_D \psi(\boldsymbol{x})d\boldsymbol{x}} + \delta_1, N_{1,2} = N \frac{\int_{\mathcal{L}} H_1(\boldsymbol{x})d\boldsymbol{x}}{\int_D \psi(\boldsymbol{x})d\boldsymbol{x}} + \delta_2 \text{ and } N_{2,2} = N \frac{\int_{\mathcal{L}} H_2(\boldsymbol{x})d\boldsymbol{x}}{\int_D \psi(\boldsymbol{x})d\boldsymbol{x}} + \delta_3.$$

Therefore,

$$\left| \frac{1}{N} \sum_{n=0}^{N-1} 1_{(-\infty,t] \cap D}(\boldsymbol{x}_n) - \frac{\int_{(-\infty,t] \cap D} \psi(\boldsymbol{x})d\boldsymbol{x}}{\int_D \psi(\boldsymbol{x})d\boldsymbol{x}} \right|$$

$$= \left| \frac{1}{N} \sum_{n=0}^{N-1} 1_{(-\infty,t] \cap D}(\boldsymbol{x}_n) - \frac{\int_{(-\infty,t] \cap \mathcal{L}} H_1(\boldsymbol{x})d\boldsymbol{x} + \int_{(-\infty,t] \cap \mathcal{L}} H_2(\boldsymbol{x})d\boldsymbol{x} + \int_{(-\infty,t] \cap \mathcal{S}} \psi(\boldsymbol{x})d\boldsymbol{x}}{\int_D \psi(\boldsymbol{x})d\boldsymbol{x}} \right|$$

$$\leq \left| \frac{1}{N} \sum_{n=0}^{N_{1,1}-1} 1_{(-\infty,t] \cap D}(\boldsymbol{x}_n^{(1,1)}) - \frac{\int_{(-\infty,t] \cap \mathcal{L}} H_1(\boldsymbol{x})d\boldsymbol{x}}{\int_D \psi(\boldsymbol{x})d\boldsymbol{x}} \right|$$

$$+ \left| \frac{1}{N} \sum_{n=0}^{N_{1,2}-1} 1_{(-\infty,t] \cap D}(\boldsymbol{x}_n^{(1,2)}) - \frac{\int_{(-\infty,t] \cap \mathcal{L}} H_2(\boldsymbol{x})d\boldsymbol{x}}{\int_D \psi(\boldsymbol{x})d\boldsymbol{x}} \right|$$

$$+ \left| \frac{1}{N} \sum_{n=0}^{N_{2,2}-1} 1_{(-\infty,t] \cap D}(\boldsymbol{x}_n^{(2,2)}) - \frac{\int_{(-\infty,t] \cap \mathcal{S}} \psi(\boldsymbol{x})d\boldsymbol{x}}{\int_D \psi(\boldsymbol{x})d\boldsymbol{x}} \right|$$

$$= \frac{N_{1,1}}{N} \left| \frac{1}{N_{1,1}} \sum_{n=0}^{N_{1,1}-1} 1_{(-\infty,t] \cap D}(\boldsymbol{x}_n^{(1,1)}) - \frac{N}{N_{1,1}} \frac{\int_{(-\infty,t] \cap \mathcal{S}} \psi(\boldsymbol{x})d\boldsymbol{x}}{\int_D \psi(\boldsymbol{x})d\boldsymbol{x}} \right|$$

$$+ \frac{N_{1,2}}{N} \left| \frac{1}{N_{1,2}} \sum_{n=0}^{N_{1,2}-1} 1_{(-\infty,t] \cap D}(\boldsymbol{x}_n^{(1,2)}) - \frac{N}{N_{1,2}} \frac{\int_{(-\infty,t] \cap \mathcal{L}} H_1(\boldsymbol{x})d\boldsymbol{x}}{\int_D \psi(\boldsymbol{x})d\boldsymbol{x}} \right|$$

