An Analysis of a Variation of Hit-and-Run for Uniform Sampling from General Regions

SEKSAN KIATSUPAIBUL
Chulalongkorn University
ROBERT L. SMITH
University of Michigan
and
ZELDA B. ZABINSKY
University of Washington

Hit-and-run, a class of MCMC samplers that converges to general multivariate distributions, is known to be unique in its ability to *mix fast* for uniform distributions over convex bodies. In particular, its rate of convergence to a uniform distribution is of a low order polynomial in the dimension. However, when the body of interest is difficult to sample from, typically a hyperrectangle is introduced that encloses the original body, and a one-dimensional acceptance/rejection is performed. The fast mixing analysis of hit-and-run does not account for this one-dimensional sampling that is often needed for implementation of the algorithm. Here we show that the effect of the size of the hyperrectangle on the efficiency of the algorithm is only a linear scaling effect. We also introduce a variation of hit-and-run that accelerates the sampler, and demonstrate its capability through a computational study.

Categories and Subject Descriptors: G.3 [**Probability and Statistics**]: —*Probabilistic algorithm (including Monte Carlo)*

General Terms: Algorithm, Performance

Additional Key Words and Phrases: Markov chain Monte Carlo, Hit-and-run algorithm, uniform sampling from a convex body

1. INTRODUCTION

The hit-and-run algorithm [Smith 1984] is a Markov Chain Monte Carlo (MCMC) method for generating points uniformly distributed on an arbitrary bounded open subset of a finite d-dimensional Euclidean space. The algorithm is relatively simple. Starting with a specific point, say x, in the open set, a point θ on the surface of

Author's address: Seksan Kiatsupaibul, Department of Statistics, Chulalongkorn Business School Chulalongkorn University, Bangkok 10330, Thailand
Robert L. Smith, Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, Michigan 48105, USA
Zelda B. Zabinsky, Department of Industrial and Systems Engineering, University of Washington, Seattle, WA 98195-2650, USA

a d-dimensional unit hypersphere centered at x is chosen at random, defining a line passing through x and θ. Then a new sampled point is chosen uniformly on the intersection of this line and the open set. This new sampled point replaces the starting point x and the process is reiterated. The distribution of the sampled points converges in total variation to the uniform distribution on the open set. This version of hit-and-run is considered to be the most efficient algorithm for generating an asymptotically uniform point if the set under consideration is convex [Lovász 1999; 2006]. With an appropriate filter, the algorithm can also be extended to sample points that converge to an arbitrary target distribution in total variation [Bélisle et al. 1993; Romeijn and Smith 1994].

Simple as it seems, direct sampling from the intersection of the random line and the open set may be difficult, because the one-dimensional set can be the union of an arbitrarily large number of open intervals. One way to circumvent the problem is to enclose the region within a hyperrectangle, or box, so that sampling on the intersection of the line with the box is easy, applying the rejection technique until a point is sampled in the union of open intervals. Performing hit-and-run on an enclosing hyperrectangle in this way was recognized as early as the conception of the hit-and-run itself. However, analysis of this technique's impact on the efficiency of hit-and-run has not been done.

In this article, we first precisely define the algorithm on an enclosing box. We then show that the impact of the box on the complexity is only a linear scaling factor and derive a bound on the actual scale factor for the case of convex bodies. This illustrates the power of hit-and-run since a direct acceptance/rejection sampling entails exponential effort in the dimension. We also describe a variation to accelerate the algorithm, and perform a computational study to compare the efficiencies between the original and the accelerated hit-and-run on an enclosing box.

2. HIT-AND-RUN WITH BOX

Let S be a bounded open subset of $\Re^d$ from which we want to sample a uniform point. We state the hit-and-run algorithm with an enclosing hyperrectangle as its sampling agent in Algorithm 2.1. The shrinking algorithm from Neal [2003] is summarized in Algorithm 2.2, and used to accelerate Algorithm 2.1 into Algorithm 2.3.

Denote the d-dimensional Lebesgue measure over $\Re^d$ by λ_d where λ denotes the one-dimensional Lebesgue measure over a line. Let D denote the d-dimensional unit sphere centered at the origin, and ∂D denote its surface. Let ν be a continuous probability distribution on ∂D with density bounded away from zero. Assume further that we know a d-dimensional hyperrectangle, referred to as box B that contains S. Let R be the diameter of B (e.g. the longest chord). The following algorithm on an enclosing box is a modification of hit-and-run with a general direction distribution, as in Bélisle et al. [1993].

Algorithm 2.1 The Hit-and-run Algorithm on a Box.

Step 0:. Let $X_0 = x_0 \in S$, *and set* $n = 0$.

Step 1:. Choose a direction Θ_n *on* ∂D *with distribution* ν *and set* $i = 1$.

