

Pattern Discrete and Mixed Hit-and-Run for Global Optimization

Huseyin Onur Mete · Yanfang Shen · Zelda B. Zabinsky · Seksan Kiatsupaibul · Robert L. Smith

Received: date / Accepted: date

Abstract We develop new Markov chain Monte Carlo samplers for neighborhood generation in global optimization algorithms based on Hit-and-Run. The success of Hit-and-Run as a sampler on continuous domains motivated Discrete Hit-and-Run with random biwalk for discrete domains. However, the potential in efficiencies in the implementation, which requires a randomization at each move to create the biwalk, lead us to a different approach that uses fixed *patterns* in generating the biwalks. We define Sphere and Box Biwalks that are pattern-based and easily implemented for discrete and mixed continuous/discrete domains. The pattern-based Hit-and-Run Markov chains preserve the convergence properties of Hit-and-Run to a target distribution. They also converge to continuous Hit-and-Run as the mesh of the discretized variables becomes finer, approaching a continuum. Moreover, we provide bounds on the finite time performance for the discrete cases of Sphere and Box Biwalks. We embed our samplers in an Improving Hit-and-Run global optimization algorithm and test their performance on a number of global optimization test problems.

Keywords Global optimization · Simulated annealing · Markov chain Monte Carlo sampling · Stochastic optimization · Adaptive search algorithms · Improving Hit-and-Run

Mathematics Subject Classification (2000) 90-08 · 90C06 · 90C59 · 68W20 · 68W40

Huseyin Onur Mete
Industrial Engineering, University of Washington, Seattle, WA 98195-2650, USA
E-mail: mete@u.washington.edu

Yanfang Shen
Citigroup Alternative Investments, New York, NY 10022, USA
E-mail: yanfang.shen@citi.com

Zelda B. Zabinsky
Industrial Engineering, University of Washington, Seattle, WA 98195-2650, USA
E-mail: zelda@u.washington.edu

Seksan Kiatsupaibul
Department of Statistics, Chulalongkorn University, Bangkok, 10330, Thailand
E-mail: seksan.kiatsupaibul@gmail.com

Robert L. Smith
Dept. of Industrial and Operations Engineering, Univ. of Michigan, Ann Arbor, MI 48109-2117, USA
E-mail: rlsmith@umich.edu

1 Introduction

We develop Markov chain Monte Carlo samplers to be used in adaptive search algorithms for global optimization problems in both discrete and mixed continuous/discrete domains. The ideal version of the Simulated Annealing algorithm requires a neighborhood generation mechanism that samples according to a sequence of Boltzmann distributions [23, 26]. For specific combinatorial problems, such as the Traveling Salesman Problem, specialized neighborhood generators have been developed to be used in random sampling in optimization algorithms [3, 10, 27]. However, for mixed continuous/discrete domains, which are prevalent in many engineering optimization problems [12, 18, 22, 28], appropriate neighborhood generators have not been developed. Many effective global and convex optimization algorithms for continuous problems have embedded the Markov chain Monte Carlo sampler known as Hit-and-Run (HR) [7, 11, 19, 20, 29, 30]. To sample a point in a set, HR generates a direction vector uniformly on a hypersphere, then samples a point from the intersection of the line along the generated direction and the sampled set [25]. A new discrete version of HR, called Discrete Hit-and-Run (DHR), creates a bidirectional random walk (biwalk) that is analogous to the random direction used in HR [4]. While DHR preserves asymptotic convergence properties to target (e.g. Boltzmann) distributions like continuous HR, the construction in [4] requires a lot of computation. We present efficient versions of DHR using pattern Sphere and Box Biwalks that maintain the theoretical properties of DHR, such as convergence to a uniform distribution in some cases in polynomial time, but are computationally much more efficient to implement. In addition, we extend Sphere and Box Biwalks as candidate point generators for mixed continuous/discrete domains. We prove that Sphere Biwalk converges to the continuous direction generated by Hit-and-Run as the discrete mesh of the mixed domain becomes finer and approaches a continuous domain. Similarly, Box Biwalk converges to Hit-and-Run with a modified continuous direction distribution.

We embed the new samplers in a mixed continuous/discrete version of Improving Hit-and-Run (IHR) [30], which is a type of Simulated Annealing algorithm that only accepts improving points instead of involving a cooling schedule. We use IHR as an optimization framework to avoid the effect of cooling schedules on the algorithm performance and allows us to contrast the effect of the neighborhood generators. We present numerical results comparing IHR with Sphere and Box Biwalks to IHR employing other candidate point generators.

The new Markov chain Monte Carlo (MCMC) samplers we propose for mixed continuous/discrete domains intend to mimic the good performance of continuous HR. Smith [25] proved convergence of HR to a uniform distribution in total variation, and Lovász [13] proved the rate of convergence to uniform is $O(n^3)$ with a warm start assumption. Lovász and Vempella extended the convergence results to log concave density functions [14] and relaxed the warm start assumption [15]. The HR algorithm with a Metropolis filter generates sample points that converge to an arbitrary multivariate distribution [6, 19], and Andersen and Diaconis [2] proposed a generalization of HR algorithms for MCMC samplers.

Discrete Hit-and-Run samples from a discrete set S , which is a subset of an integer lattice of a hyperrectangle H , by generating two independent nearest neighbor random walks from the current state of the Markov chain that stop when they step out of H [4]. The generated walks are called forward and backward walks (or paths) and we call the pair of walks a random biwalk. The ordered sequence of points visited by the biwalk are stored in a list. Then, a new state for the Markov chain is chosen uniformly from the ordered list of points that are also in S as the next candidate point. If a Metropolis filter is employed (for acceptance or rejection), then, as in HR, the limiting distribution of DHR with random biwalks is

the target distribution on the discrete set $S \subseteq H$. Thus, even for a set S with isolated points, DHR does not get trapped in isolated regions of S . Baumert et al. [4] also provide bounds on the finite time performance of DHR. For several special cases, the rate of convergence is $O(n^4)$.

In [4], DHR with random biwalks is computationally expensive to implement because it requires generating a random number for each point in the biwalk and storing the ordered list, which requires large memory capacity. Instead, we introduce *pattern* biwalk in which we create forward and backward paths, using a randomly generated pattern. The pattern biwalk does not allow loops where the random biwalk does allow loops, but pattern biwalk is more closely analogous to the random direction of HR. The pattern is formed by generating a step size for each dimension and a random order of dimensions. Therefore, this requires $2n - 1$ random number generations on the integer lattice of an n -dimensional hyperrectangle H , in contrast to an average number $O(nL^2)$ for DHR with random biwalks, where L is the length of the longest side of H . We also use the pattern to map a uniform random number to a point on the biwalk, instead of explicitly generating and storing the entire ordered list. In order to generate the discrete step sizes for a pattern, we use two different methods called Sphere and Box Biwalks. In Sphere Biwalk, we choose a uniform direction on an n -dimensional hypersphere and generate a random point on the diameter defined by this direction. The random point is rounded to the nearest integer lattice point on discrete dimensions (continuous dimensions allow real-valued points), to define the step size for each coordinate. The step sizes, together with a random permutation of coordinate directions, define the pattern of the biwalk. Similarly, Box Biwalk generates a point in a hyperrectangle which is rounded to the nearest integer lattice point, to provide the vector of step sizes for the biwalk. Then, we conceptually use the pattern to create the forward path, starting from the current state of the Markov chain, until we step out of H and use the opposite pattern to create the backward path. Similar to DHR with random biwalks, we sample uniformly from the points that are both on the biwalk and the set S .

The Sphere and Box Biwalks are easily extended to sample from a mixed continuous/discrete domain, and thus are applicable to a wide variety of real-world optimization problems. An intuitive way to handle mixed continuous/discrete domains is to treat the continuous and discrete domains separately and apply continuous HR and DHR to the separated domains. Although this method is easy to implement, the approach may lose valuable information carried by the unseparated domain. As an alternative, Romeijn et al. [20] offers a Step Function approach for mixed continuous/discrete domains. Our numerical results show that IHR with pattern biwalks provides a more efficient optimization algorithm than using either a separate approach or the Step Function approach.

The proposed pattern biwalks for both discrete and mixed continuous/discrete domains preserve the property of converging to a target distribution with conditions on their parameters. Given any target distribution π on a mixed continuous/discrete S contained in a hyperrectangle H , both pattern Sphere and Box Biwalks generate Markov chains that converge in distribution to π when using a Metropolis filter, and the radius of the utilized sphere is sufficiently large, or the lengths of the utilized box are sufficiently large.

Generating pattern biwalks are analogous to the use of random direction in HR. We prove that as the mesh of the lattice used in Sphere Biwalk gets finer, the stochastic process of the points generated with pattern Sphere Box Biwalk converges to the stochastic process of points generated by HR. In a similar way, under particular rules for choosing the lengths of the utilized box for Box Biwalk, the stochastic process of the points generated with the pattern Box Biwalk converges to the stochastic process of points generated by HR when the random direction is box generated. This provides Sphere and Box Biwalks the potential for

solving mixed continuous/discrete problems with similar performance as HR. In addition, we provide bounds on the finite time performance of the pattern Sphere and Box Biwalks on special discrete sets, which compare favorably with DHR with random biwalks.

The rest of the article is organized as follows. In Section 2, we provide the optimization setting of Improving Hit-and-Run and embed pattern biwalks for mixed continuous/discrete domains. We define our pattern Sphere and Box Biwalks in Section 3. Then, we provide the theorems and necessary conditions to show that these new MCMC samplers converge to a target distribution in Section 4. In Section 5, we show that they are mixed continuous/discrete analogs of continuous HR. We analyze the convergence rate of Sphere and Box Biwalk for discrete domains in Section 6. Finally, we provide the computational studies on the performance of the new candidate point generators on global optimization test problems in Section 7.

2 Pattern Mixed Hit-and-Run and Improving Hit-and-Run for Global Optimization

We develop our MCMC samplers to be used in global optimization algorithms. We test our approach using a discrete and mixed continuous/discrete version of Simulated Annealing, that is, with Improving Hit-and-Run [30]. An iteration of IHR starts with a random direction, and picks a candidate point on the line from the current point in this direction. Then, the candidate point is accepted if it provides a better objective function value than the current point gives. HR, as an MCMC sampler approximating a target distribution, generates a random direction and picks a point in the same way as IHR, however, instead of checking the improvement in the objective function value, HR applies a Metropolis filter with respect to the target distribution for acceptance/rejection. Thus, by providing a new method of picking a *direction* and a point on the associated *line* with pattern Sphere and Box Biwalks, we can extend IHR and HR to mixed continuous/discrete domains.

We consider the following optimization problem,

$$\begin{aligned} \min f(x) \\ \text{s.t. } x \in S \subseteq H, \end{aligned}$$

where H is a mixed continuous/discrete lattice of a hyperrectangle such that

$$H = \prod_{i=1, \dots, n} \begin{cases} [l_i, u_i] & \text{if } i \text{ is a continuous dimension} \\ \{l_i, l_i + 1, \dots, u_i\} & \text{if } i \text{ is a discrete dimension,} \end{cases}$$

provided that $l_i, u_i \in \mathbb{R}$, if i is a continuous dimension and $l_i, u_i \in \mathbb{Z}$, if i is a discrete dimension and $l_i < u_i$ for all $i = 1, \dots, n$.

An iteration for making a transition from a point x to y by the modified Improving Hit-and-Run algorithm or for Hit-and-Run as a sampler for the mixed continuous/discrete case with pattern-based biwalks is provided below.

Modified Improving Hit-and-Run and Hit-and-Run with Pattern Biwalks

- Step 0. Given a starting point $x \in S \subseteq H$.
- Step 1. Generate a pattern biwalk from x .
- Step 2. Generate a candidate point z by choosing a point uniformly distributed from the associated biwalk intersected with S .
- Step 3. Accept or reject the point z ;

- for IHR optimization,

$$y = \begin{cases} z & \text{if } f(z) < f(x) \\ x & \text{otherwise.} \end{cases}$$

- for HR with Metropolis filter for target distribution π ,

$$y = \begin{cases} z & \text{with probability } \min(1, \pi(z)/\pi(x)) \\ x & \text{otherwise.} \end{cases}$$

In Section 3, we describe how to execute Steps 1 and 2 in the description of the algorithms above, using the new pattern Sphere and Box Biwalks. In Section 4, we provide theoretical results showing that HR with the pattern biwalks converge to the target distribution that is used in the Metropolis filter given in Step 3. The accept/reject criterion for IHR optimization, given in Step 3, is used for solving the global optimization test problems in Section 7.

3 Pattern Biwalk

3.1 Sphere and Box Biwalks

We design Sphere Biwalk to be analogous to continuous Hit-and-Run [25]. We call this candidate point generator Sphere Biwalk, because as the mesh of the mixed lattice of the hyperrectangle becomes finer, the points on the biwalk generated by Sphere Biwalk converge to the direction generated by Hit-and-Run where the direction is chosen uniformly on a hypersphere. Box Biwalk differs only in the way the step-size pattern is chosen; uniformly on the interior of a box as opposed to using a hypersphere.

The first step of Sphere Biwalk is to generate n mixed continuous/discrete step length directions $\{D_1, \dots, D_n\}$, where each D_i is an n -dimensional vector and the only nonzero entry is the i^{th} entry, which is the length of step to be taken along the i^{th} coordinate axis. Note the i^{th} entry is restricted to be an integer when i is a discrete dimension, but real-valued otherwise. This set of directions is generated by first choosing a direction uniformly distributed on the surface of an n -dimensional hypersphere of radius R , $R \geq 1$, and picking a point uniformly distributed on the corresponding diameter, then rounding the point to the nearest mixed lattice point. This could be interpreted as an iteration of HR with rounding. The rounded mixed lattice point determines the step sizes in each coordinate direction. The second step of Sphere Biwalk generates a random permutation of the coordinate dimensions, which completes the pattern of the biwalk. The third step generates a forward path by repeatedly taking steps according to the mixed coordinate direction D_i in the chosen order of dimensions until stepping out of hyperrectangle H . The fourth step generates a backward path in the reverse direction. This constitutes the pattern biwalk in Step 1 of the Modified IHR and HR algorithm on a mixed continuous/discrete domain. We now define Sphere Biwalk with parameter R .

Sphere Biwalk

Step 0. Given a current point $x \in S \subseteq H$, and $R \in \mathbb{R}$, $R \geq 1$.

Step 1. Generate n coordinate step length directions $\{D_1, \dots, D_n\}$ as follows:

- 1.1. Generate a continuous direction $\tilde{D} = (\tilde{d}_1, \tilde{d}_2, \dots, \tilde{d}_n)$ uniformly distributed on the boundary of a unit hypersphere.

