J Glob Optim (2007) 38:333–365
DOI 10.1007/s10898-006-9068-2

ORIGINAL PAPER

An analytically derived cooling schedule for simulated annealing

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

Received: 21 February 2006 / Accepted: 1 July 2006 / Published online: 4 October 2006

Abstract We present an analytically derived cooling schedule for a simulated annealing algorithm applicable to both continuous and discrete global optimization problems. An adaptive search algorithm is used to model an idealized version of simulated annealing which is viewed as consisting of a series of Boltzmann distributed sample points. Our choice of cooling schedule ensures linearity in the expected number of sample points needed to become arbitrarily close to a global optimum.

Keywords Simulated annealing · Cooling schedule · Adaptive search

1 Introduction

Simulated annealing is a stochastic method for searching for global optima of discrete and continuous optimization problems [1, 7, 8, 11]. The origination of simulated annealing is from an analogy with the physical annealing process of finding low energy states of a solid in a heat bath [14]. The algorithm avoids getting trapped in local optima by allowing moves that may lead to a deterioration in objective function value. The probability of accepting a worse candidate point is controlled by a time-dependent temperature parameter, which descends asymptotically to zero in the course of the

Y. Shen
Industrial Engineering Program, University of Washington, Seattle, WA 98195-2650, USA

S. Kiatsupaibul
Department of Statistics, Chulalongkorn University, Bangkok 10330, Thailand

Z. B. Zabinsky (✉)
Industrial Engineering Program, University of Washington, Seattle, WA 98195-2650, USA
e-mail: zelda@u.washinton.edu

R. L. Smith
Department of Industrial and Operations Engineering, The University of Michigan,
Ann Arbor, MI 48109, USA

optimization process. From both practical and theoretical points of view, the cooling schedule for the temperature plays an important role in simulated annealing. Various choices for cooling schedule have been proposed and computationally tested in the literature [3, 6, 9]. Moreover, sufficient conditions for many of these cooling schedules have been established that guarantee convergence to the global optimum [4, 12, 13, 18]. Most conditions simply require the cooling not be too rapid but are otherwise not specific in specifying what rates might lead to the fastest convergence to optimal. Indeed, for globally reaching Markov chain samplers like Hit-and-Run [18, 19], any cooling schedule converging to zero at any rate guarantees convergence to optimal in the limit. In this paper, we propose an analytic cooling schedule which leads to a linear in dimension number of sample points.

2 Adaptive search for continuous and discrete problems

In this paper, we derive an adaptive cooling schedule for simulated annealing applicable to both continuous and discrete optimization problems. The cooling schedule strategies are developed based on the adaptive search algorithm introduced by Romeijn and Smith [19]. Adaptive search (AS) is an ideal algorithm that models simulated annealing by assuming points can be sampled according to a sequence of Boltzmann distributions. An attractive property of AS is that the expected number of record values generated by the algorithm increase at most linearly in dimension of the problem for a large class of continuous/discrete optimization problems [19, 23]. In addition, a natural choice of cooling schedule for AS was derived in [19], which maintained the linearity result on the expected number of sample points required for AS for solving a class of convex problems with a quadratic type of objective function. In this paper, we develop a cooling schedule for AS applied to the more general class of Lipschitz objective functions over both continuous and discrete domains. In light of the difficulty of generating Boltzmann distributed points with low temperature values, the cooling schedule attempts to keep the temperature of each AS iteration as high as possible while maintaining a linear complexity in expected number of sample points.

We consider two global optimization programs, one with a continuous domain and one with a discrete domain. The continuous problem is

$$\textbf{(P1)} \quad \max f(x) \quad \text{s.t. } x \in S,$$

where S is a convex, compact, and full dimensional subset of $\Re^n$, and the objective function f is a real-valued continuous function defined over S. Let ρ denote the diameter of S. Let x^* be an optimal solution of (P1) and f^* be the global optimum, $f^* = \max_{x \in S} f(x)$. We assume that f satisfies the *Lipschitz condition* with Lipschitz constant K, i.e.,

$$|f(x) - f(y)| \leq K\|x - y\|, \quad \forall x, y \in S, \tag{1}$$

where $\|\cdot\|$ is the Euclidean norm on $\Re^n$.