Step 2:. Choose $L_{n,i}$ *from the uniform distribution on* Ξ_n*:*

$$\Xi_n = \{r \in \Re : x_n + r\Theta_n \in B\}.$$

Step 3:. If $X_n + L_{n,i}\Theta_n$ *is not in* S*, set* $i = i+1$ *and return to Step 2. Otherwise, set* $X_{n+1} = X_n + L_{n,i}\Theta_n$*.*

Step 4:. Set $n = n + 1$*. Go to Step 1.*

In words, Ξ_n is the intersection of the line passing through X_n in direction Θ_n and the box B. Let Λ_n be the intersection of the line passing through X_n in direction Θ_n and S, i.e. $\Lambda_n = \Xi_n \cap S$. Performing Step 2 and Step 3 samples a point uniformly on Ξ_n repeatedly until we obtain a point in Λ_n. This is, in fact, the rejection technique to sample a point uniformly on Λ_n. This rejection is the only modification from the original hit-and-run, where a uniform sample on Λ_n is performed directly. Therefore, the resulting sequence of iteration points $(X_n; n \geq 0)$ from the hit-and-run algorithm on a box is the same as the original hit-and-run. Note that when S is described by linear constraints or constraints that are invertible, then sampling on the line set Λ_n can be performed directly.

The original hit-and-run, and hence the hit-and-run algorithm on a box, is uniformly ergodic. It has been shown that the limiting distribution of an iteration point X_n is the uniform distribution on S. This results from the property that the transition probability of the hit-and-run process is reversible with respect to the uniform distribution. The process is also shown to be uniformly ergodic [Bélisle et al. 1998; Diaconis and Freedman 1997].

The complexity of the hit-and-run algorithm on a box accounts for the total number of sampling points from B per iteration and the number of iterations needed to achieve the uniform distribution on S within a certain error. To make the statement precise, we need to introduce some quantities.

For any $n = 0, 1, 2, \ldots$, define C_n as the total number of sampling points in the nth iteration. Since X_n is uniformly ergodic with the uniform limiting distribution, given a fixed error $\epsilon > 0$, there exists a number N_ϵ such that, for any measurable subset A and for all $x_0 \in S$,

$$|\mathbb{P}\left[X_{N_\epsilon} \in A | X_0 = x_0\right] - \mu(A)| < \epsilon,$$

where μ is the uniform probability on S. Therefore, the total number of sampling points required before the distribution of the iteration point attains the uniform distribution on S within an ϵ error is $\sum_{n=0}^{N_\epsilon - 1} C_n$.

With a fixed error $\epsilon > 0$, define the total expected number of sampling points to get within ϵ of the limiting distribution as ET_ϵ

$$ET_\epsilon = \mathbb{E}\left[\sum_{n=0}^{N_\epsilon - 1} C_n\right] = \sum_{n=0}^{N_\epsilon - 1} \mathbb{E}\left[C_n\right]. \quad (1)$$

The second equality in (1) follows because the hit-and-run process is uniformly ergodic, so N_ϵ is a constant that is independent of the starting point $x_0 \in S$. Note that N_ϵ is the measure of complexity used in Lovász and Vempala [2006], where iteration points are assumed to be sampled directly from S. In our setting, we sample S indirectly from the box B, so we need to count not only the iteration points in S but also all the sample points in B. Therefore, N_ϵ is not enough to capture the whole complexity, and we instead use ET_ϵ for our measure of complexity. Now, if the initial distribution of X_0 is the uniform distribution on S, the hit-and-run process is stationary and $\mathbb{E}[C_n] = EC$ is the same for all n. Therefore,

$$EC = \int_S \mathbb{E}[C_n | X_n = x] \left(\frac{1}{\lambda_d(S)} \right) d\lambda_d(x). \tag{2}$$

In some applications of Markov chain sampling, one requires more than one sample from the stationary distribution of the chain, each sample independent of the others. Usually, this is accomplished not by running multiple sequences of the Markov chain from different starting points, but by running a very long sequence of the chain with the burn-in period removed. After the burn-in period, the chain is approximately stationary. EC approximates the expected number of sampling points in one iteration after the burn-in period. In that case, $ET_\epsilon = N_\epsilon EC$ approximates the expected number of sampling points required until two iteration points in the sequence generated by process can be treated as independent. These approximately independent iteration points have been used, for example, in polynomial time volume algorithm for convex bodies [Kannan et al. 1997] and also in constructing a confidence interval for some Bayesian inferences [Quinn 2004]. In simulation, the batch-means approach has been primarily employed in construction of a confidence interval of the long term mean of an output of a Markov chain sequence [Asmussen and Glynn 2007]. For example, the hit-and-run algorithm can be used with the batch-means approach to construct a confidence interval on the center of gravity of S. N_ϵ can be employed as a fixed batch size, and ET_ϵ approximates the expected number of sampling points required in one batch.

In the next section, we show that the size of the box is not a crucial factor of the computational complexity of Algorithm 2.1 by providing bounds on ET_ϵ as a linear function of the box diameter. We also provide an explicit formula for the bound on ET_ϵ when the chain is stationary and the region S is convex. This explicit bound is applicable when the burn-in period is long enough that the chain is approximately stationary.