$$+ \frac{N_{2,2}}{N} \left| \frac{1}{N_{2,2}} \sum_{n=0}^{N_{2,2}-1} 1_{(-\infty,t] \cap D}(\boldsymbol{x}_n^{(2,2)}) - \frac{N}{N_{2,2}} \frac{\int_{(-\infty,t] \cap \mathcal{L}} H_2(\boldsymbol{x})d\boldsymbol{x}}{\int_D \psi(\boldsymbol{x})d\boldsymbol{x}} \right|$$

$$= \frac{N_{1,1}}{N} \left[\left| \frac{1}{N_{1,1}} \sum_{n=0}^{N_{1,1}-1} 1_{(-\infty,\boldsymbol{t}] \cap D}(\boldsymbol{x}_n^{(1,1)}) - \frac{\int_{(-\infty,\boldsymbol{t}] \cap \mathcal{S}} \psi(\boldsymbol{x})d\boldsymbol{x}}{\int_{\mathcal{S}} \psi(\boldsymbol{x})d\boldsymbol{x}} \right| + \frac{\delta_1}{N_{1,1}} \frac{\int_{(-\infty,\boldsymbol{t}] \cap \mathcal{S}} \psi(\boldsymbol{x})d\boldsymbol{x}}{\int_{\mathcal{S}} \psi(\boldsymbol{x})d\boldsymbol{x}} \right]$$

$$+ \frac{N_{1,2}}{N} \left[\left| \frac{1}{N_{1,2}} \sum_{n=0}^{N_{1,2}-1} 1_{(-\infty,\boldsymbol{t}] \cap D}(\boldsymbol{x}_n^{(1,2)}) - \frac{\int_{(-\infty,\boldsymbol{t}] \cap \mathcal{L}} H_1(\boldsymbol{x})d\boldsymbol{x}}{\int_{\mathcal{L}} H_1(\boldsymbol{x})d\boldsymbol{x}} \right| + \frac{\delta_2}{N_{1,2}} \frac{\int_{(-\infty,\boldsymbol{t}] \cap \mathcal{L}} H_1(\boldsymbol{x})d\boldsymbol{x}}{\int_{\mathcal{L}} H_1(\boldsymbol{x})d\boldsymbol{x}} \right]$$

$$+ \frac{N_{2,2}}{N} \left[\left| \frac{1}{N_{2,2}} \sum_{n=0}^{N_{2,2}-1} 1_{(-\infty,\boldsymbol{t}] \cap D}(\boldsymbol{x}_n^{(2,2)}) - \frac{\int_{(-\infty,\boldsymbol{t}] \cap \mathcal{L}} H_2(\boldsymbol{x})d\boldsymbol{x}}{\int_{\mathcal{L}} H_2(\boldsymbol{x})d\boldsymbol{x}} \right| + \frac{\delta_3}{N_{2,2}} \frac{\int_{(-\infty,\boldsymbol{t}] \cap \mathcal{L}} H_2(\boldsymbol{x})d\boldsymbol{x}}{\int_{\mathcal{L}} H_2(\boldsymbol{x})d\boldsymbol{x}} \right]$$

$$\leq \frac{N_{1,1}}{N} D_{\mathcal{S},\psi}^* + \frac{N_{1,2}}{N} D_{\mathcal{L},H_1}^* + \frac{N_{2,2}}{N} D_{\mathcal{L},H_2}^* + \frac{1}{N} \frac{\int_{(-\infty,\boldsymbol{t}] \cap D} \psi(\boldsymbol{x})d\boldsymbol{x}}{\int_D \psi(\boldsymbol{x})d\boldsymbol{x}}$$

$$\leq \frac{N_{1,1}}{N} D_{\mathcal{S},\psi}^* + \frac{N_{1,2}}{N} D_{\mathcal{L},H_1}^* + \frac{N_{2,2}}{N} D_{\mathcal{L},H_2}^* + \frac{1}{N},$$

where $D_{\mathcal{S},\psi}^*$ is the star discrepancy of sample points in $\mathcal{S}$ associated with ψ and the same notation is also applied to $D_{\mathcal{L},H_1}^*$ and $D_{\mathcal{L},H_2}^*$. Since this result holds for arbitrary $\boldsymbol{t}$, the desired result follows then immediately. □

References

[1] AISTLEITNER, C., BRAUCHART, J.S. and DICK, J., *Point sets on the sphere* $\mathbb{S}^2$ *with small spherical cap discrepancy.* Discrete and Computational Geometry, 48, 990–1024, 2012. MR3000572

[2] BAREKAT, F. and CAFLISCH, R., *Simulation with Fluctuation and Singular Rates.* ArXiv:1310.4555 [math.NA], 2013.

