

DIRECTION CHOICE FOR ACCELERATED CONVERGENCE IN HIT-AND-RUN SAMPLING

David E. Kaufman

Department of Industrial Engineering and Operations Research
University of Massachusetts
Amherst, MA 01003-5220

Robert L. Smith

Department of Industrial and Operations Engineering
University of Michigan
Ann Arbor, Michigan 48109

August 13, 1994

Abstract

Hit-and-Run algorithms are Monte Carlo procedures for generating points that are asymptotically distributed according to general absolutely continuous target distributions G over open bounded regions S . Applications include nonredundant constraint identification, global optimization, and Monte Carlo integration. These algorithms are reversible random walks which commonly apply uniformly distributed step directions. We investigate *nonuniform* direction choice and show that under minimal restrictions on the region S and target distribution G , there exists a unique direction choice distribution, characterized by necessary and sufficient conditions depending on S and G , which optimizes a bound on the rate of convergence. We provide computational results demonstrating greatly accelerated convergence for this optimizing direction choice and for more easily implemented adaptive heuristic rules.

1 Introduction

We consider the Monte Carlo problem of generating a sample of points according to a given probability distribution G over an open, bounded region S in \mathbb{R}^n . After motivating the problem through several applications, this section discusses the limitations of exact sampling methods, describes the Hit-and-Run asymptotically exact method, and previews the contribution of the paper to Hit-and-Run direction choice.

1.1 Sampling Applications

Nonredundant constraint identification (cf. Karwan et al. [16]). Given a point x satisfying a system of linear inequalities and a unit direction u , an inequality is nonredundant if it is uniquely the nearest constraint to x in direction $+u$ or $-u$. As a special case, consider points x sampled uniformly from the relative interior of the feasible region of a linear program and directions u sampled uniformly on the unit hypersphere, with all pairs of points and directions independent. The probability that the sample fails to identify any nonredundant constraint decreases to zero as the sample size increases to infinity. We can then more rapidly solve a reduced problem including only the identified nonredundant constraints. These may not all be found by finite sampling, but even if the optimal solution to the reduced problem is infeasible for the full problem, the optimum for the dual reduced problem remains feasible and provides a good initial point for solving the full dual problem.

Global optimization (cf. Dixon and Szegö [7, 8], Rubinstein [21]). Deterministic iterative optimization strategies typically choose locally improving search directions and step sizes yielding improvement in each iteration. These methods often yield solutions which are only locally optimal for nonconvex problems. A simple stochastic alternative for global optimization is Pure Random Search [5], which samples points uniformly in the feasible region and reports the sampled point with the best objective function as the optimal solution. More practical methods commonly rely on the Multistart approach, applying local search algorithms from some or all of a group of solutions chosen randomly from the feasible region [20]. Both methods rely on efficient sampling of feasible points.

The potential of stochastic global optimization methods is illustrated by Pure Adaptive Search (PAS) [19, 30], in which a new iterate is uniformly distributed over the subset of the feasible region superior in objective value to the current iterate. The number of PAS iterations required to approach the global optimum arbitrarily closely is shown to grow only linearly in dimension, a result previously suggested by the computational experience of Solis and Wets [24]. For stochastic global methods in general and PAS in particular, efficient sampling from arbitrary regions is therefore of significant importance.

Monte Carlo Integration (cf. Hammersley and Handscomb [13]). We may evaluate $I = \int_S f(x) dx$ by noting that if the random variable X is distributed according to a positive

probability density function g on S , then

$$I = \int_S \frac{f(x)}{g(x)} g(x) dx = E \left[\frac{f(X)}{g(X)} \right].$$

The sample-mean method of Monte Carlo integration consists of generating X_1, \dots, X_N according to g and estimating I by the unbiased estimator $\hat{I} = (1/N) \sum_{i=1}^N f(X_i)/g(X_i)$. Sample-mean estimation may be superior to deterministic numerical methods for nonsmooth integrands with multidimensional domain [21]. The potentially strong dependence of the variance of \hat{I} on the choice of g motivates interest in sampling from general distributions over S .

2 Sampling methods

2.1 Exact sampling

We begin by considering some of the shortcomings associated with the exact sampling methods of transformation, composition, and rejection (Schmeiser [22]). A vector X of n numbers drawn independently from the uniform distribution over the interval $[0, 1]$ is uniformly distributed over the unit cube in \mathfrak{R}^n . Given a bijective differentiable transformation T from the unit cube onto S , then $T(X)$ is uniformly distributed on S if T has constant Jacobian determinant. However, such *transformation* techniques are known only for a small class of regions S , such as paralleleptopes, hyperspheres, and simplices. *Composition* techniques express the probability density to be sampled as a mixture of densities conditioned on a random parameter, for cases when the decomposed densities can be sampled more easily than the composite [21]. These methods tend to be intractable in even moderate dimension [27]. The most generally applicable methods are *rejection* techniques. Sampling uniformly from a region R enclosing S and rejecting those points not in S , the remaining sample is uniformly distributed over S . However, the expected number of points in R needed in order to hit S grows rapidly in dimension, making rejection techniques inefficient in high dimension for many cases of S and R . The same problems are associated with sampling nonuniformly according to a density g over S , since this is equivalent to sampling uniformly in the region under the graph of g over S .

2.2 Hit-and-Run algorithms

With each Hit-and-Run algorithm is associated a direction probability distribution H over the unit hypersphere. The algorithm proceeds from a current point x^m by generating a direction u according to H and selecting x^{m+1} according to G conditionalized on the resulting line set, i.e., the subset of S lying in the direction $\pm u$ from x^m .

The most widely known version is the Hypersphere Directions (HD) Hit-and-Run algorithm, proposed in 1979 by Boneh and Golan [4] and independently in 1980 by Smith [26]. HD selects directions according to a uniform distribution over the unit sphere and chooses iterates uniformly on the resulting line set; in [27], Smith proves that the HD iterates approach a uniform limiting distribution independently of the starting point for open bounded search regions S , and demonstrates experimentally that Hit-and-Run is potentially more efficient than rejection techniques in high dimension. Another example is Coordinate Directions (CD) Hit-and-Run, due to Telgen [29], in which the direction is chosen uniformly from among the $2n$ coordinate vectors (the vectors which parallel the coordinate axes in \mathfrak{R}^n) and the new iterate is chosen uniformly on the resulting line set as in HD. These two algorithms were analyzed and applied in Berbee et al. [2] to identify nonredundant linear constraints.

Bélisle, Romeijn, and Smith [1] have shown asymptotic convergence for the generalized Hit-and-Run class when S is open and bounded and G is absolutely continuous with a density that is positive, bounded, and continuous almost everywhere with respect to Lebesgue measure. Schmeiser and Chen [23] have generalized to unbounded S and unbounded densities of G . Under minimal restrictions on the direction distribution H , choosing new iterates conditionally according to G on the line set causes the distribution of iterates to converge in total variation to G . The direction distribution does not affect the asymptotic target distribution of the iterates, but as we shall demonstrate, it has a strong effect on convergence rate.

2.3 Other iterative sampling methods

Another iterative method for sampling according to a given asymptotic distribution is the Gibbs sampler, due to Geman and Geman [11]. The Gibbs sampler samples several random variables according to a joint density by sampling one of the random variables from its conditional distribution given the values of the other variables, which are held fixed. The random variable to be changed at each step is rotated so that each of the variables is resampled in sequence. In the Euclidean-space setting of our problems, the Gibbs sampler is the same as the CD Hit-and-Run algorithm, except that CD randomly chooses which variable to resample in each iteration rather than following a deterministic sequence. Tanner and Wong [28] propose a similar method called data augmentation. For a review of these techniques, see Gelfand and Smith [10]. Our study of random direction choice will not apply directly to the Gibbs sampling approach due to the deterministic character of the Gibbs sequential direction choice rule. Gibbs sampling tends to reduce the computational effort of each iteration, because no random direction is generated. However, the Gibbs approach is inflexible due to its dependence on the alignment of the search region with the coordinate axes. Chen and Schmeiser [6] verify empirically that for bivariate normal target distributions, Hit-and-Run is more robust than Gibbs with respect to variation in the covariance of the distribution.