For the hit-and-run algorithm with an enclosing box, one can speed up the algorithm by employing the following shrinking mechanism due to Neal [2003] to reduce the rejection time in Step 3 of Algorithm 2.1.

Algorithm 2.2 Neal's Shrinking Algorithm [Neal 2003]. *Let $S \subset \Re$ be an open set contained in an interval $I_0 = (b_0^-, b_0^+)$ and let $x \in S$. Set $i = 0$. Choose a new point Y in S as follows.*

Step 1:. Choose a point X' uniformly on I_i.

Step 2:. If X' is not in S, then shrink the interval as follows. If $X' > x$, set $b_{i+1}^+ = X'$. If $X' < x$, set $b_{i+1}^- = X'$. Let $I_{i+1} = (b_{i+1}^-, b_{i+1}^+)$. Set $i = i + 1$ and return to Step 1. Otherwise, if X' is in S, set the new point $Y = X'$.

It is proven in Neal [2003] that, for any x, y in an open set $S \subset \Re$,

$$q(x, y) = q(y, x) \quad \text{where } P(Y \in A \mid \text{starting at } x) = \int_A q(x, y)\, dy. \tag{3}$$

This shrinking algorithm is used to sample from a level set of a one-dimensional density function in the slice sampling technique. Note that Y is not necessarily uniform on S. Combining the shrinking mechanism with hit-and-run, we obtain the following accelerated algorithm as follows.

ALGORITHM 2.3 THE ACCELERATED HIT-AND-RUN ALGORITHM ON A BOX.

Step 0:. *Let $X_0 = x_0 \in S$, and set $n = 0$.*

Step 1:. *Choose a direction Θ_n on ∂D with distribution ν, defining the line Ξ_n;*

$$\Xi_n = \{r \in \Re : x_n + r\Theta_n \in B\}.$$

Set $l_1^+ = \sup\ \Xi_n$ and $l_1^- = \inf\ \Xi_n$. Set $i = 1$.

Step 2:. *Choose $L_{n,i}$ from the uniform distribution on the open interval (l_i^-, l_i^+)*

Step 3:. *Set $X_{n,i} = X_n + L_{n,i}\Theta_n$.*

If $X_{n,i}$ is not in S, set l_{i+1}^+ and l_{i+1}^- as follows. If $L_{n,i} > 0$, set $l_{i+1}^+ = L_{n,i}$ and keep $l_{i+1}^- = l_i^-$. If $L_{n,i} < 0$, set $l_{i+1}^- = L_{n,i}$ and keep $l_{i+1}^+ = l_i^+$. Set $i = i + 1$ and return to Step 2.

Otherwise, if $X_{n,i}$ is in S, set $X_{n+1} = X_{n,i}$.

Step 4:. *Set $n = n + 1$. Go to Step 1.*

The Markov chain generated by Algorithm 2.3 is reversible with respect to the uniform distribution on S, no matter if S is convex or non-convex, because of (i) the probability (density) of choosing a line passing through y starting from x is the same as that of choosing a line passing through x starting from y, for any $x, y \in S$ and (ii) given a line passing through both x and y, the probability (density) of choosing the point y on the line from x is the same as that of choosing the point x on the line from y, as stated in (3).

Algorithm 2.3 differs from Algorithm 2.1 in that the line intersecting the box is shrinking. This shrinkage increases the probability of acceptance in Step 2 and 3. Because every open subset S can still be reached in one step, the convergence property of the new Markov chain remains the same.

When S is convex, the iteration point process generated by Algorithm 2.3 distributes the same as that generated by Algorithm 2.1, so N_ϵ of the two processes are the same. However, when S is not convex, the iteration point processes from the two algorithms distribute differently, and, hence, N_ϵ of the two processes may be different. As a result, in terms of ET_ϵ, Algorithm 2.3 is faster than Algorithm 2.1 when S is convex, but unclear when S is not convex.

In what follows, we analyze the upper bounds of ET_ϵ and EC of Algorithm 2.1. We then perform a computational study to evaluate the benefit of employing Algorithm 2.3 over Algorithm 2.1.

3. ANALYSIS OF THE COMPLEXITY

The complexity of Algorithm 2.1 defined by ET_ϵ relies solely on the quantity $\mathbb{E}[C_n]$, because N_ϵ is determined by the region S and is independent of the box B. Therefore, the main analysis is developing a bound on $\mathbb{E}[C_n]$. We first derive a bound on $\mathbb{E}[C_n]$ in general to show that the bound grows linearly in the diameter of the box. When the process is stationary, then $\mathbb{E}[C_n] = EC$, and we can compute a bound for a specific case when S is the union of a finite number of convex bodies. We are interested in this specific case because the class of convex bodies is the domain of problems on which the original hit-and-run algorithm has been proven practical.