The discrete problem is

$$\textbf{(P2)} \quad \max \tilde{f}(\tilde{x}) \quad \text{s.t. } \tilde{x} \in \tilde{S},$$

where $\tilde{S}$ is the collection of integer points contained in an n-dimensional hyperrectangle $[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$, where $a_i \leq b_i$ and $a_i, b_i \in Z$ for $i = 1, \ldots, n$. Let $\tilde{\rho}$ denote the largest width of the hyperrectangle, i.e., $\tilde{\rho} = \max_{i=1,\ldots,n}(b_i - a_i)$. Let $\tilde{x}^*$ be an optimal solution of (P2) and $\tilde{f}^*$ be the global optimum, $\tilde{f}^* = \max_{\tilde{x}\in\tilde{S}} \tilde{f}(\tilde{x})$. Notice that a condition analogous to the Lipschitz condition is satisfied for (P2), i.e.

$$|\tilde{f}(\tilde{x}) - \tilde{f}(\tilde{y})| \leq \tilde{K} \max_{i=1,\ldots,n} (|\tilde{x}_i - \tilde{y}_i|), \quad \forall \tilde{x}, \tilde{y} \in \tilde{S}, \tag{2}$$

where $\tilde{K}$ is a positive constant and $\tilde{x} = (\tilde{x}_1, \ldots, \tilde{x}_n)$, $\tilde{y} = (\tilde{y}_1, \ldots, \tilde{y}_n)$. Note that because $\tilde{S}$ is a finite set, a positive constant $\tilde{K}$ exists, with $\tilde{K} \leq \tilde{f}^* - \tilde{f}_*$, where $\tilde{f}_* = \min_{\tilde{x}\in\tilde{S}} \tilde{f}(\tilde{x})$.

We center our attention on adaptive search, introduced in [19], to solve (P1) and (P2). In [19], adaptive search models simulated annealing by assuming points can be exactly sampled according to a sequence of Boltzmann distributions. An attractive property of adaptive search, as shown by Romeijn and Smith in 1994, is that the expected number of record values required for adaptive search to solve (P1) over the continuous space S grows linearly in the dimension of the problem [19, Theorem 1]. In 2003, Shen et al. [20] (see also [21, 23]) extended the linearity result to a finite domain within an n-dimensional lattice.

While this is an encouraging result, neither Romeijn and Smith [19] nor Shen et al. [20] offer algorithms for computing the cooling schedule whose existence guarantees the linearity result. In this paper, we derive formulas for an analytic cooling schedule guaranteed to result in at most a linear number of temperature changes. There is still a gap between theory and practice in that adaptive search still assumes that points can be sampled according to a Boltzmann distribution of arbitrary parameter T. However, the analytical linearity result provides motivation to develop algorithms that can approximate the theoretical performance. If a procedure were to be discovered to sample efficiently from the corresponding Boltzmann distributions, adaptive search would thereby be efficiently implementable through use of this cooling schedule.

2.1 The adaptive search algorithm

Adaptive search [19] is motivated by the idea of approximating the global optimum by generating points according to a sequence of Boltzmann distributions parameterized by decreasing temperatures (see, e.g. [15, 17]). As the temperature parameter decreases to zero, the Boltzmann distribution concentrates more around the global optimum. To be precise, let $\pi_{f,T}$ be the Boltzmann distribution corresponding to (P1),

$$\pi_{f,T}(S_i) = \frac{\int_{S_i} e^{f(x)/T}\,\mathrm{d}x}{\int_S e^{f(w)/T}\,\mathrm{d}w}$$

and let $\tilde{\pi}_{\tilde{f},\tilde{T}}$ be the Boltzmann distribution corresponding to (P2),

$$\tilde{\pi}_{\tilde{f},\tilde{T}}(\tilde{S}_i) = \frac{\sum_{\tilde{x}\in\tilde{S}_i} e^{\tilde{f}(\tilde{x})/\tilde{T}}}{\sum_{\tilde{w}\in\tilde{S}} e^{\tilde{f}(\tilde{w})/\tilde{T}}},$$

where $S_i \subseteq S$, $\tilde{S}_i \subseteq \tilde{S}$, and T and $\tilde{T}$ are referred to as the *temperature parameters*.