- 1.2. Generate a point uniformly on the line set $\tilde{L}_{\tilde{D}} = \{v : v = r\tilde{D}, r \in [-R, R]\}$, and round it to the nearest mixed lattice point, call it $V^{\tilde{D}} = (v_1^{\tilde{D}}, v_2^{\tilde{D}}, \dots, v_n^{\tilde{D}})$. If $v_i^{\tilde{D}} = 0$ for all $i = 1, 2, \dots, n$, go to Step 1.1.
- 1.3. For $i = 1, 2, \dots, n$, set $D_i = (0, \dots, 0, v_i^{\tilde{D}}, 0, \dots, 0)$ with $v_i^{\tilde{D}}$ at the i^{th} entry and 0 at the other entries.
- Step 2. Generate a random permutation of n coordinate dimensions, $\{I_1, \dots, I_n\}$, with $I_i \in \{1, \dots, n\}$ and $I_i \neq I_j$ for $i \neq j$ and $i, j = 1, \dots, n$, uniformly, i.e. each possible permutation has equal probability, $1/n!$.
- Step 3. Generate a forward path starting from x , providing the ordered list of points forming the forward path $\{W_0^x, W_1^x, \dots, W_f^x\}$, such that $W_0^x = x$, by repeatedly moving in the order of dimensions, $\{I_1, I_2, \dots, I_n\}$, with the associated step direction D_{I_i} for the I_i^{th} coordinate dimension until stepping out of H , as follows:
- 3.0. Let $W_0^x = x$. Set $j = 0$ and $t = 0$.
- 3.1. Update forward counter for coordinate dimension:

$$t = \begin{cases} 1 & \text{if } t = n \\ t + 1 & \text{otherwise.} \end{cases}$$

- 3.2. If $(W_j^x + D_{I_t}) \notin H$, go to Step 4; otherwise,
 if $D_{I_t} \neq 0$, let $W_{j+1}^x = W_j^x + D_{I_t}$, set $j = j + 1$ and go back to Step 3.1,
 if $D_{I_t} = 0$, go back to Step 3.1 directly.
- Step 4. Generate a backward path starting from x , providing the ordered list of points forming the backward path $\{W_{-b}^x, W_{-b+1}^x, \dots, W_{-1}^x, W_0^x\}$, such that $W_0^x = x$, by repeatedly moving in the reversed order of dimensions, $\{I_n, I_{n-1}, \dots, I_1\}$, with the negative associated step direction $-D_{I_i}$ for the I_i^{th} coordinate dimension until stepping out of H , as follows:
- 4.0. Set $j = 0$ and $t = n + 1$.
- 4.1. Update backward counter for coordinate dimension:

$$t = \begin{cases} 1 & \text{if } t = n \\ t - 1 & \text{otherwise.} \end{cases}$$

- 4.2. If $(W_j^x - D_{I_t}) \notin H$, stop; otherwise,
 if $D_{I_t} \neq 0$, let $W_{j-1}^x = W_j^x - D_{I_t}$, set $j = j - 1$ and go back to Step 4.1,
 if $D_{I_t} = 0$, go back to Step 4.1 directly.

Thus, the Sphere Biwalk results in an ordered list of points, $\{W_{-b}^x, W_{-b+1}^x, \dots, W_{-1}^x, W_0^x, W_1^x, \dots, W_f^x\}$ that are generated from x by the step length directions $\{D_1, \dots, D_n\}$ and permutation of coordinates $\{I_1, \dots, I_n\}$. In IHR and HR, the next step is to choose a point uniformly distributed from this biwalk, and an efficient implementation is discussed in Section 3.2.

We define the Box Biwalk in a manner similar to Sphere Biwalk. Instead of using a hypersphere to generate the continuous direction, we use an n -dimensional box, $[-c_1, c_1] \times \dots \times [-c_n, c_n]$, where $2c_i$ is the length of the i^{th} side of the box. Box Biwalk generates a point uniformly distributed in the box, and then rounds it to the nearest mixed lattice point. This rounded point determines the mixed continuous/discrete step length directions $\{D_1, \dots, D_n\}$. We provide Step 1 of Box Biwalk with parameters $c_i \in \mathbb{R}$, if i is a continuous dimension and $c_i \in \mathbb{Z}$, if i is a discrete dimension and $c_i \geq 1$ for all $i = 1, \dots, n$. The rest of the algorithm is the same as Sphere Biwalk.

Box Biwalk

Step 1. Generate n coordinate step length directions $\{D_1, \dots, D_n\}$ as follows:

- 1.1. Generate a continuous point uniformly distributed in the interior of the box $[-c_1, c_1] \times \dots \times [-c_n, c_n]$.
- 1.2. Round it to the nearest mixed lattice point, call it $V^D = (v_1^D, v_2^D, \dots, v_n^D)$. If $v_i^D = 0$ for all $i = 1, 2, \dots, n$, go to Step 1.1.
- 1.3. For $i = 1, 2, \dots, n$, set $D_i = (0, \dots, 0, v_i^D, 0, \dots, 0)$ with v_i^D at the i^{th} entry and 0 at the other entries.

Both pattern biwalks produce an ordered list of points, with the difference being the distribution used to generate the step length directions. Sphere Biwalk generates the step sizes by generating a continuous direction on the surface of a hypersphere and picking a point on the diameter (as one iteration of HR) before rounding, whereas Box Biwalk generates the random point in the interior of the box before rounding. A family of pattern biwalks could be defined by their step size distributions on a mixed continuous/discrete domain.

3.2 Implementation of Pattern Biwalk

We develop efficient ways of implementing Sphere and Box Biwalks embedded in HR and IHR. Because generating and maintaining explicit forward and backward paths has a high computational cost, we use the pattern to generate a candidate point z on the pattern biwalk, combining Steps 1 and 2 of the Modified IHR and HR, without explicitly generating the whole biwalk. Both Sphere and Box Biwalks result in an ordered list of points $\{W_{-b}^x, W_{-b+1}^x, \dots, W_{-1}^x, W_0^x, W_1^x, \dots, W_f^x\}$ that are generated from x by the step length directions $\{D_1, \dots, D_n\}$ and permutation of coordinates $\{I_1, \dots, I_n\}$. The next step is to generate a candidate point z uniformly distributed on the biwalk.

If all the dimensions are discrete, then a uniformly distributed point is generated by simply counting the number of points in the biwalk, generating a random number such that the points are all equally weighted, and mapping that random number to the corresponding point W_i^x . If all the dimensions are continuous, then the biwalk results in a sequence of intervals, with the left and right endpoints of the intervals defined by the ordered list of points $\{W_i^x\}$ and the two exit points from the hyperrectangle. In this case, a uniformly distributed point is generated by calculating the length of the biwalk, generating a random point on the length and mapping it to a corresponding point z in one of the intervals. When all the dimensions are continuous, then the chance of landing in a specific interval is weighted by the length of the interval with respect to the total length of the biwalk.

When the set S is mixed with several continuous dimensions and several discrete ones, then the biwalk consists of intermingled intervals $[W_i^x, W_{i+1}^x)$ and points W_i^x determined by the ordered list of points and whether the step direction is continuous or discrete. In this case, we generate a uniformly distributed point on the biwalk using a mixture of uniform densities on continuous intervals and point masses on discrete points.

We start by counting the number of points with discrete dimensions in the biwalk, the length of the continuous intervals in the biwalk, and by determining whether the two exit points of the biwalk are along a continuous dimension or a discrete dimension. We arbitrarily consider the points $\{W_i^x\}$ as the left-hand side of an interval, even on the backward walk, but we need to add a partial interval if the exiting dimension on the backward walk is continuous. In this way, if all the dimensions are discrete, our procedure is equivalent to a discrete

uniform distribution on the number of points, and if all the dimensions are continuous, then it is equivalent to a continuous uniform distribution on the total length of the biwalk.

We first use a modified minimum ratio test to count the number of steps that can be taken on dimension i through the forward path, μ_i^f , and the number of steps that can be taken on dimension i through the backward path, μ_i^b , irregardless of whether the dimension is discrete or continuous. For $i = 1, \dots, n$,

$$\mu_i^f = \begin{cases} \left\lfloor \frac{u_i - x_i}{v_i^{\tilde{D}}} \right\rfloor & \text{if } v_i^{\tilde{D}} > 0 \\ \left\lceil \frac{x_i - l_i}{|v_i^{\tilde{D}}|} \right\rceil & \text{if } v_i^{\tilde{D}} < 0 \end{cases} \quad \text{and} \quad \mu_i^b = \begin{cases} \left\lfloor \frac{x_i - l_i}{v_i^{\tilde{D}}} \right\rfloor & \text{if } v_i^{\tilde{D}} > 0 \\ \left\lceil \frac{u_i - x_i}{|v_i^{\tilde{D}}|} \right\rceil & \text{if } v_i^{\tilde{D}} < 0. \end{cases}$$

Then, for forward and backward paths, the number of complete cycles of all dimensions before leaving H , μ^f and μ^b respectively, are defined as follows:

$$\mu^f = \min_{i=1, \dots, n} \mu_i^f \quad \text{and} \quad \mu^b = \min_{i=1, \dots, n} \mu_i^b.$$

We also need to define \tilde{i}^f , the order of the dimension that the forward path steps out of H ,

$$\tilde{i}^f = \min_{i=1, \dots, n} \{i : \mu_i^f = \mu^f\}.$$

And then, the dimension that steps out of H on the forward path is $I_{\tilde{i}^f}$.

For the sake of clarity in the representation, let N be the set of discrete dimensions with non-zero $v_i^{\tilde{D}}$ and \tilde{N} be the set of continuous dimensions with non-zero $v_i^{\tilde{D}}$.

If all the dimensions are discrete, i.e., $|N| = n$, then the number of points on the forward path is the sum of two parts: the number of points on the complete cycles of all dimensions, which is $\mu^f n$; and the number of points on the remaining full dimension steps taken after the complete cycles before stepping out of H , which is $\tilde{i}^f - 1$. Thus, the number of points on the forward path (not including W_0^x) is

$$t^f = n\mu^f + \tilde{i}^f - 1.$$

However if the dimensions are mixed continuous/discrete, then we calculate the length of the forward path as the sum of five parts:

- i.) the length of the complete cycles of continuous dimensions on the forward path, which is $\mu^f \sum_{I_i \in \tilde{N}} |v_{I_i}^{\tilde{D}}|$;
- ii.) the number of points on the complete cycles of discrete dimensions, which is $\mu^f |N|$;
- iii.) the length of remaining full continuous dimension steps taken after the complete cycles before stepping out of H , which is $\sum_{i < \tilde{i}^f, I_i \in \tilde{N}} |v_{I_i}^{\tilde{D}}|$;
- iv.) the number of points on the remaining full discrete dimension steps taken after the complete cycles before stepping out of H , which is $|\{i : i < \tilde{i}^f, I_i \in N\}|$; and
- v.) the length of the remaining portion of the last step that resides in H , if it is a continuous dimension.

If the last step is a continuous dimension and has positive direction, that is $v_{I_{\tilde{i}^f}}^{\tilde{D}} \geq 0$, the forward path leaves H through the upper bound of dimension $I_{\tilde{i}^f}$, thus the remaining length is $u_{I_{\tilde{i}^f}} - (x_{I_{\tilde{i}^f}} + \mu^f v_{I_{\tilde{i}^f}}^{\tilde{D}})$, otherwise, if $v_{I_{\tilde{i}^f}}^{\tilde{D}} < 0$ the path leaves H through the lower bound of $I_{\tilde{i}^f}$, and the remaining length is $(x_{I_{\tilde{i}^f}} + \mu^f v_{I_{\tilde{i}^f}}^{\tilde{D}}) - l_{I_{\tilde{i}^f}}$. If the dimension that steps out of H is

discrete, we do not need to add the remaining portion. Thus, the length of the forward path in the mixed case is

$$t^f = \begin{cases} \left\{ \begin{array}{l} \mu^f \sum_{I_i \in \tilde{N}} |v_{I_i}^{\tilde{D}}| + \mu^f |N| + \sum_{i < \tilde{i}^f, I_i \in \tilde{N}} |v_{I_i}^{\tilde{D}}| + |\{i : i < \tilde{i}^f, I_i \in N\}| \\ + u_{I_{\tilde{i}^f}} - (x_{I_{\tilde{i}^f}} + \mu^f v_{I_{\tilde{i}^f}}^{\tilde{D}}) \end{array} \right. & \text{if } I_{\tilde{i}^f} \in \tilde{N} \text{ and } v_{I_{\tilde{i}^f}}^{\tilde{D}} \geq 0 \\ \left\{ \begin{array}{l} \mu^f \sum_{I_i \in \tilde{N}} |v_{I_i}^{\tilde{D}}| + \mu^f |N| + \sum_{i < \tilde{i}^f, I_i \in \tilde{N}} |v_{I_i}^{\tilde{D}}| + |\{i : i < \tilde{i}^f, I_i \in N\}| \\ + (x_{I_{\tilde{i}^f}} + \mu^f v_{I_{\tilde{i}^f}}^{\tilde{D}}) - l_{I_{\tilde{i}^f}} \end{array} \right. & \text{if } I_{\tilde{i}^f} \in \tilde{N} \text{ and } v_{I_{\tilde{i}^f}}^{\tilde{D}} < 0 \\ \left\{ \begin{array}{l} \mu^f \sum_{I_i \in \tilde{N}} |v_{I_i}^{\tilde{D}}| + \mu^f |N| + \sum_{i < \tilde{i}^f, I_i \in \tilde{N}} |v_{I_i}^{\tilde{D}}| + |\{i : i < \tilde{i}^f, I_i \in N\}| \end{array} \right. & \text{if } I_{\tilde{i}^f} \in N. \end{cases}$$

The length of the backward path is similar, where \tilde{i}^b is the order of the dimension that the backward path steps out of H ,

$$\tilde{i}^b = \max_{i=1, \dots, n} \{i : \mu_{I_i}^b = \mu^b\},$$

and $I_{\tilde{i}^b}$ is the dimension that steps out of H on the backward path. The length of the backward path for the mixed case, t^b , is given as follows:

$$t^b = \begin{cases} \left\{ \begin{array}{l} \mu^b \sum_{I_i \in \tilde{N}} |v_{I_i}^{\tilde{D}}| + \mu^b |N| + \sum_{i > \tilde{i}^b, I_i \in \tilde{N}} |v_{I_i}^{\tilde{D}}| + |\{i : i > \tilde{i}^b, I_i \in N\}| \\ + (x_{I_{\tilde{i}^b}} + \mu^b v_{I_{\tilde{i}^b}}^{\tilde{D}}) - l_{I_{\tilde{i}^b}} \end{array} \right. & \text{if } I_{\tilde{i}^b} \in \tilde{N} \text{ and } v_{I_{\tilde{i}^b}}^{\tilde{D}} \geq 0 \\ \left\{ \begin{array}{l} \mu^b \sum_{I_i \in \tilde{N}} |v_{I_i}^{\tilde{D}}| + \mu^b |N| + \sum_{i > \tilde{i}^b, I_i \in \tilde{N}} |v_{I_i}^{\tilde{D}}| + |\{i : i > \tilde{i}^b, I_i \in N\}| \\ + u_{I_{\tilde{i}^b}} - (x_{I_{\tilde{i}^b}} + \mu^b v_{I_{\tilde{i}^b}}^{\tilde{D}}) \end{array} \right. & \text{if } I_{\tilde{i}^b} \in \tilde{N} \text{ and } v_{I_{\tilde{i}^b}}^{\tilde{D}} < 0 \\ \left\{ \begin{array}{l} \mu^b \sum_{I_i \in \tilde{N}} |v_{I_i}^{\tilde{D}}| + \mu^b |N| + \sum_{i > \tilde{i}^b, I_i \in \tilde{N}} |v_{I_i}^{\tilde{D}}| + |\{i : i > \tilde{i}^b, I_i \in N\}| \end{array} \right. & \text{if } I_{\tilde{i}^b} \in N. \end{cases}$$

Once the lengths of the forward and backward paths are determined, we can generate a candidate point z as follows. We generate a random number p uniformly distributed on the interval $[-t^b, t^f]$. If $0 < p \leq t^f$, the point is on the forward path, otherwise it is on the backward path.