Another approach is Metropolis sampling, due to Metropolis et al. [18], which differs

from Hit-and-Run in that it generates random step sizes by an acceptance/rejection rule applied to a uniformly distributed point from the line set, rather than by conditionalization. This approach was extended by Hastings [14] and examined for very general settings in \mathfrak{R}^n by Bélisle, Romeijn, and Smith [1]. Hit-and-Run, Metropolis sampling, and Gibbs sampling are all instances of Markov Chain Monte Carlo; for a review, see Geyer [12].

3 The Hit-and-Run Algorithm

Denote the unit ball in \mathfrak{R}^n by $B = \{u \in \mathfrak{R}^n \mid \|u\| \leq 1\}$, whose topological boundary ∂B is the unit hypersphere. Let $S \in \mathfrak{R}^n$ be an open bounded set. Let G be the target probability distribution on S , assumed absolutely continuous with respect to Lebesgue measure V , with positive density g continuous almost everywhere and bounded from above and away from zero. Let σ be the measure giving surface area of measurable subsets of ∂B , and let H be the probability distribution of direction choice on ∂B , assumed absolutely continuous with respect to σ and having positive density h . Also, for $x, y \in S$ let $u_{xy} = (y - x)/\|y - x\|$ represent the unit direction from x to y .

We now formally define the *Hit-and-Run algorithm* in S associated with the target distribution G and the direction distribution H .

Hit-and-Run Algorithm

- 1) Choose an arbitrary starting point $x^0 \in S$ and set $m = 0$.
- 2) Generate a random direction $u_m \in \partial B$ according to H .
- 3) Select λ_m from the line set $\Lambda_m = \{\lambda \in \mathfrak{R} \mid x^m + \lambda u_m \in S\}$ according to the density

$$g_m(\lambda) = \frac{g(x^m + \lambda u_m)}{\int_{\Lambda_m} g(x^m + r u_m) dr} \quad \lambda \in \Lambda_m.$$

- 4) Set $x^{m+1} = x^m + \lambda_m u_m$ and $m = m + 1$, and go to step 2.

Because step 3 searches for both positive and negative step sizes λ in direction u , we may when convenient assume without loss of generality that $h(u) = h(-u)$ for all $u \in \partial B$.

By Theorem 5 of [1] the distribution of the Hit-and-Run iterates $\{x^m, m = 1, 2, \dots\}$ converges in total variation to G for any initial x^0 , i.e.,

$$\lim_{m \rightarrow \infty} \Pr(x^m \in A \mid x^0 = x) = G(A) \quad A \in \mathcal{B}_S, x \in S$$

uniformly in A , where \mathcal{B}_S represents the Borel σ -field on S . In fact, the cited theorem is more general, not requiring g to be bounded allowing much more general direction distributions

H . In particular, our current restrictions on H exclude the CD version of Hit-and-Run (where the direction distribution is not absolutely continuous) and other variations which have zero direction density on some direction set of positive surface measure. However, the convergence rate analysis here requires the stronger conditions.

For the HD algorithm, where G and H are uniform distributions, Smith in [27] established the following bound on convergence rate:

$$\left| \Pr(x^m \in A | x^0 = x) - \frac{V(A)}{V(S)} \right| \leq \left(1 - \frac{\gamma}{n2^{n-1}} \right)^{m-1} \quad A \in \mathcal{B}_S, x \in S \quad (1)$$

where γ is the ratio of the volume of S to the volume of the sphere in \mathfrak{R}^n whose diameter is that of S . As an example, if S is a 10-dimensional cube, $\gamma \approx 1/250$ and the number m of iterations required to upper-bound the error term on the left by 0.01 is over five million. Our goal is to generalize this bound to *nonuniform* target distribution G and direction distribution H and then find a direction distribution H^* which optimizes the bound on rate of convergence to the distribution G over the region S . Ideally, we hope moreover to achieve improvement in experimental performance (i.e., average case behavior) as well as in worst-case behavior.

4 Optimizing the Rate of Convergence

Assuming the region S and the target distribution G to be fixed and satisfying the conditions of the previous section, let $P_H(A|x)$ be the one-step transition probability distribution for the Hit-and-Run algorithm with direction distribution H . That is,

$$P_H(A|x) = \Pr(x^{m+1} \in A | x^m = x) \quad A \in \mathcal{B}_S, x \in S.$$

The following lemma expresses the transition probability density, which will contribute to the derivation of the convergence rate bound. We define $\Lambda(x, y) = \{\lambda \in \mathfrak{R} | x + \lambda u_{xy} \in S\}$ as the set of feasible step sizes in the direction between x and y . (See [1] for an independently derived expression for more general direction distributions and step size rules. The proof given here is preserved for its relative simplicity and the comparative tractability of the resulting density expression. We do not define $f_H(x|x)$; it can be an arbitrary positive number, since $V(\{x\}) = 0$.)

Lemma 1 *Given direction distribution H absolutely continuous with density h , for all $x \in S$ the transition probability distribution $P_H(\cdot|x)$ is absolutely continuous with density*

$$f_H(y|x) = \frac{(h(u_{xy}) + h(-u_{xy}))g(y)}{\|y - x\|^{n-1} \int_{\Lambda(x,y)} g(x + ru_{xy}) dr} \quad y \in S, y \neq x. \quad (2)$$

Proof: Fix x . For $0 \leq r_1 < r_2$ and measurable $D \subset \partial B$ such that $D \cap (-D) = \emptyset$, define $C_D(r_1, r_2) = \{x + ru \mid u \in D, r_1 < r < r_2\}$, and let \mathcal{C} be the class of all such sets. Let $I_S(\cdot)$ be the function indicating membership in S , and let $S_{x,u}$ be a random variable giving the Hit-and-Run step size from x in direction u . Let U be a random variable distributed according to H over ∂B .

For $C = C_D(r_1, r_2) \in \mathcal{C}$,

$$\begin{aligned}
P_H(C \cap S \mid x) &= \Pr(U \in D, S_{x,U} \in (r_1, r_2)) + \Pr(U \in -D, -S_{x,U} \in (r_1, r_2)) \\
&= \int_D \Pr(S_{x,u} \in (r_1, r_2))(h(u) + h(-u)) \sigma(du) \\
&= \int_D \int_{r_1}^{r_2} I_S(x + su) \frac{(h(u) + h(-u))g(x + su)}{s^{n-1} \int_{\Lambda(x, x+u)} g(x + ru) dr} s^{n-1} ds \sigma(du) \\
&= \int_C I_S(y) \frac{(h(u_{xy}) + h(-u_{xy}))g(y)}{\|y - x\|^{n-1} \int_{\Lambda(x,y)} g(x + ru_{xy}) dr} dy \\
&\quad \text{by change to spherical coordinates; see e.g. [15, p.4]} \\
&= \int_{C \cap S} \frac{(h(u_{xy}) + h(-u_{xy}))g(y)}{\|y - x\|^{n-1} \int_{\Lambda(x,y)} g(x + ru_{xy}) dr} dy. \tag{3}
\end{aligned}$$

Since \mathcal{B}_S is generated by the π -system \mathcal{C} restricted to S , equation (3) determines $P_H(A \mid x)$ for all $A \in \mathcal{B}_S$ as well [3, Theorem 3.3], completing the proof. ■

The following theorem bounds the rate of convergence to the target distribution by providing an upper bound on the deviation between the target distribution and the distribution of the m^{th} Hit-and-Run iterate for any m , over all $A \in \mathcal{B}_S$, as a function of the direction distribution H . We denote by \mathcal{H} the class of absolutely continuous direction distributions H with density h bounded away from zero. As stated in Section 3, we may assume without loss of generality that $h(u) = h(-u)$ for all $H \in \mathcal{H}$.