[3] BECK, J., *On the discrepancy of convex plane sets.* Monatshefte Mathematik, 105, 91–106, 1988. MR0930428

[4] BOTTS, C., HÖRMANN, W. and LEYDOLD, J., *Transformed density rejection with inflection points.* Statistics and Computing, 23, 251–260, 2013. MR3016942

[5] CAFLISCH, R.E., *Monte Carlo and quasi-Monte Carlo methods.* Acta Numerica, 1–49, 1998. MR1689431

[6] CHEN, S., *Consistency and convergence rate of Markov chain quasi Monte Carlo with examples.* PhD thesis, Stanford University, 2011.

[7] CHEN, S., DICK, J. and OWEN, A.B., *Consistency of Markov chain quasi-Monte Carlo on continuous state spaces.* Annals of Statistics, 39, 673–701, 2011. MR2816335

[8] CHENTSOV, N.N., *Pseudorandom numbers for modelling Markov chains.* Computational Mathematics and Mathematical Physics, 7, 218–233, 1967.

[9] DEVROYE, L., *Nonuniform Random Variate Generation.* Springer-Verlag, New York, 1986. MR0836973

[10] DICK, J., KUO, F. and SLOAN, I.H., *High dimensional integration-the quasi-Monte Carlo way.* Acta Numerica, 22, 133–288, 2013. MR3038697

[11] DICK, J. and PILLICHSHAMMER, F., *Digital Nets and Sequences. Discrepancy Theory and Quasi-Monte Carlo Integration.* Cambridge University Press, Cambridge, 2010. MR2683394

[12] DICK, J., RUDOLF, D. and ZHU, H., *Discrepancy bounds for*

uniformly ergodic Markov chain quasi-Monte Carlo. Available at arxiv.org/abs/1303.2423 [stat.CO], submitted, 2013.

[13] FAURE, H., *Discrépance de suites associées à un système de numération (en dimension s)*. Acta Arithmetica, 4, 337–351, 1982. MR0677547

[14] GERBER, M. and CHOPIN, N., *Sequential quasi-Monte Carlo*. Available at arXiv:1402.4039 [stat.CO], 2014.

[15] GNEWUCH, M., *Bracketing numbers for axis-parallel boxes and application to geometric discrepancy*. Journal of Complexity, 24, 154–172, 2008. MR2400314

[16] KUIPERS, L. and NIEDERREITER, H., *Uniform Distribution of Sequences*. John Wiley, New York, 2006. MR0419394

[17] HALTON, J.H., *On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals*. Numerische Mathematik, 2, 84–90, 1960. MR0121961

[18] HESSE, K., *A lower bound for the worst-case cubature error on spheres of arbitrary dimension*. Numerische Mathematik, 103, 413–433, 2006. MR2221056

[19] HÖRMANN, W., LEYDOLD, J. and DERFLINGER, G., *Automatic Nonuniform Random Variate Generation*. Statistics and Computing. Springer-Verlag, Berlin, 2004. MR2143197

[20] KRITZER, P., *Improved upper bounds on the star discrepancy of (t, m, s)-nets and (t, s)-sequences*. Journal of Complexity, 22, 336–347, 2006. MR2229897

[21] L'ECUYER, P., LECOT, C. and TUFFIN, B., *A randomized quasi-Monte Carlo simulation method for Markov chains*. Operation Research, 56, 958–975, 2008.