3.1 Bound on $\mathbb{E}[C_n]$

Let S be an open set, and B be a hyperrectangle, with diameter R, containing S. Suppose Algorithm 2.1 starts with $X_0 = x_0$. The algorithm generates two random sequences: the direction vectors $(\Theta_n; n \geq 0)$ and the iteration points $(X_n; n \geq 0)$.

For all $x \in S$ and $\theta \in \partial D$, define

$$S_{x,\theta} = \{s \in S : s = x + t\theta, t \in \Re\} \quad \text{and} \quad B_{x,\theta} = \{s \in B : s = x + t\theta, t \in \Re\}.$$

Proposition 3.1. *For all $n = 0, 1, 2, \ldots,$*

$$\mathbb{E}[C_n] \leq R \cdot \mathbb{E}\left[\frac{1}{\lambda(S_{X_n,\Theta_n})}\right]. \tag{4}$$

Proof. According to Algorithm 2.1, for each $n = 0, 1, 2, \ldots,$ given $X_n = x$, $\Theta_n = \theta$, C_n is a geometric random variable with mean $\lambda(B_{x,\theta})/\lambda(S_{x,\theta})$. Hence,

$$\mathbb{E}[C_n] = \mathbb{E}[\,\mathbb{E}[C_n | X_n, \Theta_n]\,] = \mathbb{E}\left[\frac{\lambda(B_{X_n,\Theta_n})}{\lambda(S_{X_n,\Theta_n})}\right].$$

For all $x \in S$ and $\theta \in \partial D$, $B_{x,\theta}$ is a straight line within B, and $\lambda(B_{x,\theta})$ is bounded by R, the diameter of B. Therefore,

$$\mathbb{E}[C_n] \leq \mathbb{E}\left[\frac{R}{\lambda(S_{X_n,\Theta_n})}\right] = R \cdot \mathbb{E}\left[\frac{1}{\lambda(S_{X_n,\Theta_n})}\right].$$

Observe that $\mathbb{E}[1/\lambda(S_{X_n,\Theta_n})]$ does not depend on B, since X_n is taken from hit-and-run defined by only S and ν, and Θ_n is distributed by ν. □

3.2 Computation of the Bound on EC for the Union of Convex Bodies

If X_0 is uniform, or if n is sufficiently large, the Markov chain is stationary, and then $EC = \mathbb{E}[C_n]$. In this section, we derive a bound on EC when the open region S is a finite union of convex bodies. Since the complexity bound is dictated by each single convex body in the union, we begin by analyzing $\mathbb{E}[1/\lambda(S_{X_n,\Theta_n})]$ for a single convex body. We then obtain a bound on EC for the finite union of convex bodies by applying the quantity to each single convex body of the union.

Let the dimension d be fixed, and consider first the case when S is a convex body, and B is a hyperrectangle with diameter R containing S. Construct a hit-and-run process $(X_n; n \geq 0)$ from Algorithm 2.1 where X_n is uniformly distributed on S. Consider the expectation term in Proposition 3.1,

$$\begin{aligned}\mathbb{E}\left[\frac{1}{\lambda(S_{X_n,\Theta_n})}\right] &= \int_{\partial D}\int_S \frac{1}{\lambda(S_{x,\theta})\lambda_d(S)}\, d\lambda_d(x)\, d\nu(\theta) \\ &= \frac{1}{\lambda_d(S)}\int_{\partial D}\int_S \frac{1}{\lambda(S_{x,\theta})}\, d\lambda_d(x)\, d\nu(\theta).\end{aligned} \tag{5}$$

The first equality follows by Fubini's theorem since X_n and Θ_n are independent, and because X_n is uniformly distributed on S.

Now fix $\theta \in \partial D$. Let $S_\theta^\perp$ be the projection of S onto the hyperplane perpendicular to θ. For $p \in S_\theta^\perp$, let S_θ^p be the line segment in which S intersects the line through

p in θ direction and be linearly parameterized by t. Then by Fubini's theorem,

$$\int_S \frac{1}{\lambda(S_{x,\theta})} d\lambda_d(x) = \int_{S_\theta^\perp} \int_{S_\theta^p} \frac{1}{\lambda(S_{x,\theta})} d\lambda(t)\, d\lambda_{d-1}(p) = \int_{S_\theta^\perp} d\lambda_{d-1}(p) = \lambda_{d-1}(S_\theta^\perp).$$

The second equality follows since the inner integral is equal to 1, by the definition of $\lambda(S_{x,\theta})$. Therefore,

$$\mathbb{E}\left[\frac{1}{\lambda(S_{X_n,\Theta_n})}\right] = \frac{\int_{\partial D} \lambda_{d-1}(S_\theta^\perp)\, d\nu(\theta)}{\lambda_d(S)}. \tag{6}$$

Note that (6) is true for any d-dimensional region S.