Theorem 2 For $H \in \mathcal{H}$,

$$|\Pr(x^m \in A \mid x^0 = x) - G(A)| \leq (1 - \delta_H V(S))^{m-1} \quad A \in \mathcal{B}_S, x \in S \tag{4}$$

where δ_H is the density bound for H , given by

$$\delta_H = \inf_{x,y \in S} f_H(y \mid x).$$

Proof: For $H \in \mathcal{H}$, we have S bounded, g bounded from above and away from zero, and h bounded away from zero. Hence from the preceding lemma $f_H(y \mid x)$ is bounded away from zero, i.e., $\delta_H > 0$. Then the theorem follows from a result of Doob [9, p.197, case b)]. ■

In order to minimize the error on the left of (4) regardless of iteration number m , we seek a direction distribution achieving the *optimal error* ϵ^* , given by

$$\epsilon^* = \inf_{H \in \mathcal{H}} (1 - \delta_H V(S)).$$

Therefore we formulate the convergence rate bound optimization as:

$$\text{(P)} \quad \text{Find } \delta^* = \sup_{H \in \mathcal{H}} \delta_H.$$

Since each δ_H is the value of a minimization, we call δ^* the *maximin density*. By an *optimal direction distribution* we will mean any distribution $H^* \in \mathcal{H}$ for which the maximin density δ^* is attained. Of course, our use of the term “optimal” is a shorthand, since such a distribution is optimal only in the worst-case rate of convergence as expressed by Theorem 2. We would hope to realize further improvements by applying the methodology of this paper to tighter bounds, if they should become available.

5 The Optimal Direction Distribution

We now investigate the maximin problem (P) in more detail. We will demonstrate that the problem has a unique solution H^* and characterize this solution by necessary and sufficient conditions on H .

For each $H \in \mathcal{H}$, we define $f_H^*(u)$, the *infimal transition density* in direction u , by

$$f_H^*(u) = \inf_{\substack{x, y \in S \\ u_{xy} = \pm u}} \{f_H(y|x)\} \quad u \in \partial B.$$

Then $\delta_H = \inf_{u \in \partial B} f_H^*(u)$. The infimal transition density decomposes into two parts, the first corresponding to the target distribution G and the geometry of S , which are considered fixed, and the second to the direction distribution H , which we wish to optimize.

Lemma 3 *The infimal transition density for H is given by*

$$f_H^*(u) = \frac{1}{\rho(u)} (h(u) + h(-u)) \quad u \in \partial B$$

where $\rho(u)$ is the span of S in direction u , defined independently of H by

$$\rho(u) = \sup_{\substack{x, y \in S \\ u_{xy} = \pm u}} \left\{ \frac{\int_{\Lambda(x,y)} g(x+ru) dr}{g(y)} \|y-x\|^{n-1} \right\} \quad u \in \partial B. \quad (5)$$

Proof: We have

$$\begin{aligned} \inf_{\substack{x,y \in S \\ u_{xy} = \pm u}} \{f_H(y|x)\} &= \inf_{\substack{x,y \in S \\ u_{xy} = \pm u}} \left\{ \frac{g(y)}{\int_{\Lambda(x,y)} g(x+ru) dr \|y-x\|^{n-1}} (h(u) + h(-u)) \right\} \\ &= \frac{1}{\rho(u)} (h(u) + h(-u)). \quad \blacksquare \end{aligned}$$

We now state the main result.

Theorem 4 *Let $S \subset \mathfrak{R}^n$ be an open bounded set, and let G be a probability distribution on S , absolutely continuous with respect to V with density g bounded from above and away from zero. Then*

- i) There exists a unique optimal direction distribution $H^* \in \mathcal{H}$.*
- ii) The direction distribution $H \in \mathcal{H}$ is optimal if and only if the infimal transition density for H is constant, i.e., if $f_H^* = c$ for some $c > 0$.*

Proof: We begin by proving sufficiency. Assume that for the direction distribution H with density h , there exists $c > 0$ such that $f_H^*(u) = c$ for all $u \in \partial B$. Suppose that H is not optimal, i.e., there exists a direction distribution H' with density h' such that

$$f_{H'}^*(u) \geq \delta_{H'} > \delta_H = f_H^*(u) \quad u \in \partial B.$$

Then by Lemma 3, $h'(u) + h'(-u) > h(u) + h(-u)$ for all $u \in \partial B$. But both h and h' integrate to one over ∂B , hence a contradiction ensues. No direction distribution improves on the density bound $\delta_H = c$, and H is optimal.

Now we prove existence. Let H^* be the direction distribution with density

$$h^*(u) = \frac{\rho(u)}{\int_{\partial B} \rho(v) \sigma(dv)} \quad u \in \partial B$$

and corresponding infimal transition density $f_{H^*}^*$. Observe that g bounded above and away from zero and S open and bounded implies ρ is bounded above and away from zero. Further, it is easily seen that $\rho(u) = \rho(-u)$ for all $u \in \partial B$, so h^* is likewise symmetric and hence $H^* \in \mathcal{H}$. For $u \in \partial B$,

$$f_{H^*}^*(u) = \frac{2}{\int_{\partial B} \rho(v) \sigma(dv)} \equiv c.$$

Therefore H^* satisfies the sufficient condition and is optimal, with $\delta^* = \delta_{H^*} = c$.

To prove the necessary condition, assume that direction distribution H with density h is optimal. Since H^* defined above is also optimal, $f_H^*(u) \geq \delta_H = \delta^* = f_{H^*}^*(u)$ for all $u \in \partial B$. Then by Lemma 3,

$$h(u) + h(-u) \geq h^*(u) + h^*(-u) \quad u \in \partial B. \quad (6)$$

Since both sides of this inequality must integrate over ∂B to the value 2, equation (6) must hold with equality almost everywhere on ∂B . Then by our earlier assumption that $h(u) = h(-u)$, we have $h = h^*$ almost everywhere, i.e., h^* is a density for H . Thus $f_H(u) = f_{H^*}(u) = c$ for all $u \in \partial B$ and the necessary condition is established. Furthermore, we have proven that any two optimal distributions must satisfy (6) with equality almost everywhere, i.e., that the optimal distribution is unique. ■

This main result has intuitive appeal. To maximize the minimum of a probability density, an obvious candidate is to choose a uniform distribution. The proof above shows that although the infimal transition density is not itself a probability density, it does have a certain normalization which depends on S and G through the span ρ , and to maximize the minimum, uniformity is again the solution.

From the proof of the above theorem, we have

Corollary 5 *Under the conditions of Theorem 4,*

i) The optimal distribution H^ is determined by the span ρ of S given in (5), with density*

$$h^*(u) = \frac{\rho(u)}{\int_{\partial B} \rho(v) \sigma(dv)}.$$

ii) The optimal error ϵ^ is given by*

$$\epsilon^* = 1 - \frac{2}{\int_{\partial B} \rho(v) \sigma(dv)} V(S).$$

Note also that H^* is absolutely continuous with respect to σ and has nonzero density on all of ∂B . Therefore, H^* satisfies the conditions imposed in Section 3 to ensure that the distributions of the iterates do in fact converge to the target distribution G .