The adaptive search algorithm [19] is stated below.

Adaptive Search (AS)

Step 0 Set $k = 0$. Generate X_0 uniformly over the feasible region. Set $Y_0 = f(X_0)$ and $T_0 = \tau(Y_0)$, where τ is a nonnegative real valued nonincreasing function.
Step 1 Generate Z from the Boltzmann distribution with parameter T_k over the feasible region. If $f(Z) > Y_k$, set $X_{k+1} = Z$. Otherwise, repeat Step 1.
Step 2 Set $Y_{k+1} = f(X_{k+1})$ and set the temperature parameter $T_{k+1} = \tau(Y_{k+1})$. If stopping criteria has not been met, increment k and return to Step 1.

Intermediate points Z generated in step 1 are called trial points or sample points. The sequence $(X_k; k \geq 0)$ is a sequence of record points. Defined in this way, the last sampled point generated by step 1 in each iteration is then a record point (i.e., an improving point). Lastly, the function τ generating the sequence of temperature parameters $(T_k; k \geq 0)$ is called the *cooling schedule*.

The expected number of record points of adaptive search has been shown to increase linearly in the dimension of the problem for both continuous (P1) and discrete (P2) problems [19, 23]. However, the number of sample points needed to obtain a record value needs to be considered to reflect the overall performance of the algorithm. To maintain linear performance, we would like to manipulate the cooling schedule to maintain a constant $(1 - \alpha)$ probability of achieving an improving point. This leads to our choice of cooling schedule.

2.2 Characterization of the adaptive search cooling schedule

The following principle for a cooling schedule characterizes AS.

> Choose the temperature for the next iteration of adaptive search so that the probability of generating an improving point under the Boltzmann distribution is at least $1 - \alpha$.

Defining the cooling schedule in this way, the expected number of sample points in each iteration will be $1/(1 - \alpha)$ where $0 \leq \alpha < 1$, independent of the dimension n of the problem. As a result, not only the expected number of record points, but also the expected number of sample points of the adaptive search algorithm will grow linearly in the dimension of the problem. (The task of generating Boltzmann distributed sample points remains of course a challenging task that we will not address here.)

To construct the cooling schedule, we first define the improving region corresponding to the continuous problem (P1) as follows

$$S_{f(X_k)} = \{x \in S : f(x) > f(X_k)\},$$

where X_k is the record point sampled at the kth iteration of AS. And similarly, let $\tilde{X}_k$ be the record point on the kth iteration of AS for solving the discrete problem (P2), and define the corresponding improving region as

$$\tilde{S}_{\tilde{f}(\tilde{X}_k)} = \{\tilde{x} \in \tilde{S} : \tilde{f}(\tilde{x}) > \tilde{f}(\tilde{X}_k)\}.$$

For (P1) with $0 \leq \alpha < 1$, we want to derive the temperature T_k such that the probability of generating an improving point according to the Boltzmann distribution satisfies the following cooling schedule condition

$$\pi_{f,T_k}(S_{f(X_k)}) = \frac{\int_{S_{f(X_k)}} e^{f(x)/T_k} \mathrm{d}x}{\int_S e^{f(x)/T_k} \mathrm{d}x} \geq 1 - \alpha \tag{3}$$

and similarly for (P2), find the temperature $\tilde{T}_k$ such that

$$\tilde{\pi}_{\tilde{f},\tilde{T}_k}(\tilde{S}_{\tilde{f}(\tilde{X}_k)}) = \frac{\sum_{\tilde{x}\in\tilde{S}_{\tilde{f}(\tilde{X}_k)}} e^{\tilde{f}(\tilde{x})/\tilde{T}_k}}{\sum_{\tilde{x}\in\tilde{S}} e^{\tilde{f}(\tilde{x})/\tilde{T}_k}} \geq 1-\alpha. \tag{4}$$

Note that a temperature close to zero will satisfy the required probability, as in Eqs. 3 and 4. However, for a practical algorithm, in general the lower the temperature, the higher the difficulty to achieve the Boltzmann distribution. In order to be tractable computationally, we want to find the highest temperature possible that satisfies the required probability. We also want the cooling schedule to be applicable to a broad family of optimization problems since in general little is known about the problem before us. This leads to our worst case approach discussed in the next section.