LEMMA 3.2. *For a convex body $S \subset \Re^d$ containing an open hypersphere of radius r where $d \geq 2$, if X_n is uniformly distributed on S, then*

$$\mathbb{E}\left[\frac{1}{\lambda(S_{X_n,\Theta_n})}\right] \leq \frac{d}{2r}. \tag{7}$$

PROOF. Without loss of generality, let the hypersphere of radius r contained in S be centered at the origin. We find a lower bound of $\lambda_d(S)$ in terms of $\lambda_{d-1}(S_\theta^\perp)$,

$$\lambda_d(S) = \int_S d\lambda_d(x) = \int_{S_\theta^\perp} \int_{S_\theta^p} d\lambda(t)\, d\lambda_{d-1}(p).$$

Assume that $d > 2$. For any $\theta \in \partial D$, we write $\int_{S_\theta^\perp} \cdot\, d\lambda_{d-1}(p)$ using a $(d-1)$-dimensional spherical coordinate system,

$$\begin{aligned}\int_{S_\theta^\perp} \int_{S_\theta^p} d\lambda(t)\, d\lambda_{d-1}(p) &= \int_{\partial D_{d-1}} \int_0^{\bar\rho(\phi)} \int_{\underline{t}(\rho,\phi)}^{\bar t(\rho,\phi)} dt \rho^{d-2}\, d\rho\, ds(\phi) \\ &= \int_{\partial D_{d-1}} \int_0^{\bar\rho(\phi)} \left(\bar t(\rho,\phi) - \underline{t}(\rho,\phi)\right) \rho^{d-2}\, d\rho\, ds(\phi),\end{aligned}$$

where ∂D_{d-1} is the surface of $(d-1)$-dimensional unit hypersphere and $ds(\phi)$ is the differential of the $(d-2)$-dimensional Lebesgue measure over ∂D_{d-1} parameterized by direction vector ϕ.

To obtain a bound on $\bar t(\rho,\phi) - \underline{t}(\rho,\phi)$, we establish three points in S and consider the triangle they form. For each direction ϕ, because $\bar\rho(\phi)$ is the limit of the inner integral, there exists a point $(t, \bar\rho(\phi) - \varepsilon, \phi)$ in S, for a small $\varepsilon > 0$. The points $(r - \varepsilon, 0, \phi)$ and $(-r + \varepsilon, 0, \phi)$ are also contained in S, because S contains an open ball of radius r. The triangle formed by $(t, \bar\rho(\phi) - \varepsilon, \phi)$, $(r - \varepsilon, 0, \phi)$ and $(-r + \varepsilon, 0, \phi)$ is contained in S, because S is convex. With the base being defined by $(r - \varepsilon, 0, \phi)$ and $(-r + \varepsilon, 0, \phi)$, the triangle's base length and height equal $2r - 2\varepsilon$ and $\bar\rho(\phi) - \varepsilon$, respectively. Therefore, by a similar triangle property, at each height ρ from the base of the triangle, the width is equal to

$$(2r - 2\varepsilon)\frac{\bar\rho(\phi) - \varepsilon - \rho}{\bar\rho(\phi) - \varepsilon}.$$

Because the triangle is contained in S,

$$\bar t(\rho,\phi) - \underline{t}(\rho,\phi) \geq (2r - 2\varepsilon)\frac{\bar\rho(\phi) - \varepsilon - \rho}{\bar\rho(\phi) - \varepsilon}.$$

This is true for all $\varepsilon > 0$. Therefore,

$$\bar{t}(\rho,\phi) - \underline{t}(\rho,\phi) \geq 2r\frac{\bar{\rho}(\phi)-\rho}{\bar{\rho}(\phi)},$$

and

$$\begin{aligned}
\lambda_d(S) &\geq \int_{\partial D_{d-1}} \int_0^{\bar{\rho}(\phi)} 2r\frac{\bar{\rho}(\phi)-\rho}{\bar{\rho}(\phi)}\rho^{d-2}\,d\rho\,ds(\phi) \\
&= 2r\int_{\partial D_{d-1}} \left(\int_0^{\bar{\rho}(\phi)} \rho^{d-2}\,d\rho - \frac{\bar{\rho}(\phi)^{d-1}}{d}\right) ds(\phi) \\
&= 2r\int_{\partial D_{d-1}} \int_0^{\bar{\rho}(\phi)} \left(1-\frac{d-1}{d}\right)\rho^{d-2}\,d\rho\,ds(\phi) \\
&= \frac{2r}{d}\int_{\partial D_{d-1}} \int_0^{\bar{\rho}(\phi)} \rho^{d-2}\,d\rho\,ds(\phi) \\
&= \frac{2r}{d}\lambda_{d-1}(S_\theta^\perp). \qquad (8)
\end{aligned}$$

By using a similar triangular argument, (8) is also valid for the case $d = 2$. The lemma then follows using (8) and (6). □

Now we compute a bound on EC when S is a finite union of convex bodies.