6 Optimal Direction Choice for the Uniform Target Distribution over a Class of Convex Regions

We will now examine the issue of how to generate directions according to an optimal direction distribution H^* with density h^* . We could proceed by a rejection technique, generating

directions u uniformly on ∂B and accepting only if a uniform $(0, 1)$ -variate is less than $h^*(u)/\bar{h}$, where \bar{h} is an upper bound on h^* (recall ρ is bounded from above, and hence such a bound exists). This procedure is inefficient, particularly as n becomes large and S deviates from a spherical shape. Moreover, $h^*(u)$ may be difficult to calculate. Under certain conditions on the search region and target distribution, we can establish a potentially more efficient procedure.

Assume the target distribution G is uniform over S convex. Since g is constant, the span expression simplifies to

$$\rho(u) = \sup_{\substack{x, y \in S \\ u_{xy} = u}} \|y - x\|^n \quad u \in \partial B. \quad (7)$$

We further assume that S has a *center* s , which we define to be a point such that for almost all $u \in \partial B$, the supremum defining the span in direction u is realized by points $x, y \in \partial S$ (the topological boundary of S) with $s = (x + y)/2$. The class of centered regions is quite small by this definition, including spheres, rectangles, and regular polytopes, for example, but not simplices or general polytopes. However, the importance of centered regions as a motivating case will become apparent.

We define the *radius* in direction u by $r(u) = \sup\{r \geq 0 \mid s + ru \in S\}$ for all $u \in \partial B$. By (7), we have $\rho(u) = (r(u) + r(-u))^n = (2r(u))^n$ for all $u \in \partial B$.

Theorem 6 *Let $S \subset \mathfrak{R}^n$ be open, bounded, and convex with center s , and let the target distribution G be uniform. Let Y be a random variable uniformly distributed on S and let $U = (Y - s)/\|Y - s\|$. Then U has distribution H^* on ∂B .*

Proof: For measurable $D \subset \partial B$ let $S_s(D)$ be the part of S in the directions D from s , given by

$$S_s(D) = \left\{ x \in S : \frac{x - s}{\|x - s\|} \in D \right\} .$$

Then,

$$\begin{aligned} \Pr(U \in D) &= \Pr(Y \in S_s(D)) \\ &= \frac{\int_{S_s(D)} dy}{\int_S dy} \\ &= \int_D \frac{\frac{1}{n} r^n(u)}{\int_{\partial B} \frac{1}{n} r^n(v) \sigma(dv)} \sigma(du) \text{ by a change to spherical coordinates} \\ &= \int_D \frac{\rho(u)}{\int_{\partial B} \rho(v) \sigma(dv)} \sigma(du) \\ &= \int_D h^*(u) \sigma(du) \text{ by Corollary 5.} \end{aligned}$$

Hence U is distributed according to H^* on ∂B . ■

Therefore, to evaluate the efficiency of optimal direction choice on the class of centered convex regions we may implement exact optimal direction choice by generating points uniformly distributed in the search region and normalizing the vectors from the center to these points. Of course, if we could do this efficiently we would already have solved the problem which Hit-and-Run is designed to address. However, this theoretical result motivates an easily implemented approximation to optimal directions. Since the sequence of Hit-and-Run iterates converges to uniformity, we can try choosing directions by randomly choosing one of the previous Hit-and-Run iterates and normalizing the vector from an approximate center to the iterate. A simple scheme for approximating the center would be to set the i th coordinate equal to the mean of the i th coordinates of the previous Hit-and-Run iterates. Section 8 presents further discussion and computational results for a procedure of this kind

We now show that the class of centered convex regions is characterized by a single optimal convergence rate bound.

Theorem 7 *Let $S \subset \mathfrak{R}^n$ be open, bounded, and convex with center s , and let the target distribution G be uniform. Then the optimal error is identical to that for uniform direction choice in a spherical search region, i.e.,*

$$\epsilon^* = 1 - \frac{1}{n2^{n-1}}.$$

Proof: By Corollary 5,

$$\begin{aligned} \epsilon^* &= 1 - \frac{2}{\int_{\partial B} \rho(u) \sigma(du)} V(S) \\ &= 1 - \frac{2}{2^n \int_{\partial B} r^n(u) \sigma(du)} V(S) \\ &= 1 - \frac{1}{n2^{n-1}}. \end{aligned}$$

For a spherical search region, $\gamma = 1$ in (1) and hence the same bound obtains. ■

Now, we examine the optimal direction distributions and convergence rate bounds for particular regions S . The first example verifies an intuitive result.

Example 1 Let $S = \{x \in \mathfrak{R}^n \mid \|x\| < b\}$ so that S is the open ball of radius $b > 0$ centered at the origin in \mathfrak{R}^n . By (7), $\rho = (2b)^n$ and hence from Corollary 5 the optimal direction distribution is uniform. This is natural; S looks the same from all directions, and we have no reason to favor one direction over another. ■

Example 2 Now let $S = \{x \in \mathfrak{R}^n \mid 0 < x_i < b_i, i = 1, \dots, n\}$ so that S is an open rectangular parallelepiped determined by the upper bounds b_1, b_2, \dots, b_n . S is convex and has center $s = (b_1/2, b_2/2, \dots, b_n/2)$. For any unit direction u , a similar triangles argument shows $b_i/2|u_i|$ to be the distance in direction u from s to the i th upper bound constraint. Since $r(u)$ is the greatest distance which satisfies all n of the constraints,

$$\rho(u) = (r(u) + r(-u))^n \tag{8}$$

$$= \left(\min_{\substack{i=1, \dots, n \\ u_i \neq 0}} \left\{ \frac{b_i}{|u_i|} \right\} \right)^n. \tag{9}$$

Since the optimal density is proportional to ρ , we can see either geometrically from (8) or analytically from (9) that the density favors directions along the long axes of the rectangle, and is maximized by the directions from s to the corners of S .

Let us evaluate the uniform-direction and optimal convergence rate bounds. By Theorem 7, the optimal error is $1 - 1/(n2^{n-1})$. In contrast, recalling that γ represents the ratio of the volume of S to the volume of a sphere with diameter equal to that of S , the uniform bound is determined by the term

$$\begin{aligned} 1 - \frac{\gamma}{n2^{n-1}} &= 1 - \frac{V(S)}{n2^{n-1}(\frac{1}{2}\text{diam}(S))^n V(B)} \\ &= 1 - \frac{2 \prod_{i=1}^n b_i}{n(\sum_{i=1}^n b_i^2)^{n/2} V(B)}. \end{aligned}$$

For each of the upper bound vectors $b^0 = (1, 1, 1, \dots, 1)$, $b^1 = (1, 2, 3, \dots, 10)$, and $b^2 = (1, 4, 9, 16, \dots, 100)$ in \mathfrak{R}^{10} , Table 1 shows the values of the uniform-direction and optimal-direction convergence bound terms, as well as the number of iterations required to upper-bound the error terms in (1) and (4) by 0.01. Clearly optimal direction choice is greatly superior with respect to worst-case behavior, reducing the number of iterations required by two to six orders of magnitude. ■

7 Numerical Comparison of Uniform and Optimal Direction Distributions

We executed Hit-and-Run both with uniform and with optimal direction choice in each of the three hyperrectangles of Example 2. The testing procedure was as follows: run Hit-and-Run for 10,000 iterations, sampling each tenth point in order to reduce serial correlations. The resulting 1000 points in \mathfrak{R}^{10} were then shuffled to randomize their ordering. This procedure