[22] MEYN, M.P. and TWEEDIE, T.L., *Markov Chain and Stochastic Stability*. Springer-Verlag, London, 1993. MR1287609

[23] MOROKOFF, W.J. and CAFLISCH, R.E., *Quasi-Monte Carlo integration*. Journal of Computational Physics, 122, 218–230, 1995. MR1365433

[24] MOSKOWITZ, B. and CAFLISCH, R.E., *Smoothness and dimension reduction in quasi-Monte Carlo methods*. Mathematical and Computer Modelling, 23, 37–54, 1996. MR1398000

[25] NGUYEN, N. and ÖKTEN, G., *The acceptance-rejection method for low discrepancy sequences*. Submitted, 2014.

[26] NIEDERREITER, H., *Low-discrepancy and low-dispersion sequences*. Journal of Number Theory, 30, 51–70, 1988. MR0960233

[27] NIEDERREITER, H., *Random Number Generation and Quasi-Monte Carlo Methods*. SIAM, Philadelphia, Pennsylvania, 1992. MR1172997

[28] NIEDERREITER, H. and WILLS, J.M., *Diskrepanz und Distanz von Maßen bezüglich konvexer und Jordanscher Mengen* (German). Mathematische Zeitschrift, 144, 125–134, 1975. MR0376588

[29] NOH, Y., CHOI, K.K. and DU, L., *New Transformation of Dependent Input Variables Using Copula for RBDO*. 7th World Congresses of Structural and Multidisciplinary Optimization COEX Seoul, Korea, 21–25 May 2007.

[30] OWEN, A.B., *Randomly permuted (t, m, s)-nets and (t, s)-sequences.* In: H. Niederreiter and P.J. Shiue (Eds.). Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing (Las Vegas, NV, 1994), 299–317. Lecture Notes in Statist., 106. Springer, New York, 1995. MR1445791

[31] OWEN, A.B., *Monte Carlo variance of scrambled net quadrature.* SIAM Journal on Numerical Analysis, 34, 1884–1910, 1997. MR1472202

[32] OWEN, A.B., *Scrambled net variance for integrals of smooth functions.* Annals of Statistics, 25, 1541–1562, 1997. MR1463564

[33] OWEN, A.B., *Monte Carlo theory, methods and examples (book draft).* Available at http://www-stat.stanford.edu/~owen/mc/. Last accessed on 16 May 2014.

[34] ROBERT, C. and CASELLA, G., *Monte Carlo Statistical Methods.* Springer-Verlag, New York, second edition, 2004. MR2080278

[35] ROSENBLATT, M., *Remarks on a multivariate transformation.* Annals of Mathematical Statistics, 23, 470–472, 1952. MR0049525

[36] SCHMIDT, W.M., *Irregularities of distribution.* Acta Arithmetica, 27, 385–396, 1975. MR0376593

[37] SOBOL, I.M., *Distribution of points in a cube and approximate evaluation of integrals.* Akademija Nauk SSSR. Žurnal Vyčislitel' noǐ Matematiki i Matematičeskoǐ Fiziki 7, 784–802, 1967 (in Russian); U.S.S.R Computational Mathematics and Mathematical Physics 7, 86–112, 1967 (in English). MR0219238

[38] STUTE, W., *Convergence rates for the isotrope discrepancy.* Annals of Probability, 5, 707–723, 1977. MR0455096

[39] TRIBBLE, S.D., *Markov chain Monte Carlo algorithms using completely uniformly distributed driving sequences.* PhD thesis, Stanford University, 2007. MR2710331

[40] TRIBBLE, S.D. and OWEN, A.B., *Constructions of weakly CUD sequences for MCMC.* Electronic Journal of Statistics, 2, 634–660, 2008. MR2426105

[41] WANG, X., *Quasi-Monte Carlo integration of characteristic functions and the rejection sampling method.* Comupter Physics Communication, 123, 16–26, 1999.

[42] WANG, X., *Improving the rejection sampling method in quasi-Monte Carlo methods.* Journal of Computational and Applied Mathematics, 114, 231–246, 2000. MR1737076

[43] WYNER, A.D., *Capabilities of bounded discrepancy decoding.* The Bell System Technical Journal, 44, 1061–1122, 1965. MR0180417