PROPOSITION 3.3. *Assume that $S \in \Re^d$ is a finite union of m convex bodies S_i contained in a hyperrectangle B with diameter R, e.g.*

$$S = \cup_{i=1}^m S_i,$$

where each S_i contains a hypersphere of radius r. Then

$$EC \leq K\frac{dR}{2r}, \qquad (9)$$

for $d \geq 2$ where $K = \sum_{i=1}^m \lambda_d(S_i)/\lambda_d(S)$.

PROOF. Let X_n be uniformly distributed on S. Define $T_1 = S_1$ and, for $i = 2,\ldots,m$,

$$T_i = S_i \backslash (\cup_{k=1}^{i-1} T_k)$$

where $A\backslash B$ denotes A excluding B. Then

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{\lambda(S_{X_n,\Theta_n})}\right] &= \sum_{i=1}^m \mathbb{E}\left[\frac{1}{\lambda(S_{X_n,\Theta_n})}\ \middle|\ X_n \in T_i\right]\mathbb{P}(X_n \in T_i) \\
&= \sum_{i=1}^m \mathbb{E}\left[\frac{1}{\lambda(S_{X_n,\Theta_n})}\ \middle|\ X_n \in T_i\right]\frac{\lambda_d(T_i)}{\lambda_d(S)}. \qquad (10)
\end{aligned}$$

The second equality follows because X_n is uniformly distributed over S. Now consider only the conditional expectation term,

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{\lambda(S_{X_n,\Theta_n})}\;\middle|\; X_n \in T_i\right] &= \int_{\partial D}\int_{T_i} \frac{1}{\lambda(S_{x,\theta})\lambda_d(T_i)} d\lambda_d(x)\, d\nu(\theta) \\
&= \frac{\lambda_d(S_i)}{\lambda_d(T_i)} \int_{\partial D}\int_{T_i} \frac{1}{\lambda(S_{x,\theta})\lambda_d(S_i)} d\lambda_d(x)\, d\nu(\theta) \\
&\leq \frac{\lambda_d(S_i)}{\lambda_d(T_i)} \int_{\partial D}\int_{S_i} \frac{1}{\lambda(S_{x,\theta})\lambda_d(S_i)} d\lambda_d(x)\, d\nu(\theta) \\
&\leq \frac{\lambda_d(S_i)}{\lambda_d(T_i)} \int_{\partial D}\int_{S_i} \frac{1}{\lambda(S_{i_{x,\theta}})\lambda_d(S_i)} d\lambda_d(x)\, d\nu(\theta) \\
&\leq \frac{\lambda_d(S_i)}{\lambda_d(T_i)} \frac{d}{2r}. \qquad (11)
\end{aligned}$$

The first inequality follows from $T_i \subset S_i$ and the integrand is a nonnegative function. The second inequality follows from $S_i \subset S$ and, hence $\lambda(S_{i_{x,\theta}}) \leq \lambda(S_{x,\theta})$. The third inequality follows from Lemma 3.2. Substituting (11) into (10),

$$\mathbb{E}\left[\frac{1}{\lambda(S_{X_n,\Theta_n})}\right] \leq \left(\frac{\sum_{i=1}^m \lambda_d(S_i)}{\lambda_d(S)}\right)\frac{d}{2r} = K\frac{d}{2r}, \qquad (12)$$

where $K = \sum_{i=1}^m \lambda_d(S_i)/\lambda_d(S)$. Since X_n is uniformly distributed on S, the hit-and-run process is stationary and $EC = \mathbb{E}[C_n]$. Equation (12), Equation (5) and Proposition 3.1 imply Proposition 3.3. □

Observe that the constant K in Proposition 3.3 is the ratio between the sum of volumes of the convex bodies constituting S and the volume of S. If the convex bodies in S do not intersect with one another, then K is equal to one.

3.3 The Bound on the Complexity

Let $(X_n; n \geq 0)$ be generated from Algorithm 2.1. Then let N_ϵ be the number of iterations required such that the distribution of the iteration point attains the uniform distribution on S within ϵ error, as defined in Section 2.

THEOREM 3.4. *There exists a bound for the expected number of sampling points and the conditional expected number of sampling points of Algorithm 2.1 that grows linearly in R, the diameter of B. In particular,*

$$ET_\epsilon \leq R \cdot \sum_{n=0}^{N_\epsilon - 1} \mathbb{E}\left[\frac{1}{\lambda(S_{X_n,\Theta_n})}\right]. \qquad (13)$$

PROOF. From (1) and Proposition 3.1,

$$ET_\epsilon = \sum_{n=0}^{N_\epsilon - 1} \mathbb{E}[C_n] \leq \sum_{n=0}^{N_\epsilon - 1} R \cdot \mathbb{E}\left[\frac{1}{\lambda(S_{X_n,\Theta_n})}\right].$$