Table 1: Convergence bound values in \mathfrak{R}^{10}

| Upper bound vector | Error terms | | # iters m required so that $\Pr(x^m \in \cdot x^0) - G(\cdot) \leq 0.01$ | |
|-------------------------------|--|---|--|-------------------|
| | Uniform dir: $\epsilon = 1 - \gamma/n2^{n-1}$ | Optimal direction: $\epsilon^* = 1 - 1/n2^{n-1}$ | Uniform dir. | Optimal direction |
| $b^0 = (1, 1, 1, \dots, 1)$ | $1 - 7.84 \cdot 10^{-7}$ | $1 - 1.95 \cdot 10^{-4}$ | 5.9 million | 24 thousand |
| $b^1 = (1, 2, 3, \dots, 10)$ | $1 - 3.36 \cdot 10^{-8}$ | $1 - 1.95 \cdot 10^{-4}$ | 137 million | 24 thousand |
| $b^2 = (1, 4, 9, \dots, 100)$ | $1 - 9.90 \cdot 10^{-11}$ | $1 - 1.95 \cdot 10^{-4}$ | 47 billion | 24 thousand |

produces ten samples of 1000 real-valued points, one sample in each coordinate direction. We performed two-tailed χ^2 frequency and serial correlation tests [17, pp. 59-60] to test the hypothesis that the samples are uniformly distributed with respect to each of the ten coordinate directions. The coordinate directions are each broken into 10 equal cells so that each frequency test has 9 degrees of freedom. Smith [27] performed the same computational test for HD on a cube in \mathfrak{R}^{10} (i.e., uniform directions with b^0 as the upper bound) and reported that seven of the ten coordinates passed the frequency test for uniformity at a significance level of 10% and nine of the ten coordinates passed the serial correlation test at the 10% significance level.

Tables 2 and 3 show the test statistics for the frequency and serial correlation tests respectively. They show that with upper bound vector b^0 , for which the optimal direction distribution is the closest to uniform of the three upper bounds tested, the results of uniform and optimal direction choice are comparable. However, when we elongate the region somewhat by using the upper bound b^1 , only four of the ten coordinate samples pass the frequency test with uniform directions, while nine of the ten coordinates pass with optimal direction choice. The very elongated region with bound b^2 yields results close to those for b^1 with respect to the number of tests passed by each direction choice method, but note that most of the coordinates which fail under uniform direction choice do so with spectacularly poor χ^2 values. The serial correlation test statistics tell a similar tale; although the serial tests for uniform direction choice are largely passed with bound b^1 , when we further elongate the search region with bound b^2 , uniform directions badly fails half of the coordinate tests. Taken together, these results demonstrate that optimal direction choice accelerates convergence of Hit-and-Run to a uniform target distribution on regions whose geometry makes clear distinctions between search directions.

We can gain further insight by seeking to establish a connection between the test results and the convergence rate bound values which apply to the regions tested. Theorem 7 and Table 1 indicate that the convergence rate bound for Hit-and-Run with optimal direction choice is the same for any upper bound vector. That is, the convergence bound analysis suggests that Hit-and-Run with optimal direction choice should perform the same in any rectangular region. Revisiting Tables 2 and 3, we see that this predicted behavior does

Table 2: Frequency test statistics for Hit-and-Run in rectangular regions.^a

| Coordinate i | χ^2 Statistics ^b (i th coord. range $(0, b_i)$ broken into 10 equal cells) | | | | | |
|---|---|---------|---------|---------|---------|---------|
| | Upper bound vector/Direction choice method | | | | | |
| | b^0 | | b^1 | | b^2 | |
| | Uniform | Optimal | Uniform | Optimal | Uniform | Optimal |
| 1 | 5.7 • | 10.6 • | 9.3 • | 11.3 • | 6.6 • | 8.0 • |
| 2 | 10.9 • | 6.6 • | 2.5 | 8.0 • | 8.7 • | 26.4 |
| 3 | 8.3 • | 10.7 • | 13.9 • | 14.1 • | 30.8 | 12.4 • |
| 4 | 8.1 • | 7.2 • | 5.6 • | 12.3 • | 16.8 • | 4.0 • |
| 5 | 10.4 • | 17.6 | 20.9 | 13.0 • | 27.6 | 10.0 • |
| 6 | 22.0 | 20.8 | 50.4 | 14.3 • | 498.8 | 5.3 • |
| 7 | 19.5 | 7.5 • | 19.3 | 18.2 | 843.2 | 18.1 |
| 8 | 8.6 • | 15.8 • | 13.0 • | 4.8 • | 563.0 | 15.5 • |
| 9 | 9.7 • | 16.6 • | 27.3 | 7.7 • | 401.2 | 7.1 • |
| 10 | 16.0 • | 9.4 • | 44.0 | 8.5 • | 773.1 | 7.6 • |
| # passing uniformity at $\alpha = 10\%$ | 8 | 8 | 4 | 9 | 3 | 8 |

^aStatistics passing the frequency test are marked by •.

^bUpper and lower χ^2 values for $\alpha = 10\%$, $\nu = 9$: (3.3, 16.9).

occur. The numbers of frequency tests passing for the three regions tested are 8, 9, and 8 respectively; the difference is not statistically significant. The differences among the number of serial tests passing (10, 7, and 10 respectively) might be considered significant, but two of the three tests which failed for b^1 did so just barely; a slightly lower test significance would have made the serial results as stable as the frequency results. The correspondence of stability between convergence rate bounds and experimental performance for the differing test regions provides empirical justification for the approach of accelerating convergence by optimizing the worst-case performance bound.

8 An adaptive direction choice rule

For the purpose of evaluating the effects of optimal direction choice in section 7, we were able to efficiently generate optimally distributed directions by applying a simple transformation to points drawn from the uniform distribution over the rectangular region S , as stated in Theorem 6. However, this method has no practical value; if a sample from the target distribution over S is readily available, there is no need for further application of Hit-and-Run or any other sampling method.

The convergence of Hit-and-Run to its uniform target distribution leads us to consider a heuristic approach. In keeping with Theorem 6, let us try choosing directions by randomly selecting one of the previous Hit-and-Run iterates and normalizing the vector from an ap-

Table 3: Serial correlation test statistics for Hit-and-Run in rectangular regions.^a

| Coordinate i | χ_{99}^2 Statistics ^b | | | | | |
|---|--|---------|---------|---------|---------|---------|
| | Upper bound vector/Direction choice method | | | | | |
| | b^0 | | b^1 | | b^2 | |
| | Uniform | Optimal | Uniform | Optimal | Uniform | Optimal |
| 1 | 98.8 • | 79.6 • | 89.6 • | 107.6 • | 98.4 • | 94.4 • |
| 2 | 83.6 • | 92.4 • | 89.2 • | 102.4 • | 126.0 | 105.2 • |
| 3 | 101.2 • | 92.8 • | 116.4 • | 117.2 • | 123.2 • | 108.0 • |
| 4 | 100.4 • | 101.6 • | 87.6 • | 124.4 | 101.2 • | 77.2 • |
| 5 | 110.8 • | 99.6 • | 113.6 • | 103.6 • | 111.6 • | 91.6 • |
| 6 | 136.8 | 115.6 • | 140.8 | 101.6 • | 693.2 | 95.2 • |
| 7 | 124.4 | 88.8 • | 106.8 • | 134.4 | 1223.6 | 96.8 • |
| 8 | 88.4 • | 105.6 • | 105.2 • | 80.0 • | 836.4 | 101.2 • |
| 9 | 115.2 • | 88.0 • | 97.2 • | 102.8 • | 574.0 | 103.6 • |
| 10 | 79.6 • | 107.2 • | 127.2 | 124.4 | 1064.4 | 106.4 • |
| # passing uniformity at $\alpha = 10\%$ | 8 | 10 | 8 | 7 | 4 | 10 |

^aStatistics passing the serial correlation test are marked by •.