First, suppose p is positive, $0 < p \leq t^f$, indicating the candidate point z is on the forward path. To map p to z , we have to identify whether z falls on a continuous interval, or a point associated with a discrete dimension. We first determine μ_p^f , the number of complete cycles of all dimensions from x to the candidate point, which is given by

$$\mu_p^f = \left\lfloor \frac{p}{\sum_{I_i \in \tilde{N}} |v_{I_i}^{\tilde{D}}| + |N|} \right\rfloor.$$

Starting from x , after moving μ_p^f complete cycles on the forward path, the remaining distance to z is $p - \mu_p^f (\sum_{I_i \in \tilde{N}} |v_{I_i}^{\tilde{D}}| + |N|)$. We find the order of coordinate dimension that encompasses the point z , say \tilde{i}_p^f , given as

$$\tilde{i}_p^f = \min_{i=1, \dots, n} \left\{ i : \sum_{j \leq i, I_j \in \tilde{N}} |v_{I_j}^{\tilde{D}}| + |\{j : j \leq i, I_j \in N\}| > p - \mu_p^f \left(\sum_{I_i \in \tilde{N}} |v_{I_i}^{\tilde{D}}| + |N| \right) \right\}.$$

Then, we move $\tilde{i}_p^f - 1$ steps on the forward path and the remaining distance to z is $p - \mu_p^f (\sum_{I_i \in \tilde{N}} |v_{I_i}^{\tilde{D}}| + |N|) - \sum_{i < \tilde{i}_p^f, I_i \in \tilde{N}} |v_{I_i}^{\tilde{D}}| - |\{i : i < \tilde{i}_p^f, I_i \in N\}|$.

If \tilde{i}_p^f is the order of a continuous dimension, then we move the remaining distance to z through the unit direction of $D_{I_{\tilde{i}_p^f}}$, which is $\frac{D_{I_{\tilde{i}_p^f}}}{|D_{I_{\tilde{i}_p^f}}|}$.

Thus, if $I_{\tilde{i}_p^f}$ is a continuous dimension, the candidate point is

$$z = x + \mu_p^f \sum_{i=1}^n D_{I_i} + \sum_{i=1}^{\tilde{i}_p^f-1} D_{I_i} + \left(p - \mu_p^f \left(\sum_{I_i \in \tilde{N}} |v_{I_i}^{\tilde{D}}| + |N| \right) - \sum_{i < \tilde{i}_p^f, I_i \in \tilde{N}} |v_{I_i}^{\tilde{D}}| - |\{i : i < \tilde{i}_p^f, I_i \in N\}| \right) \frac{D_{I_{\tilde{i}_p^f}}}{|D_{I_{\tilde{i}_p^f}}|}.$$

If $I_{\tilde{i}_p^f}$ is a discrete dimension, we stop the move before the step \tilde{i}_p^f , yielding,

$$z = x + \mu_p^f \sum_{i=1}^n D_{I_i} + \sum_{i=1}^{\tilde{i}_p^f-1} D_{I_i}.$$

If $-t^b \leq p \leq 0$, then the candidate point z is on the backward path and is calculated similarly. Let μ_p^b be the number of complete cycles from x to the candidate point, which is

$$\mu_p^b = \left\lfloor \frac{-p}{\sum_{I_i \in \tilde{N}} |v_{I_i}^{\tilde{D}}| + |N|} \right\rfloor.$$

Then let \tilde{i}_p^b be the order of coordinate dimension that encompasses the candidate point z ,

$$\tilde{i}_p^b = \max_{i=1, \dots, n} \left\{ i : \sum_{j \geq i, I_j \in \tilde{N}} |v_{I_j}^{\tilde{D}}| + |\{j : j \geq i, I_j \in N\}| > -p - \mu_p^b \left(\sum_{I_i \in \tilde{N}} |v_{I_i}^{\tilde{D}}| + |\tilde{N}| \right) \right\}.$$

If $I_{\tilde{i}_p^b}$ is a continuous dimension, the candidate point is

$$z = x - \mu_p^b \sum_{i=1}^n D_{I_i} - \sum_{i=\tilde{i}_p^b+1}^n D_{I_i} - \left(-p - \mu_p^b \left(\sum_{I_i \in \tilde{N}} |v_{I_i}^{\tilde{D}}| + |N| \right) - \sum_{i > \tilde{i}_p^b, I_i \in \tilde{N}} |v_{I_i}^{\tilde{D}}| - |\{i : i > \tilde{i}_p^b, I_i \in N\}| \right) \frac{D_{I_{\tilde{i}_p^b}}}{|D_{I_{\tilde{i}_p^b}}|}$$

If $I_{\tilde{i}_p^b}$ is a discrete dimension,

$$z = x - \mu_p^b \sum_{i=1}^n D_{I_i} - \sum_{i=\tilde{i}_p^b+1}^n D_{I_i}.$$

4 Convergence to a Target Distribution

We provide a theorem with necessary conditions to show that both Sphere and Box Biwalks, when used with a Metropolis filter in HR, converge to the target distribution used in the Metropolis filter. We first provide a theorem for the discrete case, and then for the mixed case.

First, we consider discrete HR with Sphere and Box Biwalks operated on $S \subseteq H$, where H is an integer lattice of a hyperrectangle. We let π be the target distribution associated with the Metropolis filter. We assume that π is a strictly positive probability mass function defined on S . We define the transition matrix, $Q = \{q_{xy} : x, y \in S\}$ of the Markov chain associated with Sphere or Box Biwalk for candidate points and the transition matrix $P = \{p_{xy} : x, y \in S\}$ of the Markov chain associated with discrete HR and Sphere or Box Biwalk after applying the Metropolis filter for target distribution π . Thus,

$$p_{xy} = \min \left\{ 1, \frac{\pi(y)}{\pi(x)} \right\} q_{xy} \quad \text{for all } x, y \in S. \quad (1)$$

In order to prove convergence to the target distribution we prove in Lemma 1 that the transition matrices associated with Sphere and Box Biwalks are symmetric, and then in Theorem 1 that the Markov chain for discrete HR with the Metropolis filter and Sphere or Box Biwalk converges to π .

Lemma 1 *Consider Sphere and Box Biwalks operated on $S \subseteq H$, where H is an integer lattice of a hyperrectangle such that $H = \prod_{i=1, \dots, n} \{l_i, l_i + 1, \dots, u_i\}$ and $l_i, u_i \in \mathbb{Z}$ with $l_i < u_i$ for all $i = 1, \dots, n$. The transition matrices Q are symmetric.*

Proof Let an ordered set of points $r = \{r_1, \dots, x, \dots, r_m\}$ denote a bidirectional walk, generated with Sphere or Box Biwalk from a starting point x ; and r_1 and r_m being the end of backward and forward paths respectively. Let $P(r | x)$ denote the probability of generating path r from the starting point x . The probability of generating the path only depends on the choice of discrete step directions and the order of dimensions, which are independent of the starting point. Therefore any path r generated by Sphere or Box Biwalk satisfies $P(r | x) = P(r | z)$ for all $x, z \in r$. Now consider any $x, z \in S$. Let $R_{x,z}$ be the set of all possible paths with starting point x that contain z and let $\tilde{P}(z | r)$ be the probability of selecting point z uniformly from the points in path r . Then,

$$q_{xz} = \sum_{r \in R_{x,z}} P(r | x) \tilde{P}(z | r).$$

We also have that $R_{x,z} = R_{z,x}$, because for any $r \in R_{x,z}$ there is a corresponding path $r' \in R_{z,x}$ with reversed direction vectors and permutations (i.e. if z is on the forward path of r , then x is on the backward path of r'), and vice versa. Finally, given path $r \in R_{x,y}$, since all points on the path are equally likely to be chosen, $\tilde{P}(z | r) = \tilde{P}(x | r)$. Hence

$$q_{zx} = \sum_{r \in R_{z,x}} P(r | z) \tilde{P}(x | r) = q_{xz}.$$

Thus, Q is symmetric. ■

Theorem 1 *Consider HR with Sphere and Box Biwalks operated on $S \subseteq H$, where H is an integer lattice of a hyperrectangle such that $H = \prod_{i=1, \dots, n} \{l_i, l_i + 1, \dots, u_i\}$ and $l_i, u_i \in \mathbb{Z}$ with $l_i < u_i$ for all $i = 1, \dots, n$. Let π be the target distribution associated with the Metropolis filter.*

i.) For Sphere Biwalk, if $R \geq (\sum_{i=1}^n (u_i - l_i)^2)^{\frac{1}{2}}$, or if $R \geq 1$ and $S = H$;

ii.) for Box Biwalk, if $c_i = u_i - l_i$ for $i = 1, \dots, n$, or if $c_i \geq 1$ and $S = H$,

then the Markov chains generated by HR with Sphere and Box Biwalks converge in distribution to the target distribution π over S .

Proof We first prove the irreducibility of $P = \{p_{xy} : x, y \in S\}$, the transition matrix associated with discrete HR with Metropolis filter π and Sphere or Box Biwalk. For Sphere Biwalk, if $R \geq (\sum_{i=1}^n (u_i - l_i)^2)^{\frac{1}{2}}$, the probability of generating any point $y \in S$ from a current point, say $x \in S$ is positive; because the probability of generating \tilde{D} that is in the direction $y - x$ in Step 1.1 and rounding to y after generating a uniform point on the line set $\tilde{L}_{\tilde{D}} = \{v : v = r\tilde{D}, r \in [-R, R]\}$, due to R chosen sufficiently big, is positive. Similarly, for Box Biwalk, if $c_i = u_i - l_i$ for all $i = 1, \dots, n$, then the probability of generating any point $y \in S$ after rounding a uniform point in the box $[-c_1, c_1] \times \dots \times [-c_n, c_n]$ in Step 1.2 is positive. Moreover, if $S = H$, the probability of generating any point from a current point is also positive as long as $R \geq 1$ for Sphere Biwalk, and $c_i \geq 1$ for Box Biwalk. Therefore, for both biwalks, given conditions *i* and *ii*, $q_{xy} > 0$ for all $x, y \in S$. And given that π is strictly positive, $p_{xy} = \min\{1, \pi(y)/\pi(x)\}q_{xy} > 0$ for all $x, y \in S$, therefore, P is irreducible. The generated biwalks include the current point, therefore the probability of staying at the current point is positive. Thus, P is aperiodic.

Next, we prove that P is reversible relative to π , that is,

$$\pi(x)p_{xy} = \pi(y)p_{yx} \quad \text{for all } x, y \in S.$$

Since Q is symmetric as provided in Lemma 1,

$$q_{xy} = q_{yx} \quad \text{for all } x, y \in S \quad (2)$$

and, we observe that

$$\pi(x) \min\left\{1, \frac{\pi(y)}{\pi(x)}\right\} = \pi(y) \min\left\{1, \frac{\pi(x)}{\pi(y)}\right\}, \quad (3)$$

then combining (1), (2), and (3), we conclude that

$$\begin{aligned} \pi(x)p_{xy} &= \pi(x) \min\left\{1, \frac{\pi(y)}{\pi(x)}\right\} q_{xy} = \pi(y) \min\left\{1, \frac{\pi(x)}{\pi(y)}\right\} q_{yx} \\ &= \pi(y)p_{yx} \quad \text{for all } x, y \in S. \end{aligned}$$

From reversibility, we can take the summation,

$$\sum_{x \in S} \pi(x)p_{xy} = \sum_{x \in S} \pi(y)p_{yx} \quad \text{for all } y \in S$$

and because P is a probability transition matrix, $\sum_{y \in S} p_{xy} = 1$; thus,

$$\sum_{x \in S} \pi(x)p_{xy} = \pi(y)$$

and given that the Markov chain is irreducible and aperiodic and S is finite,

$$\sum_{x \in S} \pi(x)p_{xy} = \lim_{n \rightarrow \infty} P^n = \pi(y) \quad \text{for all } y \in S.$$

Thus, we conclude that for every starting point $x \in S$, the Markov chain associated with discrete HR with Sphere or Box Biwalk when used with a Metropolis filter with target distribution π converges to π . ■

To prove a similar convergence theorem for HR with Sphere and Box Biwalks on mixed continuous/discrete domains, we have to consider biwalks with intermingled intervals and points. While the current point x is always on the left-hand side of an interval, the next point z may be a point interior to an interval on a continuous dimension. However, for any path $r \in R_{x,z}$ that starts at x and includes z , with corresponding step length directions $\{D'_1, \dots, D'_n\}$ and order of permutations $\{I_1, \dots, I_n\}$ there exists a unique path $r' \in R_{z,x}$, where $R_{z,x}$ denotes the set of all paths starting from z and passing through x . The path r' is determined by the negative step directions $-D'_i$ for all i except for the dimension that includes the point z , $\{-D'_1, \dots, -D'_{k-1}, D'_k, -D'_{k+1}, \dots, -D'_n\}$ with the opposite order of r . We replace that step direction D'_k , with D'_k whose step size equals $|z_k - x_k|$ divided by the number of D'_k steps taken from x to z on r . In this way, the total length of the biwalk on the k^{th} dimension is equal for both paths r and r' , thus the probability of choosing z given x on r equals the probability of choosing x given z and path r' . Thus, symmetry of Q is also satisfied for the mixed case. The convergence theorem for mixed domains is now given.

Theorem 2 Consider HR with Sphere and Box Biwalks operated on $S \subseteq H$, where H is a mixed continuous/discrete lattice of a hyperrectangle such that

$$H = \prod_{i=1, \dots, n} \begin{cases} [l_i, u_i] & \text{if } i \text{ is a continuous dimension} \\ \{l_i, l_i + 1, \dots, u_i\} & \text{if } i \text{ is a discrete dimension,} \end{cases}$$

provided that $l_i, u_i \in \mathbb{R}$, if i is a continuous dimension and $l_i, u_i \in \mathbb{Z}$, if i is a discrete dimension and $l_i < u_i$ for all $i = 1, \dots, n$. Let π be the target distribution associated with the Metropolis filter used in HR with pattern biwalk.

i.) For Sphere Biwalk, if $R \geq (\sum_{i=1}^n (u_i - l_i)^2)^{\frac{1}{2}}$,

ii.) for Box Biwalk, if $c_i = u_i - l_i$ for $i = 1, \dots, n$,

then the Markov chains generated by HR with Sphere and Box Biwalks converge in distribution to the target distribution π over S .