Observe that the term $\sum_{n=0}^{N_\epsilon-1} \mathbb{E}[1/\lambda(S_{X_n,\Theta_n})]$ depends on the stochastic process $(X_n; n \geq 0)$ and $(\Theta_n; n \geq 0)$, which do not depend on B. Therefore, the bound in Theorem 3.4 grows linearly in R, the diameter of B. □

COROLLARY 3.5. *Assume that $S \subset \Re^d$ is a finite union of m convex bodies contained in a hyperrectangle B with radius R, and each convex body constituting S contains a hypersphere with radius r. Assume that X_0 is uniformly distributed over S. Then*

$$ET_\epsilon \leq N_\epsilon K \frac{dR}{2r}, \tag{14}$$

for $d \geq 2$ and $K = \sum_{i=1}^{m} \lambda_d(S_i)/\lambda_d(S)$ where K and N_ϵ are independent of B.

PROOF. Since X_0 is uniformly distributed over S, the process is stationary and $\mathbb{E}[C_n] = EC$. Substituting (9) in Proposition 3.3 into (1), yields

$$ET_\epsilon = \sum_{n=0}^{N_\epsilon - 1} \mathbb{E}[C_n] \leq \sum_{n=0}^{N_\epsilon - 1} K \frac{dR}{2r} = N_\epsilon K \frac{dR}{2r}.$$

□

In Lovász and Vempala [2006], it is shown that, if S is a well rounded convex body, then N_ϵ is of order $O^*(d^4)$. With this result, Lovász and Vempala point out that the original hit-and-run is one of the fastest MCMC uniform samplers for a well rounded convex body. A well rounded convex body contains a unit hypersphere and is contained in a hypersphere with radius d, so $r = 1$ and $R = 2d^{3/2}$. For a single convex body, $K = 1$. Therefore, ET_ϵ is of order $O^*(d^{6\frac{1}{2}})$. Since any convex body can be transformed into a well rounded one in polynomial time by an affine transformation [Lovász 1999], this bound is applicable to problems involving a convex body where an affine transformation of the original convex body is allowed. Therefore, the overall complexity of hit-and-run on a box applied to a convex body is only slightly worse than that of the original hit-and-run, and the power of the original hit-and-run is still carried over to the hit-and-run on a box.

4. A COMPUTATIONAL STUDY

We compare the performance of Algorithm 2.1, hit-and-run on a box (HRB), with Algorithm 2.3, accelerated hit-and-run on a box (AHRB), on two types of nonconvex bodies, a union of hyperspheres and a union of hypercubes. The objective is to evaluate the potential benefit of employing AHRB over HRB.

Consider first a union of two unit hyperspheres whose centers are one unit apart,

$$S = S_1 \cup S_2, \text{ where } S_i = \{x \in \Re^d : \|x - x_i\|_2 < 1\}, \text{ for } i = 1, 2,$$

where $x_1 = [0.5, 0, 0, \ldots, 0]^T$ and $x_2 = [-0.5, 0, 0, \ldots, 0]^T$, and $\|\cdot\|_2$ denotes the Euclidean norm. We run HRB and AHRB to compute the center of gravity of the region. The solution to this simple problem is obviously the origin. We apply HRB and AHRB by enclosing S in a box $[-b, b]^d$. For each box size b and dimension d, 10 simulations are run in multiples of 1000 hit-and-run iterations and stop when the 95% confidence interval of the first component of the estimate contains 0 and the width of the interval is smaller than 0.1. The performance measure is defined as the average number of sample points (both accepted and rejected) over the 10 simulation runs. Figure 1 shows the performances of HRB and AHRB at different values of b and d.

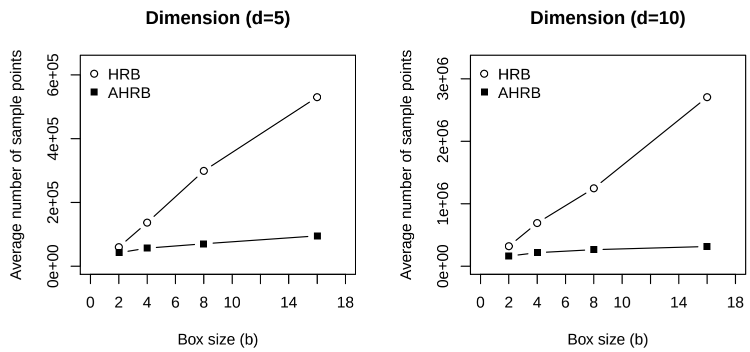

Fig. 1. Performances of HRB and AHRB on the union of 2 hyperspheres at different b and d

For each fixed dimension, it is obvious that AHRB requires significantly fewer sampling points than HRB does, and the advantage increases with the box size.