^bUpper and lower χ^2 values for $\alpha = 10\%$, $\nu = 99$: (77.0, 123.2).

proximate center to the iterate. A simple scheme for approximating the center is to take the i th coordinate to be the mean of the i th coordinates of the previous Hit-and-Run iterates. We call this algorithm Artificial Centering Hit-and-Run (ACHR), formally stated as follows:

Artificial Centering Hit-and-Run

- 1) Choose an arbitrary starting point $x^0 \in S$ and a number of *warmup* sample points $W \geq n$. Set $m = 0$ and $\hat{s} = x^0$.
- 2) If $m < W$ (warmup phase), generate a random direction u^m according to the uniform distribution on ∂B . Otherwise (main phase), select a number a from the uniform distribution on $\{0, 1, \dots, m\}$ and set $u^m = (x^a - \hat{s})/\|x^a - \hat{s}\|$.
- 3) Select the step size λ_m as in the prior definition of Hit-and-Run.
- 4) Set $x^{m+1} = x^m + \lambda_m u^m$ and $m = m + 1$.
- 5) Set $\hat{s} = (m\hat{s} + x^m)/(m + 1)$. Go to step 2.

The algorithm first goes through a warmup phase, using uniform (i.e., HD) direction choice, in order to generate an initial sample. The requirement $W \geq n$ ensures that after the warmup phase, the set of directions $\{(x^a - \hat{s})/\|x^a - \hat{s}\| \mid a = 0, \dots, W\}$ spans ∂B with probability one. Then in the main phase, the search direction is chosen as the vector from the experimental

center \hat{s} to a randomly selected previous iterate. When the warmup period is short enough to be attractive with respect to computational effort, the warmup sample points may be generated with too few iterations to be nearly uniformly distributed, so we will discard them from the sample.

Obviously, the sequence of ACHR iterates is not a Markov chain, due to the dependence of directions on prior iterates and directions. Thus the convergence theorems from the literature do not apply, and hence it is not assured that the sequence of iterates converges in distribution to the target distribution.

8.1 Numerical results: centered convex case

We applied ACHR to the rectangular sampling problems, generating $W = 100$ warmup sample points. Recall that the uniform target distribution over a hyperrectangle is a case covered by Theorem 6, which motivated the ACHR algorithm.

As before, we iterate ten times between each sample point. The results of frequency and serial correlation tests are shown in Table 4. Comparing to the results for optimal-direction Hit-and-Run in Tables 2 and 3, *we find little or no degradation in performance when using ACHR*. Thus, we have implemented approximate optimal direction choice yielding performance on a par with exact optimal direction choice while keeping computational effort roughly equivalent to uniform direction choice. The extra effort is simply in the warmup iterations; for a 1000-point sample, the 100 warmup sample points simply require 10% more iterations. Furthermore, the added effort for ACHR is less than W/N because ACHR iterations are cheaper. Main-phase ACHR direction choice requires only one uniformly distributed number from $[0, 1)$, while HD direction choice requires n samples from the univariate standard normal distribution.

8.2 Numerical results: a noncentered case

Although ACHR succeeded in accelerating convergence on the rectangular sampling problems considered above, a strong question remains as to whether adaptive rules of this kind may have application in more general settings. To investigate this question, we now look at applying ACHR to a problem outside the class of centered search regions.

Let S be a simplex defined by

$$S = \{x \in \mathbb{R}^n | x_i > 0, \sum_{i=1}^n x_i/b_i < 1\}$$

given an upper-bounding vector $b > 0$ in \mathbb{R}^n . Let the target distribution G be uniform.

We will again compare uniform, optimal, and adaptive direction choice by simulating each method in the simplex S . Just as in our experiment with rectangular regions, we will generate optimally distributed directions by a method which requires points uniformly

Table 4: Test statistics for ACHR in rectangular regions.^a

| Coordinate i | Freq. χ^2 Statistics ^b | | | Ser. corr. χ^2 Statistics ^c | | |
|---|--|--------|--------|---|---------|---------|
| | Upper bound vector | | | b^0 | b^1 | b^2 |
| 1 | 28.0 | 8.7 • | 16.5 • | 142.8 | 100.0 • | 110.4 • |
| 2 | 36.6 | 25.9 | 9.3 • | 138.4 | 109.2 • | 84.8 • |
| 3 | 13.9 • | 2.7 | 14.6 • | 91.6 • | 98.0 • | 99.2 • |
| 4 | 14.0 • | 41.4 | 7.6 • | 108.8 • | 124.8 | 104.0 • |
| 5 | 9.1 • | 8.4 • | 5.5 • | 104.8 • | 71.2 | 110.8 • |
| 6 | 11.7 • | 15.9 • | 13.1 • | 98.8 • | 103.2 • | 122.0 • |
| 7 | 6.7 • | 9.7 • | 12.9 • | 85.2 • | 117.2 • | 82.0 • |
| 8 | 10.3 • | 12.7 • | 8.0 • | 102.0 • | 98.0 • | 90.8 • |
| 9 | 22.8 | 15.7 • | 31.8 | 98.4 • | 100.8 • | 134.0 |
| 10 | 9.1 • | 8.7 • | 9.5 • | 89.6 • | 92.8 • | 86.4 • |
| # passing uniformity at $\alpha = 10\%$ | 7 | 7 | 9 | 8 | 8 | 9 |

^aStatistics passing tests are marked by •.

^bUpper and lower χ^2 values for $\alpha = 10\%$, $\nu = 9$: (3.3, 16.9).

^cUpper and lower χ^2 values for $\alpha = 10\%$, $\nu = 99$: (77.0, 123.2).

distributed over S . Since S is noncentered, however, the simple transformation of Theorem 6 does not apply. Compared to the relative simplicity of that theorem, the derivation of the technique for generating optimal directions in the simplex (which derivation we consign to Appendix A) illustrates the potential complexity of general optimal direction distributions, and the need for practical adaptive rules.

The frequency test statistics for uniform and optimal direction choice and ACHR are shown in Table 5, for each of the three upper bound vectors $b^0 = (1, 1, 1, \dots, 1)$, $b^1 = (1, 2, 3, \dots, 10)$, and $b^2 = (1, 4, 9, \dots, 100)$. We generate 1000-point samples taking every twentieth iterate, with 100 warmup points for the ACHR adaptive algorithm. With the easiest upper-bound vector b^0 , uniform directions unexpectedly outperform optimal directions, illustrating that the optimality is with respect to a bound on the convergence rate, rather than an actual convergence rate. However, by comparing performance on the more difficult regions we see that optimal direction choice is far more robust.