5 Convergence to Hit-and-Run

Sphere Biwalk is a mixed continuous/discrete analog of the original continuous HR, where the direction is chosen uniformly on the surface of a hypersphere. And Box Biwalk is a mixed continuous/discrete analog of a variation of continuous HR where the direction is chosen uniformly in the interior of a hyperrectangle. We provide the necessary conditions and theorems to prove the convergence of the mixed continuous/discrete form to the continuous form as the mixed lattice gets finer.

Consider the Sphere Biwalk on a finite space S_δ contained in a mixed lattice of an hyperrectangle H_δ with δ being the minimum distance between two lattice points for discrete dimensions in H_δ . That is, $H_\delta = \{(x_1, x_2, \dots, x_n) \in \mathbb{R}^n : l_i \leq x_i \leq u_i \text{ for } i = 1, \dots, n \text{ and } \frac{x_i}{\delta} \in \mathbb{Z} \text{ for discrete dimensions } i\}$. In order to prove convergence, we introduce a continuous hyperrectangle $\tilde{H} = \{(x_1, x_2, \dots, x_n) \in \mathbb{R}^n : l_i \leq x_i \leq u_i \text{ for } i = 1, \dots, n\}$ and let $\tilde{S} \subseteq \tilde{H}$ be a measurable continuous subset of \tilde{H} , such that

$$\tilde{H} = \lim_{\delta \rightarrow 0} H_\delta \text{ and } \tilde{S} = \lim_{\delta \rightarrow 0} S_\delta.$$

Let $\tilde{D} = (\tilde{d}_1, \tilde{d}_2, \dots, \tilde{d}_n)$ be a continuous unit direction generated uniformly on the surface of an n -dimensional unit hypersphere with center being the origin, as in Step 1.1 of Sphere Biwalk. For a given pair of R and δ , where $R \geq \delta$, let $V_{(R,\delta)}^{\tilde{D}} = (v_{(R,\delta)1}^{\tilde{D}}, \dots, v_{(R,\delta)n}^{\tilde{D}})$ be the mixed continuous/discrete lattice point obtained by rounding \tilde{D} , as in Step 1.2 of Sphere Biwalk. We denote the set of step length directions, as in Step 1.3, by $D_{(R,\delta)}^{\tilde{D}} = \{D_{(R,\delta)1}^{\tilde{D}}, \dots, D_{(R,\delta)n}^{\tilde{D}}\}$. Given a starting point $x_0 \in S_\delta$, let $S_{D_{(R,\delta)}^{\tilde{D}}}(x_0)$ be the set of mixed continuous/discrete points contained in S_δ that are on at least one of the biwalks generated from x_0 with $D_{(R,\delta)}^{\tilde{D}}$ and any permutation of the dimensions. Let $\tilde{L}_{\tilde{D}}(x_0)$ be the line set of points generated by continuous Hit-and-Run from x_0 with the direction \tilde{D} over \tilde{S} , i.e.,

$$\tilde{L}_{\tilde{D}}(x_0) = \{x \in \tilde{S} \subseteq \tilde{H} : x = x_0 + \beta \tilde{D}, \beta \text{ is a real scalar}\}.$$

The following theorem shows that as $\frac{\delta}{R} \rightarrow 0$ and $R \rightarrow 0$, the biwalk generated by Sphere Biwalk converges to the direction vector generated by continuous Hit-and-Run.

Theorem 3 Consider Sphere Biwalk on the set S_δ contained in H_δ and $R \geq \delta$. Given a current point $x_0 \in S_\delta$, let \tilde{D} be a unit continuous direction generated by Hit-and-Run, and let $D_{(R,\delta)}^{\tilde{D}}$ be the set of step length directions generated by Sphere Biwalk whose corresponding unit continuous direction is \tilde{D} . Then as $\frac{\delta}{R} \rightarrow 0$ and $R \rightarrow 0$, the biwalk set $S_{D_{(R,\delta)}^{\tilde{D}}}(x_0)$ converges to the line set $\tilde{L}_{\tilde{D}}(x_0)$, i.e.,

- (i) for every subsequence of points $\{x_{(R,\delta)}\}$ such that $x_{(R,\delta)} \in S_{D_{(R,\delta)}^{\tilde{D}}}(x_0)$ for all R and δ , if $\{x_{(R,\delta)}\} \rightarrow x$ as $\frac{\delta}{R} \rightarrow 0$ and $R \rightarrow 0$, then $x \in \tilde{L}_{\tilde{D}}(x_0)$.
- (ii) for every $x \in \tilde{L}_{\tilde{D}}(x_0)$, there is a sequence of points $\{x_{(R,\delta)}\}$ with $x_{(R,\delta)} \in S_{D_{(R,\delta)}^{\tilde{D}}}(x_0)$ for all R and δ , such that $x_{(R,\delta)} \rightarrow x$ as $\frac{\delta}{R} \rightarrow 0$ and $R \rightarrow 0$.

Proof See Appendix. ■

Similar to Sphere Biwalk, under a particular rule of choosing maximum step length c_i for each dimension, the biwalk generated by Box Biwalk converges to a direction vector of a version of continuous HR algorithm, that is generated uniformly in the interior of an n -dimensional hyperrectangle.

Theorem 4 Consider Box Biwalk on the set S_δ contained in H_δ . For $i = 1, 2, \dots, n$, let c_i be an arbitrary real number between $[1, u_i - l_i]$, and let γ_δ be any value between $\left[\frac{\delta}{\min_{i=1, \dots, n} c_i}, 1\right]$. Set the maximum step length $c_{\delta i}$ associated with Box Biwalk to $c_{\delta i} = \gamma_\delta c_i$, for $i = 1, \dots, n$. Given a current point $x_0 \in S_\delta$, let $\tilde{D} = (\tilde{d}_1, \dots, \tilde{d}_n)$ be a continuous point generated uniformly in the box $[-c_1, c_1] \times \dots \times [-c_n, c_n]$ and let $V_{(\gamma_\delta)}^{\tilde{D}} = (v_{(\gamma_\delta)1}^{\tilde{D}}, \dots, v_{(\gamma_\delta)n}^{\tilde{D}})$ be the mixed continuous/discrete lattice point obtained by rounding \tilde{D} , where $v_{(\gamma_\delta)i}^{\tilde{D}}/\tilde{d}_i = \gamma_\delta$ for $i = 1, \dots, n$. Also, let $D_{(\gamma_\delta)}^{\tilde{D}}$ be the set of step length directions generated by Box Biwalk corresponding to $V_{(\gamma_\delta)}^{\tilde{D}}$.

As in Theorem 3, let $S_{D_{(\gamma_\delta)}^{\tilde{D}}}(x_0)$ be the set of mixed continuous/discrete points contained in S_δ that are on at least one of the biwalks generated from x_0 with $D_{(\gamma_\delta)}^{\tilde{D}}$ and any permutation of the dimensions. Let $\tilde{L}_{\tilde{D}}(x_0)$ be the line set of points generated by Hit-and-Run from x_0 with the direction \tilde{D} over \tilde{S} , i.e.,

$$\tilde{L}_{\tilde{D}}(x_0) = \{x \in \tilde{S} \subseteq \tilde{H} : x = x_0 + \beta \tilde{D}, \beta \text{ is a real scalar}\}.$$

Then as $\frac{\delta}{\gamma_\delta} \rightarrow 0$ and $\gamma_\delta \rightarrow 0$, the biwalk set $S_{D_{(\gamma_\delta)}^{\tilde{D}}}(x_0)$ converges to the line set $\tilde{L}_{\tilde{D}}(x_0)$, i.e.,

(i) for every subsequence of points $\{x_{(\gamma_\delta)}\}$ such that $x_{(\gamma_\delta)} \in S_{D_{(\gamma_\delta)}^\delta}(x_0)$ for all γ_δ and δ , if

$$\{x_{(\gamma_\delta)}\} \rightarrow x \text{ as } \frac{\delta}{\gamma_\delta} \rightarrow 0 \text{ and } \gamma_\delta \rightarrow 0, \text{ then } x \in \tilde{L}_D(x_0).$$

(ii) for every $x \in \tilde{L}_D(x_0)$, there is a sequence of points $\{x_{(\gamma_\delta)}\}$ with $x_{(\gamma_\delta)} \in S_{D_{(\gamma_\delta)}^\delta}(x_0)$ for all

$$\gamma_\delta \text{ and } \delta, \text{ such that } x_{(\gamma_\delta)} \rightarrow x \text{ as } \frac{\delta}{\gamma_\delta} \rightarrow 0 \text{ and } \gamma_\delta \rightarrow 0.$$

Proof The proof is very similar to the proof for Theorem 3 and is not repeated. ■

6 Finite time performance

We analyze the convergence rate of the discrete HR sampling with a target distribution π using Sphere Biwalk and Box Biwalk over a finite domain S , where S is an n -dimensional hyperrectangle integer lattice with $S = H = \{(x_1, x_2, \dots, x_n) \in \mathbb{Z}^n : l_i \leq x_i \leq u_i, \text{ with } l_i < u_i, i = 1, \dots, n\}$. We provide bounds on the rate of convergence to general π , and then give specific bounds when π is uniform and when π is a Boltzmann distribution for Sphere Biwalk (Corollary 1) and Box Biwalk (Corollary 2). Global optimization algorithms such as Simulated Annealing often use a Boltzmann Metropolis filter associated with π_T , the Boltzmann T distribution on the function f over S , $\pi_T(x) = \frac{e^{-f(x)/T}}{\sum_{z \in S} e^{-f(z)/T}}$ in their cooling schedules.

As the measure of convergence, we use the absolute version of the maximal relative error introduced by Behrends [5] for irreducible, aperiodic, and reversible transition matrices of Markov chains defined on a finite state space with stationary distributions. In order to build this measure for the discrete HR sampling with pattern biwalks, we start with defining $P = \{p_{xx'}\}_{x, x' \in S}$ as the transition matrix of the Markov chain associated with discrete HR with Sphere and Box Biwalk with Metropolis filter for target distribution π on a finite state space S . Denote the k^{th} transition matrix by $P^{(k)} = \{p_{xx'}^{(k)}\}$. We use the maximal absolute error, denoted by $d(k)$ to measure the distance of the k^{th} transition of the Markov chain from the stationary distribution, which is defined as

$$d(k) = \max_{x, x' \in S} |p_{xx'}^{(k)} - \pi(x')|. \quad (4)$$

Since the Markov chain associated with discrete HR with Sphere or Box Biwalk is irreducible, aperiodic and it has a stationary distribution (from Theorem 1), then all of its states are positive recurrent and it is ergodic [21, Theorem 4.3.3. p. 175]. Moreover, ergodicity provides us with the property that P has real eigenvalues that can be ordered as $1 = \lambda_{p_1} > \lambda_{p_2} \geq \dots \geq \lambda_{p_n} > -1$ and the rate of convergence to the stationary distribution is governed by the second-largest eigenvalue in absolute value, say λ_p^* [24]. A lower bound for the number of iterations to limit the maximal absolute distance in an $\varepsilon > 0$ boundary is provided by Lemma 2 below.

Lemma 2 Given $\varepsilon > 0$, if $k \geq \frac{\ln\left(\frac{\max_{x \in S} \pi(x)}{\min_{x \in S} \pi(x)}\right) - \ln(\varepsilon)}{1 - \lambda_p^*}$, then $d(k) = \max_{x, x' \in S} |p_{xx'}^{(k)} - \pi(x')| \leq \varepsilon$.

Proof In [5, Theorem 10.3 p. 81], a bound on the maximal relative error is given as

$$\max_{x, x' \in S} \frac{|p_{xx'}^{(k)} - \pi(x')|}{\pi(x')} \geq \frac{\lambda_p^*}{\min_{x \in S} \pi_x} \quad (5)$$

We utilize this bound to get a bound for $d(k)$. The relation between maximal absolute and relative errors can be derived as

$$\max_{x,x' \in S} \frac{|p_{xx'}^{(k)} - \pi(x')|}{\pi(x')} \geq \frac{\max_{x,x' \in S} |p_{xx'}^{(k)} - \pi(x')|}{\max_{x \in S} \pi(x)} = \frac{d(k)}{\max_{x \in S} \pi(x)}.$$

Thus, using the bound on maximal relative error,

$$d(k) \leq \left(\max_{x,x' \in S} \frac{|p_{xx'}^{(k)} - \pi(x')|}{\pi(x')} \right) \max_{x \in S} \pi(x) \leq \frac{\max_{x \in S} \pi(x)}{\min_{x \in S} \pi(x)} \lambda_p^* \quad \text{for all } k. \quad (6)$$

By assumption of the lemma, $k \geq \frac{\ln\left(\frac{\max_{x \in S} \pi(x)}{\min_{x \in S} \pi(x)} / \varepsilon\right)}{1 - \lambda_p^*}$. Hence,

$$\ln\left(\varepsilon / \left(\frac{\max_{x \in S} \pi(x)}{\min_{x \in S} \pi(x)}\right)\right) \geq (\lambda_p^* - 1)k \geq k(\ln(\lambda_p^*)),$$

since $0 < \lambda_p^* < 1$. Thus, $(\lambda_p^*)^k \leq \varepsilon / \left(\frac{\max_{x \in S} \pi(x)}{\min_{x \in S} \pi(x)}\right)$. Then, by combining with (6), we get $d(k) \leq \varepsilon$. ■

As Lemma 2 shows, an upper bound on λ_p^* provides a bound on the rate of convergence for a Markov chain. We next develop a bound on λ_p^* for Discrete Hit-and-Run with Sphere Biwalk and Box Biwalk. Because the eigenvalues are ordered, λ_p^* will be either $|\lambda_{p_2}|$ or $|\lambda_{p_n}|$. Therefore an upper bound on both $|\lambda_{p_2}|$ and $|\lambda_{p_n}|$ provides a bound on λ_p^* . We first develop a bound on $|\lambda_{p_n}|$.

Lemma 3 Consider DHR with Metropolis filter π with Sphere Biwalk or Box Biwalk being the candidate point generator over a finite state space S , where S is an n -dimensional hyper-rectangle integer lattice, $S = H$. Let Q be the candidate transition matrix with eigenvalues $1 = \lambda_{Q_1} > \lambda_{Q_2} \geq \dots \geq \lambda_{Q_n} > -1$, and let P be the DHR transition matrix with eigenvalues $1 = \lambda_{P_1} > \lambda_{P_2} \geq \dots \geq \lambda_{P_n} > -1$, where P is determined by Q and π . Then both $|\lambda_{Q_n}|$ and $|\lambda_{P_n}|$ are bounded by

$$|\lambda_{Q_n}| \leq 1 - \frac{2}{n(L+1)} \quad \text{and} \quad |\lambda_{P_n}| \leq 1 - \frac{2}{n(L+1)} \quad (7)$$

where $L = \max_{i=1}^n (u_i - l_i)$.