Now consider a union of two hypercubes

$$S = S_1 \cup S_2, \text{ where } S_i = \{x \in \Re^d : \ \|x - x_i\|_\infty < 1\}, \text{ for } i = 1, 2,$$

where $x_1 = [0.5, 0.5, \ldots, 0.5]^T$, $x_2 = [-0.5, -0.5, \ldots, -0.5]^T$, and $\| \cdot \|_\infty$ is the max norm. This S is more difficult because of the corners, and the overlapping part of the two hypercubes decreases as the dimension increases, causing the two cubes to be less connected. We repeat the same simulation experiment on the union of the two cubes. The results are shown in Figure 2, and again AHRB dominates HRB.

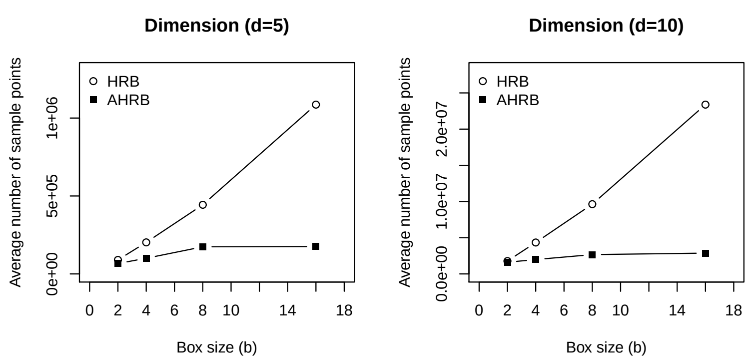

Fig. 2. Performances of HRB and AHRB on the union of 2 hypercubes at different b and d

Table I shows the number of sampling points and the number of iteration points required in each problem instance when $d = 10$. It is worth noting that, even though the average number of sampling points of AHRB is much less than that of HRB, the average number of iteration points of AHRB is greater than that of HRB. This is because the shrinking scheme of AHRB causes the transition to be more localized, while a high dimensional nonconvex S requires a higher degree of global reaching to obtain a good mixing rate.

Table I. Number of sampling points and iteration points of HRB and AHRB on each problem instance when $d = 10$ (unit in 1000 points)

Box size	Sampling points (both accepted and rejected)				Iteration points (accepted only)			
	2 hyperspheres		2 hypercubes		2 hyperspheres		2 hypercubes	
	HRB	AHRB	HRB	AHRB	HRB	AHRB	HRB	AHRB
2	320.6	162.5	1,777.2	1,606.7	39.7	42.6	293.7	534.1
4	691.5	217.2	4,342.1	1,986.9	41.1	42.3	271.4	426.1
8	1,247.5	265.1	9,635.1	2,668.4	36.9	41.0	281.4	441.0
16	2,706.6	315.7	23,382.9	2,872.1	39.8	40.4	337.5	387.5

5. CONCLUSION

We analyze the effect of a hyperrectangle as a sampling agent in obtaining a hit-and-run process. We show that its effect on the computational complexity is of linear order in its diameter. The bound implies that it is not crucial to find the hyperrectangle that best fits the support. We also demonstrate in a computational study that, by shrinking the line set at each rejection step, hit-and-run on a box can be sped up considerably.

ACKNOWLEDGMENTS

This work is partially supported by the Faculty of Commerce and Accountancy of Chulalongkorn University research fund under Chulalongkorn University Centennial. Zelda Zabinsky would like to thank the Faculty of Commerce and Accountancy, Chulalongkorn University, for sponsoring her as a Visiting Professor during her sabbatical and supporting in part this paper.

REFERENCES

Asmussen, S. and Glynn, P. W. 2007. *Stochastic Simulation: Algorithms and Analysis*. Springer-Verlag.

Bélisle, C., Boneh, A., and Caron, R. J. 1998. Convergence properties of hit-and-run samplers. *Stochastic Models 14*, 767–800.

Bélisle, C. J. P., Romeijn, H. E., and Smith, R. L. 1993. Hit-and-Run algorithm for generating multivariate distributions. *Mathematics of Operations Research 18*, 255–266.

Diaconis, P. and Freedman, D. 1997. On the hit and run process. Tech. Rep. 497, Department of Statistics, University of California, Berkeley.

Kannan, R., Lovász, L., and Simonovits, M. 1997. Random walks and an $o^*(n^5)$ volume algorithm for convex bodies. *Random Structures and Algorithms 11*, 1–50.

Lovász, L. 1999. Hit-and-run mixes fast. *Mathematical Programming 86*, 443–461.

Lovász, L. and Vempala, S. 2006. Hit-and-run from a corner. *SIAM Journal of Computing 35(4)*, 985–1005.

Neal, R. M. 2003. Slice sampling. *The Annals of Statistics 31*, 705–767.

Quinn, K. 2004. Bayesian factor analysis for mixed ordinal and continuous responses. *Political Analysis 12*, 338–353.

Romeijn, H. E. and Smith, R. L. 1994. Simulated annealing and adaptive search in global optimization. *Probability in the Engineering and Informational Sciences 8*, 571–590.

Smith, R. L. 1984. Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions. *Operations Research 32*, 1296–1308.

Received; revised; accepted