Remarkably, the adaptive direction rule is both robust and also much more successful than the worst-case optimal rule, even though its motivating principle, asymptotic equivalence to the optimal direction distribution as stated in Theorem 6, does not apply to the simplex because that region is noncentered. Our tests have been restricted to regions which can be handled more efficiently by transformation techniques, because in these regions we are able to efficiently implement exact optimal direction choice for purposes of comparison. The success of adaptive direction choice on a noncentered region suggests that the ACHR algorithm

Table 5: Frequency test statistics for Hit-and-Run in simplices^a

| Coord. <i>i</i> | χ_9^2 Statistics ^b | | | | | | | | |
|---------------------------------------|--|--------|--------|-------------------|--------|--------|-------------------|--------|--------|
| | Upper bound vector/Direction choice method | | | | | | | | |
| | \underline{b}^0 | | | \underline{b}^1 | | | \underline{b}^2 | | |
| | Unif. | Opt. | Adapt. | Unif. | Opt. | Adapt. | Unif. | Opt. | Adapt. |
| 1 | 21.7 | 15.7 • | 4.1 • | 4.5 • | 20.5 | 12.2 • | 27.1 | 29.4 | 8.4 • |
| 2 | 16.5 • | 10.4 • | 8.0 • | 3.0 | 11.3 • | 14.3 • | 21.3 | 21.8 | 9.8 • |
| 3 | 30.6 | 11.9 • | 9.3 • | 15.3 • | 10.6 • | 6.8 • | 14.8 • | 29.6 | 5.5 • |
| 4 | 11.7 • | 19.9 | 22.5 | 25.2 | 25.7 | 9.2 • | 99.8 | 13.2 • | 10.1 • |
| 5 | 14.1 • | 9.5 • | 24.5 | 11.9 • | 7.8 • | 5.8 • | 93.1 | 9.0 • | 12.7 • |
| 6 | 24.1 | 27.1 | 18.3 | 12.0 • | 17.2 • | 13.8 • | 89.6 | 10.9 • | 24.3 |
| 7 | 12.8 • | 28.3 | 4.0 • | 64.5 | 20.5 | 6.0 • | 107.7 | 16.3 • | 9.2 • |
| 8 | 16.0 • | 12.4 • | 4.9 • | 20.6 | 18.0 | 10.3 • | 302.5 | 28.1 | 12.4 • |
| 9 | 6.5 • | 50.4 | 15.9 • | 85.3 | 44.5 | 12.7 • | 541.6 | 10.8 • | 7.1 • |
| 10 | 14.9 • | 32.7 | 15.2 • | 68.3 | 29.8 | 10.0 • | 274.8 | 24.1 | 19.5 |
| # pass- -ing at $\alpha = 10\%$ | 7 | 5 | 7 | 4 | 4 | 10 | 1 | 5 | 8 |

^aStatistics passing the frequency test are marked by •.

^bUpper and lower χ^2 values for $\alpha = 10\%$, $\nu = 9$: (3.3, 16.9).

has potential application to general sampling problems for which rejection is the only exact sampling approach.

9 Conclusion

The problem of generating points according to a probability distribution G with density g over an open bounded region S in \mathfrak{R}^n has application to Monte Carlo methods of simulation, numerical methods, and optimization. Even for simple regions S and a uniform distribution G , the computational effort of rejection techniques for exact sampling grows rapidly in dimension. When S and G are complicated, the performance is much worse.

Hit-and-Run algorithms offer an efficient method for generating points which asymptotically approach the desired probability distribution G . However, experimental and worst-case performance of HD Hit-and-Run is substantially degraded as S becomes nonspherical and G nonuniform. In this paper we have generalized the Hit-and-Run convergence rate bound, known previously for HD, to nonuniform direction distributions. We have constructed a unique bound-optimal direction distribution which significantly accelerates convergence to the target distribution, to a degree consistent with the corresponding improvement in convergence rate bound. Although exact implementation of this optimal direction choice may be very difficult, the ACHR heuristic algorithm suggests the possibility of implementing adaptive direction choice rules which automatically approximate optimal direction choice. The

success of ACHR on the simple problems considered here should motivate formulation and testing of adaptive direction choice for more general settings, with an eye towards seeking adaptive methods for which convergence to the target distribution can be guaranteed.

References

- [1] Bélisle, C.J.P., H.E. Romeijn and R.L. Smith. 1993. Hit-and-Run Algorithms for Generating Multivariate Distributions. *Mathematics of Operations Research* **18** 255–266.
- [2] Berbee, H.C.P., C.G.E. Boender, A.H.G. Rinnooy Kan, C.L. Scheffer, R.L. Smith, and J. Telgen. 1987. Hit-and-Run Algorithms for the Identification of Nonredundant Linear Inequalities. *Mathematical Programming* **37**, 184–207.
- [3] Billingsley, P. 1986. *Probability and Measure*. John Wiley & Sons, New York.
- [4] Boneh, A., and A. Golan. 1979. Constraints' Redundancy and Feasible Region Boundedness by Random Feasible Point Generator (RFPG). Presented at the Third European Congress on Operations Research (EURO III), Amsterdam.
- [5] Brooks, S.H. 1958. A Discussion of Random Methods for Seeking Maxima. *Operations Research* **6**, 244–251.
- [6] Chen, M. and B. Schmeiser. 1993. Performance of the Gibbs, Hit-and-Run and Metropolis Samplers. *Journal of Computational and Graphical Statistics* **2**, 251–272.
- [7] Dixon, L.C.W. and G.P. Szegö (eds.). 1975. *Towards Global Optimization*. North-Holland, Amsterdam.
- [8] Dixon, L.C.W. and G.P. Szegö (eds.). 1978. *Towards Global Optimization 2*. North-Holland, Amsterdam.
- [9] Doob, J.L. 1953. *Stochastic Processes*. John Wiley & Sons, New York.
- [10] Gelfand, A.E. and A.F.M. Smith. 1990. Sampling-Based Approaches to Calculating Marginal Densities. *J. American Statistical Association* **85** 398–409.
- [11] Geman, S. and D. Geman. 1984. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6** 721–741.
- [12] Geyer, C.J. 1992. Practical Markov Chain Monte Carlo. *Statistical Science* **7** 4 473–483.
- [13] Hammersley, J.M. and D.C. Handscomb. 1964. *Monte Carlo Methods*. Methuen & Co. Ltd., London.
- [14] Hastings, W.K. 1970. Monte Carlo Sampling Methods Using Markov Chains and their Applications. *Biometrika* **57** 97–109.
- [15] Helms, L.L. 1969. *Introduction to Potential Theory*. John Wiley & Sons, New York.
- [16] Karwan, M.H., V. Lotfi, J. Telgen, and S. Zionts (eds.). 1983. *Redundancy in Mathematical Programming*. Springer-Verlag, Berlin.

- [17] Knuth, D.E. 1981. *The Art of Computer Programming, Second Edition*, Vol. 2. Addison-Wesley, Reading, Massachusetts.
- [18] Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller. 1953. Equations of state calculations by fast computing machines. *The Journal of Chemical Physics* **21**, 1087–1092.
- [19] Patel, N.R., R.L. Smith, and Z.B. Zabinsky. 1988. Pure Adaptive Search in Monte Carlo Optimization. *Mathematical Programming* **43**, 317–328.
- [20] Rinnooy Kan, A.H.G. and G.T. Timmer. 1987. Stochastic Global Optimization Methods — Part I: Clustering Methods. *Mathematical Programming* **39**, 27–56.
- [21] Rubinstein, R.Y. 1981. *Simulation and the Monte Carlo Method*. John Wiley & Sons, New York.
- [22] Schmeiser, B.W. 1981. Random Variate Generation: A Survey. In T.I. Oren, C.M. Shub and P.F. Roth (eds.), *Simulation with Discrete Models: A State of the Art View*, 79–104. IEEE, New York.
- [23] Schmeiser, B.W. and M. Chen. 1991. On Hit-and-Run Monte Carlo Sampling for Evaluating Multidimensional Integrals. Report SMS91-1, Department of Statistics, Purdue University.
- [24] Solis, F.J. and R. J.-B. Wets. Minimization by Random Search Techniques. 1981. *Mathematics of Operations Research* **6**, 19–30.
- [25] Sommerville, D.M.Y. 1958. *An Introduction to the Geometry of N Dimensions*. Dover Publications, Inc., New York.
- [26] Smith, R.L. 1980. Monte Carlo Techniques for Generating Random Feasible Solutions to Mathematical Programs. Presented at the ORSA/TIMS Conference, Washington D.C.
- [27] Smith, R.L. 1984. Efficient Monte Carlo Procedures for Generating Points Uniformly Distributed over Bounded Regions. *Operations Research* **32**, 1296–1308.
- [28] Tanner, M.A. and W.H. Wong. 1987. The Calculation of Posterior Distributions by Data Augmentation. *J. American Statistical Association* **82** 528–541.
- [29] Telgen, J. 1980. Private communication with A. Boneh.
- [30] Zabinsky, Z.B. and R.L. Smith. 1992. Pure Adaptive Search in Global Optimization. *Mathematical Programming* **53** 323–338.
- [31] Zabinsky, Z.B., R.L. Smith, J.F. McDonald, H.E. Romeijn, and D.E. Kaufman. 1993. Improving Hit and Run for Global Optimization. *Journal of Global Optimization* **3** 171–192.