Proof The minimal holding probabilities as defined in [17], $\alpha_Q, \alpha_P \in [0, 1]$ for matrices Q and P respectively, provides that $q_{xx} \geq \alpha_Q$ and $p_{xx} \geq \alpha_P$ for all $x \in S$. Since Q and P are reversible, the smallest eigenvalues satisfy the following relations:

$$-\lambda_{Q_n} \leq 1 - 2\alpha_Q \quad \text{and} \quad -\lambda_{P_n} \leq 1 - 2\alpha_P, \quad (8)$$

which are given in [17, p. 16]. According to the process of DHR, the probability of staying at the same point only depends on the number of candidate points along the bidirectional path, which implies $q_{xx} = p_{xx}$ for all $x \in S$. It can also be shown that the points in any bidirectional path generated by Sphere Biwalk or Box Biwalk are distinct and the number of candidate points along the path is no more than $n(L+1)$. Therefore, the one-step transition probabilities q_{xx} and p_{xx} are not less than $\frac{1}{n(L+1)}$. Thus,

$$\alpha_Q = \alpha_P \geq \frac{1}{n(L+1)}. \quad (9)$$

Combining (8) and (9), we get

$$\lambda_{Q_n} \geq -1 + \frac{2}{n(L+1)}, \quad \text{and} \quad \lambda_{P_n} \geq -1 + \frac{2}{n(L+1)}, \quad (10)$$

which implies $|\lambda_{Q_n}| \leq 1 - \frac{2}{n(L+1)}$ and $|\lambda_{P_n}| \leq 1 - \frac{2}{n(L+1)}$. ■

Lemma 3 provides a bound on the smallest eigenvalues of candidate transition matrix Q and transition matrix P . We next develop bounds on the second largest eigenvalue for both Q and P respectively.

Lemma 4 Consider DHR with Metropolis filter π with Sphere Biwalk or Box Biwalk being the candidate point generator over a finite state space S , where S is an n -dimensional hyperrectangle integer lattice, $S = H$. Let λ_{Q_2} and λ_{P_2} be the second largest eigenvalues of Q and P respectively, as in Lemma 3. Then for DHR with either Sphere or Box Biwalk, λ_{Q_2} is bounded by

$$|\lambda_{Q_2}| \leq 1 - 4 \left(\min_{x,y \text{ adjacent}} q_{xy} \right) (n(L+1)^2)^{-1},$$

and when π is a uniform distribution, $\lambda_{P_2} = \lambda_{Q_2}$. When π is a Boltzmann T distribution,

$$|\lambda_{P_2}| \leq 1 - 2e^{-2(y^* - y_*)/T} \left(\min_{x,y \text{ adjacent}} q_{xy} \right) (n(L+1)^2)^{-1},$$

where $f^* = \max_{x \in S} f(x)$, $f_* = \min_{x \in S} f(x)$, and $L = \max_{i=1}^n (u_i - l_i)$.

Proof The proof for the bound on λ_{Q_2} uses the geometry bound on the second largest eigenvalue of an irreducible, aperiodic, reversible Markov chain developed by Diaconis and Strook [8] and a technique which was explored by Sinclair [24] called the method of the canonical paths. The bound on λ_{P_2} follows directly from the relationship between λ_{Q_2} and λ_{P_2} provided by Behrends in [5, Proposition 21.3, p. 210]. ■

Now we are ready to provide the convergence rate of Discrete Hit-and-Run with Boltzmann Metropolis filter and using Sphere Biwalk and Box Biwalk candidate point generators respectively. We start with an analysis of DHR with Sphere Biwalk.

Consider the Markov chain $\{X_k, k = 0, 1, 2, \dots\}$ generated by DHR with Sphere Biwalk. The following lemma provides a bound on the one step candidate transition probability between two adjacent points in S .

Lemma 5 Consider DHR with Metropolis filter π with Sphere Biwalk and parameter $R \geq 1$ being the candidate point generator over a finite state space S , where S is an n -dimensional hyperrectangle integer lattice, $S = H$. Let (x, z) be a pair of adjacent points in S . Then the one step candidate transition probability q_{xz} is bounded by $q_{xz} \geq \frac{1}{n^2(1+nL)(R+1)}$, where $L = \max_{i=1, \dots, n} (u_i - l_i)$.

Proof See Appendix. ■

By applying Lemma 5 to Lemma 3 and Lemma 4, the second largest eigenvalue modulus λ_Q^* and λ_P^* are bounded by

$$\lambda_Q^* = \max\{|\lambda_{Q_2}|, |\lambda_{Q_n}|\} \leq 1 - \frac{4}{n^3(1+nL)(L+1)^2(R+1)}, \quad (11)$$

$$\lambda_P^* = \max\{|\lambda_{P_2}|, |\lambda_{P_n}|\} \leq 1 - \frac{2e^{-2(f^*-f_*)/T}}{n^3(1+nL)(L+1)^2(R+1)}. \quad (12)$$

By providing the bounds on λ_Q^* and λ_P^* to Lemma 2, we can make the following conclusion on the convergence rate of DHR with Sphere Biwalk to its target distribution.

Corollary 1 *Consider DHR with Metropolis filter π with Sphere Biwalk and parameter $R \geq 1$ being the candidate point generator over a finite state space S , where S is an n -dimensional hyperrectangle integer lattice, $S = H$. If the stationary distribution associated with the Metropolis filter is a uniform distribution, then the number of iterations required for the distribution of the chain to be within maximal absolute error ε of the uniform distribution on S is at most,*

$$(1/4)n^3(1+nL)(L+1)^2(R+1)\ln(\varepsilon^{-1}). \quad (13)$$

If the stationary distribution associated with the Metropolis filter is a Boltzmann T distribution, then the number of iterations required for the distribution of the chain to be within maximal absolute error ε of the Boltzmann T distribution on S is at most,

$$(1/2)n^3(1+nL)(L+1)^2(R+1)e^{2(f^*-f_*)/T} \left(\frac{f^*-f_*}{T} - \ln \varepsilon \right). \quad (14)$$

Proof The bound on the convergence rate to a uniform distribution follows from (11), Lemma 2 and the fact that the two probability transition matrices Q and P are equal when the stationary distribution of P is a uniform distribution. The bound on the convergence rate to a Boltzmann distribution follows from (12) and Lemma 2. ■

Hence if S is the integer lattice of a hyperrectangle of dimension n , the Discrete Hit-and-Run with Sphere Biwalk and with a uniform filter produces an approximately uniformly distributed sample point on S in polynomial time, $O(n^4)$; the Discrete Hit-and-Run with Sphere Biwalk and with a Boltzmann T filter produces an approximately Boltzmann T distributed sample point on S in polynomial time, $O(n^4)$, when holding the other parameters constant.

To show the difference between Box Biwalk and Sphere Biwalk, we now consider the Markov chain generated by DHR with Box Biwalk. Let Q denote the candidate transition matrix associated with the Box Biwalk. The following lemma provides a bound on the one step candidate transition probability between two adjacent points in S .

Lemma 6 *Consider DHR with Metropolis filter π with Box Biwalk and parameter $1 \leq c_i \leq u_i - l_i$ for all $i = 1, \dots, n$ being the candidate point generator over a finite state space S , where S is an n -dimensional hyperrectangle integer lattice, $S = H$. Then the one step candidate transition probability q_{xz} is bounded by $q_{xz} \geq \frac{1}{n(1+nL)L}$, where $L = \max_{i=1, \dots, n}(u_i - l_i)$.*

Proof See Appendix. ■

By applying Lemma 6 to Lemma 3 and Lemma 4 for DHR with Box Biwalk, the second largest eigenvalue modulus λ_Q^* and λ_P^* are bounded by

$$\lambda_Q^* = \max\{|\lambda_{Q_2}|, |\lambda_{Q_n}|\} \leq 1 - \frac{4}{n^2(1+nL)L(L+1)^2}, \quad (15)$$

$$\lambda_P^* = \max\{|\lambda_{P_2}|, |\lambda_{P_n}|\} \leq 1 - \frac{2e^{-2(f^*-f_*)/T}}{n^2(1+nL)L(L+1)^2}. \quad (16)$$

By providing the bounds on λ_Q^* and λ_P^* to Lemma 2, we can make the following conclusion on the convergence rate of DHR with Box Biwalk to its target distribution.

Corollary 2 Consider DHR with Metropolis filter π with Box Biwalk and parameter $1 \leq c_i \leq u_i - l_i$ for all $i = 1, \dots, n$ being the candidate point generator over a finite state space S , where S is an n -dimensional hyperrectangle integer lattice, $S = H$. If the stationary distribution associated with the Metropolis filter is a uniform distribution, then the number of iterations required for the distribution of the chain to be within maximal absolute error ε of the uniform distribution on S is at most,

$$(1/4)n^2L(L+1)^2(1+nL)\ln(\varepsilon^{-1}). \quad (17)$$

If the stationary distribution associated with the Metropolis filter is a Boltzmann T distribution, then the number of iterations required for the distribution of the chain to be within maximal absolute error ε of the Boltzmann T distribution on S is at most,

$$(1/2)n^2L(L+1)^2(1+nL)e^{2(f^*-f_*)/T} \left(\frac{f^* - f_*}{T} - \ln \varepsilon \right). \quad (18)$$

Proof The bound on the convergence rate to a uniform distribution follows from (15), Lemma 2 and the fact that the two probability transition matrices Q and P are equal when the stationary distribution of P is a uniform distribution. The bound on the convergence rate to a Boltzmann distribution follows from (16) and Lemma 2. ■

Hence if S is the integer lattice of a hyperrectangle of dimension n , the Discrete Hit-and-Run with Box Biwalk and with a uniform filter produces an approximately uniformly distributed sample point on S in polynomial time, $O(n^3)$; the Discrete Hit-and-Run with Box Biwalk and with a Boltzmann T filter produces an approximately Boltzmann T distributed sample point on S in polynomial time, $O(n^3)$, holding the other parameters constant.

7 Computational Study

We tested and compared modified Improving Hit-and-Run with several candidate point generators: our new pattern biwalks, DHR with random biwalk [4] for discrete domain, separate continuous HR and discrete HR with random biwalk for mixed domains, and the Step Function approach [20] for both discrete and mixed domains.

We solved 18 global optimization test problems which are used in [16] and given in [1]. The problems are listed in Table 1, with dimensions and parameter settings we used. The radius parameter (R) for Sphere Biwalk and box length parameter (c) for Box Biwalk are determined according to the longest side of H and have three values: a small, medium and large value. For the mixed continuous/discrete domain problems, we take half of the

dimensions as continuous and the other half as discrete. We perform 5,000 iterations for 100 runs for each test case.

For each problem, we evaluate the runs according to the m -fold improvement (scaled distance to the optimal function value f^* after a number of iterations) measure, which is developed by Ali et al. [1] by modifying performance profile given by Dolan and Moré [9]. Accordingly, the m -fold improvement of each method (with its parameter if needed) is: $\frac{\hat{f}_s - f^*}{f_w - f^*}$ where \hat{f}_s is the average incumbent function value after 5,000 iterations and f_w is the worst average incumbent function value of 100 runs obtained among the methods. Then, we calculate the performance ratio for each method by dividing m -fold improvement by the minimum of m -fold improvements. Then, the method/parameter setting that provides performance ratio one can be seen as the best. The performance ratios for each test problem and method/parameter setting are given in Table 2. For the problems with 20 dimensions, we provide the graphs of the average incumbent function values for 5,000 iterations in Figure 1 for the discrete domains and Figure 2 for the mixed continuous/discrete domains. We excluded the Rosenbrock function in 20 dimensions from the figures because the differences among the methods are not visible in the graphs.

Table 1 Dimensions and Parameters (R for Sphere Biwalk, c for Box Biwalk) for Test Problems

Problems	n	R_s and c_s	R_m and c_m	R_l and c_l
Ackley	20	2	30	60
Epistatic Michalewicz	5	1	2	3
Epistatic Michalewicz	10	1	2	3
Griewank	10	2	600	1,200
Griewank	20	2	600	1,200
Levy and Montalvo 2	10	2	5	10
Levy and Montalvo 2	20	2	5	10
Paviani	10	2	5	8
Rosenbrock	10	2	30	60
Rosenbrock	20	2	30	60
Salomon	10	2	100	200
Salomon	20	2	100	200
Schwefel	10	2	500	1,000
Shekel Foxholes	10	2	5	10
Sinusoidal, Centered ($z = 0$)	10	2	90	180
Sinusoidal, Centered ($z = 0$)	20	2	90	180
Sinusoidal, Shifted ($z = 30$)	10	2	90	180
Sinusoidal, Shifted ($z = 30$)	20	2	90	180

For the discrete and mixed domains, Sphere and Box Biwalks outperform Random Biwalk and the Step Function approach in almost all of the 36 test cases. We observe that Sphere Biwalk with small R values and Box Biwalk with large c values performed better than the other methods/parameter settings overall.