Appendix A Optimal direction choice for uniform sampling in a simplex

Let S be a simplex defined by

$$S = \{x \in \mathfrak{R}^n | x > 0, \sum_{i=1}^n x_i/b_i < 1\}$$

given an upper-bounding vector $b > 0$ in \mathfrak{R}^n . Let the target distribution G be uniform.

By Corollary 5, the optimal direction density h^* is proportional to the span function of the problem, which we now derive. Let $D_1 = \{x \in \mathfrak{R}^n | x_i > 0, i = 1, \dots, n\}$ be the first orthant of \mathfrak{R}^n .

Proposition 1 *The span function for the uniform target distribution on S is given by*

$$\rho(u) = \begin{cases} \left(\sum_{i=1}^n \frac{u_i}{b_i} \right)^{-n} & u \in \pm D_1 \\ \left(\min_{\substack{i=1, \dots, n \\ u_i \neq 0}} \left\{ \frac{b_i}{|u_i|} \right\} \right)^n & u \notin \pm D_1. \end{cases}$$

Proof: Recall from equation (7) that $\rho(u) = \sup\{\|y - x\|^n | x, y \in S, u_{xy} = u\}$ since S is convex. Thus $\rho(u)$ is determined by the greatest distance across S in direction u .

If $u \in \pm D_1$, then one endpoint of any segment across S in direction u must lie on the facet bounding S above; see Figure 1(a). Call the other endpoint y and let α be the distance between them. Then $\sum_{i=1}^n (y_i + \alpha u_i)/b_i = 1$. Solving the expression for α , we see that the distance is maximized by $y = 0$, confirming the proposition for $u \in \pm D_1$.

If $u \notin \pm D_1$, it is evident pictorially that one endpoint of the longest segment must lie in a corner other than the origin, and that the same segment would maximize the distance across the enclosing rectangle completed in dashed lines in Figure 1(b). Thus the span for such directions is equal to its value on the rectangle of Example 2. ■

Propositions 2 and 3 give equivalences for the optimal direction distribution H^* conditioned on whether or not the search direction lies in $\pm D_1$. Proposition 4 then gives the probability that an optimally distributed direction lies in $\pm D_1$, allowing sampling from H^* by a composition technique.

Proposition 2 *Let U be distributed according to H^* on ∂B and let Y be uniformly distributed on the simplex S . Then $Y/\|Y\|$ is distributed identically to U given $U \in \pm D_1$.*

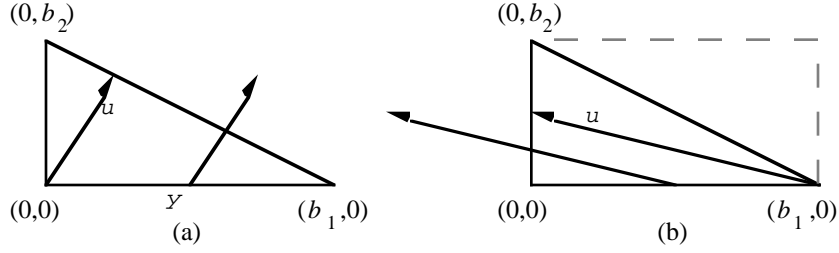


Figure 1: Spans of the simplex

Proof: For measurable $D \subset D_1$, let $S_0(D)$ be the part of S in the directions D from 0, as defined in the proof of Theorem 6. Also, let $r_0(u)$ be the distance in direction u from 0 to the facet bounding S above. Then

$$\begin{aligned} \Pr(Y/\|Y\| \in D) &= \Pr(Y \in S_0(D)) \\ &\propto \int_D r_0^n(u) \sigma(du) && \text{by change to spherical coordinates} \\ &= \int_D \rho(u) \sigma(du), \end{aligned}$$

the last equality being evident from the proof of Proposition 1. Thus $Y/\|Y\|$ has probability density proportional to h^* over the first orthant. ■

Let $S' = \{x \in \mathfrak{R}^n \mid 0 < x_i < b_i, i = 1, \dots, n\}$ be the hyperrectangle enclosing S in the obvious fashion, as shown in dashed lines in Figure 1(b), and note that S' has center $s = b/2$.

Proposition 3 *Let U be distributed according to H^* on ∂B and let Y be uniformly distributed on S' . Then U and $(Y - s)/\|(Y - s)\|$ have the same distribution, given that neither lies in $\pm D_1$.*

Proof: Let $r(u)$ be the distance in direction u from s to $\partial S'$. For measurable $D \subset (\partial B - \pm D_1)$,

$$\Pr\left(\frac{Y - s}{\|Y - s\|} \in D\right) \propto \int_D r^n(u) \sigma(du) \tag{10}$$

$$= \int_D \frac{1}{2} \rho(u) \sigma(du) \tag{11}$$

where the factor of $1/2$ occurs because now the distance is from s to the boundary rather than from boundary to boundary. Thus, $(Y - s)/\|(Y - s)\|$ has density proportional to h^* outside the first orthant. ■

Proposition 4 *Let U be distributed according to H^* over S . Then*

$$\Pr(U \in \pm D_1) = \frac{1}{1 + n!(2^{n-1} - 1)}.$$

Proof: Let $R = \int_{\partial B} \rho(u) \sigma(du)$ and $p = \Pr(U \in \pm D_1)$. Then

$$\begin{aligned} p &= (2/R) \int_{D_1} \rho(u) \sigma(du) \\ &= (2n/R) \int_{D_1} (1/n)r_0(u) \sigma(du) \\ &= (2n/R)V(S) \quad \text{by change to spherical coordinates} \end{aligned}$$

and

$$\begin{aligned} 1 - p &= (1/R) \int_{\partial B - \pm D_1} \rho(u) \sigma(du) \\ &= \frac{n 2^n - 2}{R 2^n} \int_{\partial B} (1/n)(2r(u))^n \sigma(du) \\ &= 2n(2^{n-1} - 1)V(S')/R. \end{aligned}$$

Since $V(S')/V(S) = n!$ (see e.g. equation (VIII.4.4) of [25]), we have $(1 - p)/p = n!(2^{n-1} - 1)$, which yields the desired result when solved for p . ■

We can therefore generate optimally distributed directions by a composition technique, selecting a random number from $[0, 1)$ to determine whether or not the direction will lie in $\pm D_1$. If so, we generate a direction by sampling uniformly from S as in Proposition 2. (This is easy, since the marginal distribution function of the first coordinate Y_1 of a point uniformly distributed in S is $1 - (1 - y_1)^n$, the conditional distribution function of Y_2 given $Y_1 = y_1$ is $y_1(1 - (1 - y_2)^{n-1})$, and so on.) Otherwise, we generate a direction by sampling uniformly from S' as specified in Proposition 3, rejecting directions which search the first orthant.