8 Conclusion

We introduce two new Markov chain Monte Carlo samplers for neighborhood generation in global optimization algorithms. Sphere and Box Biwalks are motivated by the success of Hit-and-Run and its discrete version, however they utilize the patterns for generating the

Table 2 Performance Ratios

Discrete Problems	n	Ran. B.	Step Fun.	Sphere Biwalk			Box Biwalk		
				R_s	R_m	R_l	c_s	c_m	c_l
Ackley	20	3.78	1.86	1	2.29	2.68	1.65	3.67	1.27
E. Michalewicz	5	1	1.01	1	1	1	1	1	1
E. Michalewicz	10	1.02	1.17	1.02	1.05	1.08	1	1.2	1.25
Griewank	10	59.74	21.82	1	7.68	4.14	5.43	4.13	2.48
Griewank	20	12.01	13.29	1	10.22	9.7	10.07	2.35	2.15
Levy and M. 2	10	> 1000	> 1000	1	1	1	1	> 1000	1
Levy and M. 2	20	> 1000	> 1000	1	> 1000	> 1000	> 1000	> 1000	1
Paviani	10	1	3.03	1	1	1	1	1	1
Rosenbrock	10	7.14	2.36	1	1.38	1.72	1.51	2.51	1.1
Rosenbrock	20	21.76	5.28	1	4.71	7.69	3.7	15.98	1.75
Salomon	10	2.15	1.11	1	1.54	1.83	1.08	2.51	1.56
Salomon	20	1.93	1.26	1	1.46	1.67	1.22	2.1	1.39
Schwefel	10	174.13	171.43	1	57.1	49.7407	123.92	52.09	2.4
Shekel Foxholes	10	1	1	1	1	1	1	1	1.01
Sinusoidal Cen.	10	15	7.88	1	4.46	3.7	2.91	4.44	3.48
Sinusoidal Cen.	20	2.83	2.18	1.39	2.03	2.33	1.9	2.25	1
Sinusoidal Shi.	10	10.71	9.27	1	8.31	6.18	4.4	2.85	3.58
Sinusoidal Shi.	20	2.01	2	1.39	1.85	1.79	1.8	1.12	1
Mixed Problems	n	Sep. Dom.	Step Fun.	Sphere Biwalk			Box Biwalk		
				R_s	R_m	R_l	c_s	c_m	c_l
Ackley	20	2.95	2.71	3.19	1.52	1.38	2.70	1.40	1
E. Michalewicz	5	1.00	1.35	1.07	1.01	1	1.01	1.08	1.00
E. Michalewicz	10	1.25	1.46	1.23	1.05	1.14	1	1.31	1.37
Griewank	10	5.29	1.52	1	1.65	1.98	3.12	4.58	4.12
Griewank	20	2.89	2.97	2.56	2.57	2.35	2.47	1.12	1
Levy and M. 2	10	99.45	12.25	1	2.08	2.55	4.89	5.39	6.58
Levy and M. 2	20	14.98	5.40	2.07	1	1.65	1.95	2.42	1.97
Paviani	10	1.00	2.94	1	1	1	1	1	1.03
Rosenbrock	10	14.66	1.39	1.11	1	1.52	1.01	4.83	3.97
Rosenbrock	20	27.79	1.24	1.09	1.34	1.93	1	10.61	4.32
Salomon	10	3.17	1.03	1.07	1.40	1.65	1	2.74	2.28
Salomon	20	2.27	1	1.19	1.17	1.34	1.01	1.85	1.59
Schwefel	10	9.78	11.68	9.72	4.44	4.08	10.77	4.11	1
Shekel Foxholes	10	1.02	1.02	1.02	1.01	1.01	1.01	1	1.01
Sinusoidal Cen.	10	3.25	1.37	2.36	1.91	1	1.76	2.11	2.01
Sinusoidal Cen.	20	1.85	1.02	1	1.28	1.06	1.17	1.44	1.31
Sinusoidal Shi.	10	2.07	1.92	1.74	1.80	1.53	2.06	1	1.25
Sinusoidal Shi.	20	1.60	1.52	1.37	1.30	1.37	1.41	1.01	1

biwalks in a more efficient way. In order to cover a wide range of real optimization problems, we also define Sphere and Box Biwalks for mixed continuous/discrete domains. We show that the new pattern biwalks converge to essentially any target distribution when they are used with a Metropolis filter. We provide the finite time performance for uniform distribution and Boltmann T distribution and show it is polynomial for special discrete domains. In addition, we prove that new candidate point generators are discrete and mixed analogies of continuous Hit-and-Run. We tested Sphere and Box Biwalks within a modified Improving Hit-and-Run setting on several global optimization problems and the numerical results support that they are successful against other sampling methods such as random biwalk, the Step Function approach and separating the discrete and continuous domains.

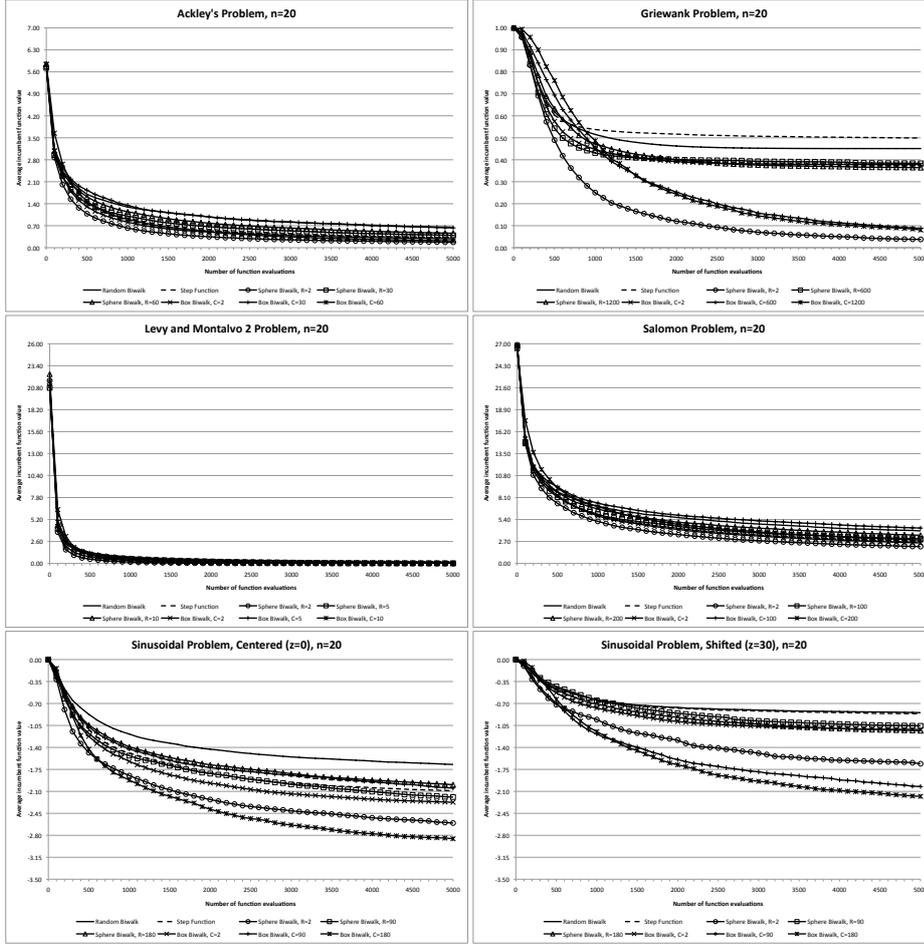


Fig. 1 Average Incumbent Values vs. Number of Iterations for Discrete Problems

Appendix

Proof (Theorem 3)

Consider Sphere Biwalk defined on the mixed continuous/discrete set $S_\delta \in H_\delta$. Given R, δ and the continuous unit direction $\vec{D} = (\vec{d}_1, \vec{d}_2, \dots, \vec{d}_n)$, let $V_{(R,\delta)}^{\vec{D}} = (v_{(R,\delta)1}^{\vec{D}}, \dots, v_{(R,\delta)n}^{\vec{D}})$ be the mixed continuous/discrete point generated in Step 1.2. Therefore we have

$$v_{(R,\delta)j}^{\vec{D}} = r\vec{d}_j \text{ if } j \text{ is a continuous dimension} \quad (19)$$

$$v_{(R,\delta)j}^{\vec{D}} = \text{Round}\left(\frac{r\vec{d}_j}{\delta}\right) \delta \text{ if } j \text{ is a discrete dimension,} \quad (20)$$

for some $r \in [-R, R]$, where $\text{Round}(\cdot)$ represents rounding to the nearest integer number. We also let $D_{(R,\delta)}^{\vec{D}} = \{D_{(R,\delta)1}^{\vec{D}}, \dots, D_{(R,\delta)n}^{\vec{D}}\}$ be the set of step length directions, as determined in Step 1.3.

Let $S_{D_{(R,\delta)}^{\vec{D}}}(x_0)$ be the set of mixed continuous/discrete set points contained in S_δ that are on at least one of the bidirectional paths generated from x_0 with $D_{(R,\delta)}^{\vec{D}}$ and any permutation of the dimensions. Let $S_{\vec{D}}(x_0)$

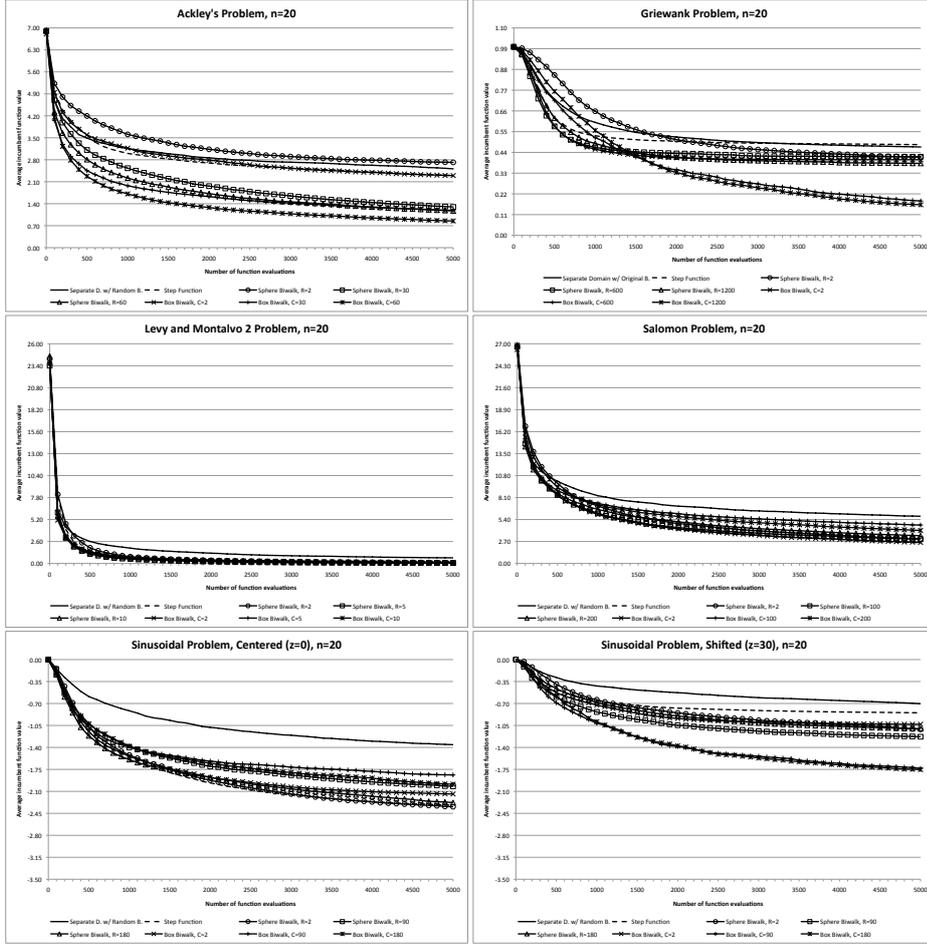


Fig. 2 Average Incumbent Values vs. Number of Iterations for Mixed Continuous/Discrete Problems

be a set of points generated by continuous Hit-and-Run from x_0 with the direction \vec{D} over \vec{S} , i.e.,

$$S_{\vec{D}}(x_0) = \{x \in \vec{S} \subseteq \vec{H} : x = x_0 + \beta \vec{D}, \beta \text{ is a real scalar}\}.$$

To prove (i), we first prove that $x \in \vec{S}$. Considering that $x_{(R,\delta)} \in S_{D_{(R,\delta)}}(x_0)$, we have $x_{(R,\delta)} \in S_{\delta}$. And according to the supposition that $\{x_{(R,\delta)}\} \rightarrow x$ as $\frac{\delta}{R} \rightarrow 0$ and $R \rightarrow 0$ and the fact $\vec{S} = \lim_{\delta \rightarrow 0} S_{\delta}$, we can claim that $x \in \vec{S}$.

We next prove that x is a point on the line with direction \vec{D} passing through x_0 , $\vec{L}_{\vec{D}}(x_0)$.

Given $r \in [-R, R]$, considering (19), we have

$$\frac{v_{(R,\delta)}^{\vec{D}}}{r} \rightarrow \vec{d}_j \text{ as } \frac{\delta}{R} \rightarrow 0 \text{ for continuous dimensions } j, \quad (21)$$

and

$$|v_{(R,\delta)}^{\vec{D}}| \rightarrow 0 \text{ as } R \rightarrow 0 \text{ for continuous dimension } j. \quad (22)$$

In addition, (20) implies,

$$\left(\frac{r\vec{d}_j}{\delta} - 1\right) \delta \leq v_{(R,\delta)}^{\vec{D}} \leq \left(\frac{r\vec{d}_j}{\delta} + 1\right) \delta \text{ for discrete dimensions } j, \quad (23)$$

which is equivalent to

$$\tilde{d}_j - \frac{\delta}{r} \leq \frac{v_{(R,\delta)}^{\tilde{D}}}{r} \leq \tilde{d}_j + \frac{\delta}{r} \text{ for discrete dimensions } j. \quad (24)$$

Combining (24) with the fact that r is a scalar with arbitrary value in $[-R, R]$ and (21), we conclude that

$$\frac{v_{(R,\delta)}^{\tilde{D}}}{r} \rightarrow \tilde{D} \text{ as } \frac{\delta}{R} \rightarrow 0. \quad (25)$$

Now let $I = \{I_1, I_2, \dots, I_n\}$ be any random permutation of n coordinate dimensions. Given x_0 , R , δ , \tilde{D} , $D_{(R,\delta)}^{\tilde{D}}$ and I , Sphere Biwalk generates a forward path and a backward path starting from x_0 until stepping out of H_δ . Without losing generality, assume the point $x_{(R,\delta)}$ is located on the forward path. Let $\mu_{R,\delta}$ be the total number of times the complete cycles of n coordinate dimensions, $\{I_1, I_2, \dots, I_n\}$, is repeated to generate the forward path from x_0 to $x_{(R,\delta)}$. Then we have

$$\mu_{R,\delta} \sum_{j=1}^n D_{(R,\delta)I_j}^{\tilde{D}} - \sum_{j=1}^n |D_{(R,\delta)I_j}^{\tilde{D}}| \leq x_{(R,\delta)} - x_0 \leq \mu_{R,\delta} \sum_{j=1}^n D_{(R,\delta)I_j}^{\tilde{D}} + \sum_{j=1}^n |D_{(R,\delta)I_j}^{\tilde{D}}|,$$

which is equivalent to

$$r\mu_{R,\delta} \frac{V_{(R,\delta)}^{\tilde{D}}}{r} - \tilde{V}_{(R,\delta)}^{\tilde{D}} \leq x_{(R,\delta)} - x_0 \leq r\mu_{R,\delta} \frac{V_{(R,\delta)}^{\tilde{D}}}{r} + \tilde{V}_{(R,\delta)}^{\tilde{D}}, \quad (26)$$

where $\tilde{V}_{(R,\delta)}^{\tilde{D}} = (|v_{(R,\delta)1}^{\tilde{D}}|, \dots, |v_{(R,\delta)n}^{\tilde{D}}|)$. Applying (22) and (25) and the fact that $\{x_{(R,\delta)}\} \rightarrow x$ as $\frac{\delta}{R} \rightarrow 0$ and $R \rightarrow 0$, we have as $\frac{\delta}{R} \rightarrow 0$ and $R \rightarrow 0$, $(\{x_{(R,\delta)}\} - x_0) \rightarrow (x - x_0) = (r\mu_{R,\delta})\tilde{D}$, and combining with the fact that $x \in \tilde{S}$, we can conclude that $x \in S_{\tilde{D}}(x_0)$ and $x \in \tilde{L}_{\tilde{D}}(x_0)$.

Next we prove (ii). We first choose a proper sequence of points $\{x_{(R,\delta)}\}$. Given R and δ , let $S_{D_{R,\delta}}^m$ be a set such that $S_{D_{R,\delta}}^m = \{x_{(R,\delta)}^m \in S_\delta : x_{(R,\delta)}^m = mV_{(R,\delta)}^{\tilde{D}} + x_0, \text{ for } m = \{\dots, -1, 0, 1, \dots\}\}$. It is obvious that $S_{D_{R,\delta}}^m \subset S_\delta$. Choose $x_{(R,\delta)}$ to be the point in $S_{D_{R,\delta}}^m$ which is closest to point x , i.e., $x_{(R,\delta)} \in S_{D_{R,\delta}}^m$ such that $\|x_{(R,\delta)} - x\| \leq \|x_{(R,\delta)}^m - x\|$ for all $x_{(R,\delta)}^m \in S_{D_{R,\delta}}^m$. Let \hat{m} be the integer associated with $x_{(R,\delta)}$, i.e., $x_{(R,\delta)} = \hat{m}V_{(R,\delta)}^{\tilde{D}} + x_0$, where $\hat{m} \in \{\dots, -1, 0, 1, \dots\}$. Then we have

$$x_{(R,\delta)} - x = \hat{m}V_{(R,\delta)}^{\tilde{D}} + x_0 - x.$$

Considering that $x \in \tilde{L}_{\tilde{D}}(x_0)$, we have $x = \beta_x \tilde{D} + x_0$ for some $\beta_x \in \mathbb{R}$. Therefore

$$x_{(R,\delta)} - x = \hat{m}V_{(R,\delta)}^{\tilde{D}} - \beta_x \tilde{D}.$$

And considering (19) and (23), which can be combined as

$$\frac{v_{(R,\delta)j}^{\tilde{D}} - \delta}{r} \leq \tilde{d}_j \leq \frac{v_{(R,\delta)j}^{\tilde{D}} + \delta}{r}, \text{ for all } j = 1, \dots, n,$$

and without losing generality, we assume $\beta_x \geq 0$, and then we have

$$\hat{m}v_{(R,\delta)j}^{\tilde{D}} - \beta_x \frac{v_{(R,\delta)j}^{\tilde{D}} + \delta}{r} \leq x_{(R,\delta)j} - x_j \leq \hat{m}v_{(R,\delta)j}^{\tilde{D}} - \beta_x \frac{v_{(R,\delta)j}^{\tilde{D}} - \delta}{r}$$

for $j = 1, \dots, n$, which is equivalent to

$$\left(\hat{m} - \frac{\beta_x}{r}\right)v_{(R,\delta)j}^{\tilde{D}} - \beta_x \frac{\delta}{r} \leq x_{(R,\delta)j} - x_j \leq \left(\hat{m} - \frac{\beta_x}{r}\right)v_{(R,\delta)j}^{\tilde{D}} + \beta_x \frac{\delta}{r} \quad (27)$$

for $j = 1, \dots, n$. And considering the fact that $S_\delta \rightarrow \tilde{S}$ as $\frac{\delta}{R} \rightarrow 0$ and $R \rightarrow 0$, there exists small enough ε such that for all $\frac{\delta}{R} < \varepsilon$ and $R < \varepsilon$, \hat{m} satisfies $(\hat{m} - \frac{\beta_x}{r}) < \infty$ and $(\hat{m}v_{(R,\delta)}^{\tilde{D}} + x_0) \in S_\delta$, which implies $(\hat{m}v_{(R,\delta)}^{\tilde{D}} + x_0) \in S_{D_{R,\delta}}^m$. Now considering (27) and applying the fact that $(\hat{m} - \frac{\beta_x}{r}) < \infty$ and applying the fact that as $\frac{\delta}{R} \rightarrow 0$ and $R \rightarrow 0$, $|v_{(R,\delta)j}^{\tilde{D}}| \rightarrow 0$ for $j = 1, \dots, n$, we have $(x_{(R,\delta)} - x) \rightarrow 0$ as $\frac{\delta}{R} \rightarrow 0$ and $R \rightarrow 0$, i.e., $x_{(R,\delta)} \rightarrow x$ as $\frac{\delta}{R} \rightarrow 0$ and $R \rightarrow 0$. ■

The following lemma is used to prove Lemma 5 and Lemma 6. The lemma states that any bidirectional path generated in a hyperrectangle H by Sphere Biwalk or Box Biwalk needs no more than $1 + nL$ steps.

Lemma 7 For any bidirectional path

$$W^x = \{W_{-m_2}^x, W_{-m_2+1}^x, \dots, W_{-1}^x, W_0^x, W_1^x, \dots, W_{m_1-1}^x, W_{m_1}^x\}$$

generated by Sphere Biwalk or Box Biwalk over a hyperrectangle H with

$$H = \{(x_1, x_2, \dots, x_n) \in \mathbb{Z}^n : l_i \leq x_i \leq u_i, i = 1, \text{ with } l_i < u_i, i = \dots, n\},$$

the length of bidirectional path is no more than $1 + nL$, i.e., $m_1 + m_2 + 1 \leq 1 + nL$, where $L = \max_{i=1, \dots, n} (u_i - l_i)$.

Proof Let $\{D_1, \dots, D_k, \dots, D_n\}$ be the set of discrete directions generated by Sphere Biwalk or Box Biwalk associated with the continuous direction \vec{D} , and let $V^{\vec{D}} = (v_1^{\vec{D}}, \dots, v_k^{\vec{D}}, \dots, v_n^{\vec{D}})$ be the corresponding discrete point.

First consider the forward path $\{W_0^x, W_1^x, \dots, W_{m_1-1}^x, W_{m_1}^x\}$. Let m_1^i be the number of forward path steps that have been taken along coordinate i , where $i = 1, 2, \dots, n$. Then

$$\sum_{i=1}^n m_1^i = m_1.$$

For the ending point $W_{m_1}^x = (W_{m_1,1}^x, \dots, W_{m_1,n}^x)$ and the starting point $W_{0,1}^x = (W_{0,1}^x, \dots, W_{0,n}^x)$ on the forward path, we have

$$W_{m_1,i}^x = W_{0,i}^x + m_1^i v_i^{\vec{D}}, \text{ for } i = 1, 2, \dots, n.$$

Similarly, Let m_2^i be the number of backward path steps that have been taken along coordinate i , where $i = 1, 2, \dots, n$. Then

$$\sum_{i=1}^n m_2^i = m_2.$$

For the ending point $W_{-m_2}^x = (W_{-m_2,1}^x, \dots, W_{-m_2,n}^x)$ on the backward path and the starting point $W_{0,1}^x$, we have

$$W_{-m_2,i}^x = W_{0,i}^x - m_2^i v_i^{\vec{D}}, \text{ for } i = 1, 2, \dots, n.$$

Therefore, we have

$$|W_{m_1,i}^x - W_{-m_2,i}^x| = |(m_1^i + m_2^i) v_i^{\vec{D}}|, \text{ for } i = 1, 2, \dots, n,$$

and considering that $l_i \leq W_{m_1,i}^x, W_{-m_2,i}^x \leq u_i$ for $i = 1, 2, \dots, n$, we have

$$|W_{m_1,i}^x - W_{-m_2,i}^x| \leq u_i - l_i, \text{ for } i = 1, 2, \dots, n,$$

and hence

$$m_1^i + m_2^i \leq \frac{u_i - l_i}{|v_i^{\vec{D}}|}$$

if $|v_i^{\vec{D}}| \geq 1$; and when $v_i^{\vec{D}} = 0$, $m_1^i + m_2^i = 0$. Therefore,

$$m_1^i + m_2^i \leq (u_i - l_i) \text{ for all } i = 1, 2, \dots, n$$

Then,

$$\begin{aligned} m_1 + m_2 &= \sum_{i=1}^n (m_1^i + m_2^i) \\ &\leq \sum_{i=1}^n (u_i - l_i) \\ &\leq nL, \end{aligned}$$

where $L = \max_{i=1, \dots, n} (u_i - l_i)$, and considering that the length of the bidirectional path W^x is $m_1 + m_2 + 1$, we have $m_1 + m_2 + 1 \leq 1 + nL$. ■

Proof (Lemma 5)

Let $\{D_1, \dots, D_n\}$ be the set of discrete directions generated by Sphere Biwalk associated with the continuous direction \bar{D} , and let $V^{\bar{D}} = (v_1^{\bar{D}}, \dots, v_n^{\bar{D}})$ be the corresponding discrete point generated by one iteration of Hit-and-Run with rounding from the center of an n -dimensional ball with radius R .

Let $\{I_1, \dots, I_n\}$ be a random permutation of n coordinate dimensions. Given two adjacent points $x, z \in S$, define the event A_x^z as,

$$A_x^z = \begin{cases} 1 & \text{if the algorithm generates a bidirectional path from } x \text{ going through } z \\ 0 & \text{if the algorithm generates a bidirectional path from } x \text{ without going through } z. \end{cases}$$

Recall that the Markov chain associated with Sphere Biwalk for candidate points consists of two main steps to generate a candidate point from x , generating a bidirectional path from x and picking a point uniformly along the path. The one step transition probability q_{xz} can be written as

$$q_{xz} = P(Z = z | A_x^z = 1) P(A_x^z = 1).$$

Considering that each bidirectional path generated from x can reach z by either the forward path or the backward path, and according to the process of the algorithm, it is impossible for both of them to go through z , we have,

$$P(A_x^z = 1) = P(A_x^z = 1 \text{ by the forward path}) + P(A_x^z = 1 \text{ by the backward path}).$$

First consider the case that the path passes through z by the forward path. Considering that x and z are adjacent, without losing generality, let $x = (x_1, \dots, x_{j-1}, x_j, x_{j+1}, \dots, x_n)$, $z = (x_1, \dots, x_{j-1}, x_j + 1, x_{j+1}, \dots, x_n)$. Obviously a path generated from x passes through z by the forward path if and only if the j^{th} entry of D_j equals one and the j^{th} dimension is the first nonzero dimension that Sphere Biwalk chooses to move to during a forward path from x , i.e., $v_k^{\bar{D}} = 1$ and if $I_k = j$, then $v_m^{\bar{D}} = 0$ for all $m = 1, 2, \dots, k-1$. To simplify the notation, for $i = 1, 2, \dots, n$, define the event a_i such that if the i^{th} dimension is the first nonzero dimension that Sphere Biwalk chooses to move to during a forward path from x , then $a_i = 1$, i.e.,

$$a_i = \begin{cases} 1 & \text{if } v_{I_1}^{\bar{D}} = \dots = v_{I_{k-1}}^{\bar{D}} = 0, I_k = i, \text{ for } k = 1, \dots, n, \text{ and } |v_i^{\bar{D}}| \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

Without loss of generality, we take $v_i^{\bar{D}} \geq 1$, when $a_i = 1$. Therefore, $P(A_x^z = 1 \text{ by forward path}) = P(v_j^{\bar{D}} = 1 \text{ and } a_j = 1) = P(a_j = 1 | v_j^{\bar{D}} = 1) P(v_j^{\bar{D}} = 1)$. By the definition of a_j , we have

$$P(a_j = 1 | v_j^{\bar{D}} = 1) \geq P(I_1 = j | v_j^{\bar{D}} = 1) \quad (28)$$

considering that the total number of nonzero-step length dimension is no more than n , we have

$$P(a_j = 1 | v_j^{\bar{D}} = 1) \geq \frac{1}{n}. \quad (29)$$

Thus, $P(A_x^z = 1 \text{ by forward path}) \geq \frac{1}{n} P(v_j^{\bar{D}} = 1)$. Conditioning the probability $P(v_j^{\bar{D}} = 1)$ on $v_j^{\bar{D}}$, since $P(v_j^{\bar{D}} = 1 | v_j^{\bar{D}} = 0) = 0$, we get,

$$P(A_x^z = 1 \text{ by forward path}) \geq \frac{1}{n} P(v_j^{\bar{D}} = 1 | v_j^{\bar{D}} \geq 1) P(v_j^{\bar{D}} \geq 1) \quad (30)$$

Next we will discuss $P(v_j^{\bar{D}} \geq 1)$ and $P(v_j^{\bar{D}} = 1 | v_j^{\bar{D}} \geq 1)$ respectively.

Recall that $a_j = 1$ represents the event that the first nonzero step length dimension is the j^{th} dimension, which implies that the discrete direction $v_j^{\bar{D}} \neq 0$. Thus one has,

$$P(v_j^{\bar{D}} \geq 1) \geq P(a_j = 1). \quad (31)$$

Hence if we can find a lower bound on the probability $P(a_j = 1)$, it will provide a lower bound on $P(v_j^{\bar{D}} \geq 1)$.

By conditioning the probability $P(a_j = 1)$ on $v_j^{\bar{D}}$, we get

$$P(a_j = 1) = P(a_j = 1 | v_j^{\bar{D}} \geq 1) P(v_j^{\bar{D}} \geq 1) + P(a_j = 1 | v_j^{\bar{D}} = 0) \cdot P(v_j^{\bar{D}} = 0) \quad (32)$$

$$= P(a_j = 1 | v_j^{\bar{D}} \geq 1) P(v_j^{\bar{D}} \geq 1). \quad (33)$$

What is the value of $P(a_j = 1 | v_j^{\bar{D}} \geq 1)$? Intuitively, this value should be the same as $P(a_i = 1 | v_i^{\bar{D}} \geq 1)$ for all $i = 1, 2, \dots, n$. To show this, for $k = 1, 2, \dots, n$, define a set $S_{v_k^{\bar{D}}}$ to be a collection of discrete points as follows,

$$S_{v_k^D} = \{V^D = (v_1^D, \dots, v_k^D, \dots, v_n^D) : v_k^D \geq 1 \text{ and } \sum_{i=1}^n (v_i^D)^2 \leq R^2\}.$$

Note that $S_{v_k^D}$ is a finite set. Now consider $\forall i, j \in \{1, 2, \dots, n\}$ and $i \neq j$. Without losing generality, assume that $i > j$. Let $V^D = (v_1^D, \dots, v_{i-1}^D, v_i^D, v_{i+1}^D, \dots, v_{j-1}^D, v_j^D, v_{j+1}^D, \dots, v_n^D)$ be any discrete point in set $S_{v_i^D}$. By exchanging v_i^D and v_j^D , we get another discrete point $\tilde{V}^D = (v_1^D, \dots, v_{i-1}^D, \tilde{v}_i^D = v_j^D, v_{i+1}^D, \dots, v_{j-1}^D, \tilde{v}_j^D = v_i^D, v_{j+1}^D, \dots, v_n^D)$. Obviously, $\tilde{V}^D \in S_{v_j^D}$. On the other hand, for any $\tilde{V}^D = (\tilde{v}_1^D, \dots, \tilde{v}_i^D, \dots, \tilde{v}_j^D, \dots, \tilde{v}_n^D) \in S_{v_j^D}$, one has $V^D = (v_1^D, \dots, v_i^D = \tilde{v}_j^D, \dots, v_j^D = \tilde{v}_i^D, \dots, v_n^D) \in S_{v_i^D}$. Hence for any such pairs of V^D and \tilde{V}^D , $V^D \in S_{v_i^D}$ iff $\tilde{V}^D \in S_{v_j^D}$, and for any discrete point in set $S_{v_i^D}$ there is a corresponding discrete point in set $S_{v_j^D}$ and visa versa. Given such a pair of V^D and \tilde{V}^D , where $V^D \in S_{v_i^D}$, $\tilde{V}^D \in S_{v_j^D}$ and is obtained from V^D by exchanging the position of v_i^D and v_j^D in V^D , let $m(V^D)$ denote the total number of nonzero coordinates of the point V^D and let $m(\tilde{V}^D)$ denote the total number of nonzero coordinates of the point \tilde{V}^D . Obviously, $m(V^D) = m(\tilde{V}^D)$. Hence,

$$P(a_i = 1 | V^D) = \frac{1}{m(V^D)!} = \frac{1}{m(\tilde{V}^D)!} = P(a_j = 1 | \tilde{V}^D),$$

which implies

$$P(a_i = 1 | v_i^D \geq 1) = \sum_{V^D \in S_{v_i^D}} P(a_i = 1 | V^D) \quad (34)$$

$$= \sum_{\tilde{V}^D \in S_{v_j^D}} P(a_j = 1 | \tilde{V}^D) \quad (35)$$

$$= P(a_j = 1 | v_j^D \geq 1). \quad (36)$$

Next, consider the probability $P(v_j^D \geq 1)$. Recall that the discrete direction V^D is generated by operating one iteration of Hit-and-Run algorithm with rounding in the hypersphere domain with radius R and with the origin O as the center of the hypersphere, we have,

$$P(v_j^D \geq 1) = \int_{A(v_j^D \geq 1)} f(u|O) du,$$

where $A(v_j^D \geq 1)$ denotes the continuous subset of the hypersphere domain such that any point in $A(v_j^D \geq 1)$ will round to a discrete point $V^D = (v_1^D, \dots, v_n^D)$ with $v_j^D \geq 1$, i.e., $V^D \in S_{v_j^D}$, and $f(z|O)$ is the density function with the form [25],

$$f(u|O) = \frac{1}{S_n(r_u)R_r},$$

where r_u is the distance from u to the origin O , $S_n(r_u)$ denotes the surface of hypersphere with radius r_u and $R_r = \text{Round}(R)$. By the observation that the density function $f(z|O)$ just depends on the distance between O and z and the fact that the hypersphere domain is symmetric, we claim that,

$$P(v_j^D \geq 1) = \int_{A(v_j^D \geq 1)} f(z|O) dy = \int_{A(v_i^D \geq 1)} f(z|O) dy = P(v_i^D \geq 1). \quad (37)$$

By applying equations (34) and (37) to equation (32), we get that for any $i, j \in \{1, 2, \dots, n\}$ and $i \neq j$, one has

$$P(a_j = 1) = P(a_j = 1 | v_j^D \geq 1) P(v_j^D \geq 1) = P(a_i = 1 | v_i^D \geq 1) P(v_i^D \geq 1) = P(a_i = 1).$$

Combining this with the fact that $\sum_{j=1}^n P(a_j = 1) = 1$, we conclude that for any $j \in \{1, 2, \dots, n\}$, $P(a_j = 1) = \frac{1}{n}$. Thus, by applying the conclusion to the (31), we have,

$$P(v_j^D \geq 1) \geq \frac{1}{n}. \quad (38)$$

The next step is to find the lower bound of $P(v_j^D = 1 | v_j^D \geq 1)$. It can be done by proving that, for any $k_1, k_2 \in \{1, 2, \dots, R_r\}$, if $k_1 < k_2$, $P(v_j^D = k_1 | v_j^D \geq 1) > P(v_j^D = k_2 | v_j^D \geq 1)$. The proof is as follows. For

$k \in \{1, 2, \dots, \text{Round}(R)\}$, let $A(v_j^{\bar{D}} = k)$ be the continuous subset of the hypersphere domain with radius R such that any point in $A(v_j^{\bar{D}} = k)$ will round to a discrete point $V^{\bar{D}} = (v_1^{\bar{D}}, \dots, v_n^{\bar{D}})$ with $v_j^{\bar{D}} = k$. Consider $k_1, k_2 \in \{1, 2, \dots, \text{Round}(R)\}$ such that $k_1 < k_2$. Let $y = \{y_1, \dots, y_j, \dots, y_n\}$ be any point in $A(v_j^{\bar{D}} = k_2)$, i.e., $\text{Round}(y_j) = k_2$. one has for any point Consider the point $y' = \{y_1, \dots, y'_j = y_j - (k_2 - k_1), \dots, y_n\}$, i.e., $\text{Round}(y'_j) = k_1$. Since y is a point in the hypersphere with radius R and $k_1 < k_2$, which implies y' is also a point in the same hypersphere, and therefore, $y' \in A(v_j^{\bar{D}} = k_1)$. Since $r'_y < r_y$, one has $S_n(r'_y) < S_n(r_y)$. Hence,

$$\begin{aligned} P(v_j^{\bar{D}} = k_2 | v_j^{\bar{D}} \geq 1) &= \frac{P(v_j^{\bar{D}} = k_2)}{P(v_j^{\bar{D}} \geq 1)} \\ &= \frac{\int_{A(v_j^{\bar{D}} = k_2)} \frac{1}{S_n(r_y)R_r} dy}{P(v_j^{\bar{D}} \geq 1)} \\ &< \frac{\int_{A(v_j^{\bar{D}} = k_1)} \frac{1}{S_n(r'_y)R_r} dy'}{P(v_j^{\bar{D}} \geq 1)} \\ &= \frac{P(v_j^{\bar{D}} = k_1)}{P(v_j^{\bar{D}} \geq 1)} \\ &= P(v_j^{\bar{D}} = k_1 | v_j^{\bar{D}} \geq 1) \end{aligned}$$

Thus, $P(v_j^{\bar{D}} = 1 | v_j^{\bar{D}} \geq 1) > P(v_j^{\bar{D}} = 2 | v_j^{\bar{D}} \geq 1) > \dots > P(v_j^{\bar{D}} = R_r | v_j^{\bar{D}} \geq 1)$. Considering the fact that $\sum_{k=1}^{R_r} P(v_j^{\bar{D}} = \pm k | v_j^{\bar{D}} \geq 1) = 1$, and $P(v_j^{\bar{D}} = k | v_j^{\bar{D}} \geq 1) = P(v_j^{\bar{D}} = -k | v_j^{\bar{D}} \geq 1)$ for $k = 1, 2, \dots, R_r$, one has,

$$P(v_j^{\bar{D}} = 1 | v_j^{\bar{D}} \geq 1) = P(v_j^{\bar{D}} = -1 | v_j^{\bar{D}} \geq 1) > \frac{1}{2R_r}. \quad (39)$$

By applying (38) and (39) to (30), we get,

$$P(A_x^z = 1 \text{ by forward path}) \geq \frac{1}{R_r 2n^2}$$

Similarly, we can prove that $P(A_x^z = 1 \text{ by backward path}) \geq \frac{1}{R_r 2n^2}$. Thus,

$$\begin{aligned} P(A_x^z = 1) &= P(A_x^z = 1 \text{ by forward path}) + P(A_x^z = 1 \text{ by backward path}) \\ &\geq \frac{1}{R_r n^2} \end{aligned}$$

Hence for any two adjacent points $x, z \in S$, one has $q_{xz} = P(Z = z | A_x^z = 1)P(A_x^z = 1)$, which is greater than equal to

$$P(Z = z | A_x^z = 1) \frac{1}{R_r n^2}.$$

And applying Lemma 7 which proving an upper bound on the number of candidate points along a bidirectional path, we have

$$\begin{aligned} q_{zx} &\geq \frac{1}{(1+nL)} \frac{1}{R_r n^2} \\ &\quad \text{and considering that } R_r = \text{Round}(R) \leq R+1, \text{ we have} \\ &\geq \frac{1}{(1+nL)} \frac{1}{(R+1)n^2} \end{aligned}$$

where $L = \max_{i=1, \dots, n} (u_i - l_i)$. ■

Proof (Lemma 6)

Given two adjacent points $x, z \in S$, without losing generality, let $x = (x_1, \dots, x_j, \dots, x_n)$ and $z = (x_1, \dots, x_j + 1, \dots, x_n)$. Define the event A_x^z as,

$$A_x^z = \begin{cases} 1 & \text{if the algorithm generates a bidirectional path from } x \text{ going through } z \\ 0 & \text{if the algorithm generates a bidirectional path from } x \text{ without going through } z. \end{cases}$$

As in the proof of Lemma 5, each bidirectional path generated by Box Biwalk from x can reach z by either the forward path or the backward path, and it is impossible for both of them to go through z . Thus, according to the process of Box Biwalk, if the j^{th} entry of the n -dimensional vector D_j is not equal to one, there is no chance for the forward path starting from x to reach z . The probability of this step-direction choice is $\frac{1}{2c_j}$. Given the j^{th} entry of D_j is one, the point z will be reached by the forward path from x if the permutation of n coordinate dimensions, $I = \{I_1, \dots, I_n\}$, satisfies $I_1 = j$. The probability of choosing $I_1 = j$ is $1/n$. Thus the probability of generating a forward path from x that reaches z is bounded by $\frac{1}{2nc_j}$. By the same argument, it is also true that the probability of generating a backward path from x that reaches z is bounded by $\frac{1}{2nc_j}$. And considering that $c_j \leq u_j - l_j \leq L$, where $L = \max_{i=1, \dots, n} (u_i - l_i)$, we have $P(A_x^z = 1) \geq \frac{1}{nL}$. Therefore

$$q_{xz} = P(Z = z | A_x^z = 1) P(A_x^z = 1)$$

$$\geq P(Z = z | A_x^z = 1) \frac{1}{nL}$$

and applying Lemma 7 which provides an upper bound on the number of candidate points along a bidirectional path, we have

$$\geq \frac{1}{(1+nL)} \frac{1}{nL}.$$

■

References

1. M. Montaz Ali, Charoenchai Khompatraporn, and Zelda B. Zabinsky. A numerical evaluation of several stochastic algorithms on selected continuous global optimization test problems. *Journal of Global Optimization*, 31(4):635–672, 2005.
2. Hans C. Andersen and Persi Diaconis. Hit and run as a unifying device. *Journal de la societe francaise de statistique & revue de statistique appliquee*, 148(4):5–28, 2007.
3. David L. Applegate. *The traveling salesman problem: a computational study*. Princeton University Press, Princeton, 2006.
4. Stephen Baumert, Archis Ghate, Seksan Kiatsupaibul, Yanfang Shen, Robert L. Smith, and Zelda B. Zabinsky. Discrete hit-and-run for sampling points from arbitrary distributions over subsets of integer hyper-rectangles. *Operations Research*, forthcoming.
5. Ehrhard Behrends. *Introduction to Markov chains: with special emphasis on rapid mixing*. Vieweg, Braunschweig/Wiesbaden, 2000.
6. Claude J. P. Belisle, H. Edwin Romeijn, and Robert L. Smith. Hit-and-Run algorithms for generating multivariate distributions. *Mathematics of Operations Research*, 18(2):255–266, 1993.
7. Dimitris Bertsimas and Santosh Vempala. Solving convex programs by random walks. *Journal of the ACM*, 51(4):0–556, 2004.
8. Persi Diaconis and Daniel Stroock. Geometric bounds for eigenvalues of Markov chains. *The Annals of Applied Probability*, 1(1):36–61, 1991.
9. Elizabeth D. Dolan and Jorge J. Morè. Benchmarking optimization software with performance profiles. *Mathematical Programming*, 91(2):201–213, 2002.
10. Keld Helsgaun. An effective implementation of the Lin-Kernighan traveling salesman heuristic. *European Journal of Operational Research*, 126(1):106–130, 2000.
11. Adam Tauman Kalai and Santosh Vempala. Simulated annealing for convex optimization. *Mathematics of Operations Research*, 31(2):253–266, 2006.
12. Birna P. Kristinsdottir, Zelda B. Zabinsky, Mark E. Tuttle, and Sudipto Neogi. Optimal design of large composite panels with varying loads. *Composite Structures*, 51(1):93–102, 2001.
13. László Lovász. Hit-and-run mixes fast. *Mathematical Programming*, 86(3):443–461, 1999.

14. László Lovász and Santosh Vempala. Simulated annealing in convex bodies and an $O^*(n^4)$ volume algorithm. *Proceedings of Symposium on Foundations of Computer Science*, 44:650–659, 2003.
15. László Lovász and Santosh Vempala. Hit-and-Run from a corner. *SIAM Journal on Computing*, 35(4): 985–1005, 2006.
16. Orcun Molvalioglu, Zelda B. Zabinsky, and Wolf Kohn. The interacting-particle algorithm with dynamic heating and cooling. *Journal of Global Optimization*, forthcoming.
17. Ravi R. Montenegro and Prasad Tetali. *Mathematical aspects of mixing times in Markov chains*. Now Publishers, Hanover, MA, 2005.
18. János D. Pintér. *Global optimization: scientific and engineering case studies*. Springer, New York, 2006.
19. H. Edwin Romeijn and Robert L. Smith. Simulated annealing for constrained global optimization. *Journal of Global Optimization*, 5(2):101–126, 1994.
20. H. Edwin Romeijn, Zelda B. Zabinsky, Douglas L. Graesser, and Sudipto Neogi. New reflection generator for simulated annealing in mixed-integer/continuous global optimization. *Journal of Optimization Theory and Applications*, 101(2):403–427, 1999.
21. Sheldon M. Ross. *Stochastic processes*. Wiley, New York, 1996.
22. Vesna Savic, Mark E. Tuttle, and Zelda B. Zabinsky. Optimization of composite I-sections using fiber angles as design variables. *Composite Structures*, 53(3):265–277, 2001.
23. Yanfang Shen, Seksan Kiatsupaibul, Zelda B. Zabinsky, and Robert L. Smith. An analytically derived cooling schedule for simulated annealing. *Journal of Global Optimization*, 38(3):333–365, 2007.
24. Alistair Sinclair. Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combinatorics, Probability and Computing*, 1(04):351–370, 1992.
25. Robert L. Smith. Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions. *Operations Research*, 32(6):1296–1308, 1984.
26. James C. Spall. *Introduction to stochastic search and optimization: estimation, simulation, and control*. Wiley-Interscience, Hoboken, N.J., 2003.
27. Gunther Stattenberger, Markus Dankesreiter, Florian Baumgartner, and Johannes Schneider. On the neighborhood structure of the traveling salesman problem generated by local search moves. *Journal of Statistical Physics*, 129(4):623–648, 2007.
28. Zelda B. Zabinsky. Stochastic methods for practical global optimization. *Journal of Global Optimization*, 13(4):433–444, 1998.
29. Zelda B. Zabinsky. *Stochastic adaptive search for global optimization*. Kluwer Academic Publishers, Boston, 2003.
30. Zelda B. Zabinsky, Robert L. Smith, J. Fred McDonald, H. Edwin Romeijn, and David E. Kaufman. Improving Hit-and-Run for global optimization. *Journal of Global Optimization*, 3(2):171–192, 1993.