3 An analytical cooling schedule

In this section, we develop cooling schedules for (P1) and (P2) that satisfy inequalities (3) and (4), respectively, based on a worst case analysis.

3.1 Worst case functions

Given the current record value, the temperature consistent with the principle of an adaptive search cooling schedule will be calculated based upon the worst case function among those consistent with the Lipschitz constants K and $\tilde{K}$, respectively.

The worst case functions for the continuous and discrete problems are constructed as follows.

Definition 1 Given the record point X_k generated at the kth iteration of AS for solving (P1), define a function $h(x)$ over $\Re^n$ as follows

$$h(x) = \max\left\{f(X_k), \left(f^* - K||x-x^*||\right)\right\}$$

for all $x \in \Re^n$.

Definition 2 Given the record point $\tilde{X}_k$ generated at the kth iteration of AS for solving (P2), define a function $\tilde{h}$ over Z^n as follows

$$\tilde{h}(\tilde{x}) = \max\left\{\tilde{f}(\tilde{X}_k), \left(\tilde{f}^* - \tilde{K}\max_{i=1,\ldots,n}(|\tilde{x}_i - \tilde{x}_i^*|)\right)\right\}$$

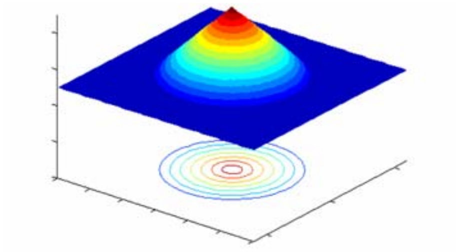
for all $\tilde{x} \in Z^n$.

An illustration of $h(x)$ and $\tilde{h}(\tilde{x})$ in two dimensions is given in Fig. 1.

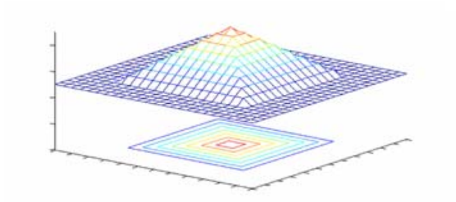
Let $S_{h(X_k)}$ and $\tilde{S}_{\tilde{h}(\tilde{X}_k)}$ denote the improving regions of the h function and $\tilde{h}$ function, respectively, i.e.,

$$S_{h(X_k)} = \{x \in S : h(x) > h(X_k)\},$$
$$\tilde{S}_{\tilde{h}(\tilde{X}_k)} = \{\tilde{x} \in \tilde{S} : \tilde{h}(\tilde{x}) > \tilde{h}(\tilde{X}_k)\}.$$

The following theorem states the functions h and $\tilde{h}$ are worse than the original functions f and $\tilde{f}$, respectively, in the sense that the probability of sampling an improving



(a) Worst case function for continuous problem in two dimensions.



(b) Worst case function for discrete problem in two dimensions

Fig. 1 Worst case functions in two dimensions

point with the worst case function is smaller than the probability of sampling an improving point with the original function.

Theorem 1 *Let X_k and $\tilde{X}_k$ be the record points at the kth iteration of AS for solving* (*P*1) *and* (*P*2)*, respectively. Under the same temperature and feasible region, the probability of sampling a point on the improving region according to the Boltzmann distribution corresponding to the worst function is less than that of sampling a point on the improving region according to the Boltzmann distribution with respect to the original functions $f(x)$ and $\tilde{f}(\tilde{x})$, i.e.,*

$$\pi_{f,T_k}(S_{f(X_k)}) \geq \pi_{h,T_k}(S_{h(X_k)}) \quad \text{and} \tag{5}$$

$$\tilde{\pi}_{\tilde{f},\tilde{T}_k}(\tilde{S}_{\tilde{f}(\tilde{X}_k)}) \geq \tilde{\pi}_{\tilde{h},\tilde{T}_k}(\tilde{S}_{\tilde{h}(\tilde{X}_k)}). \tag{6}$$

Proof We only show the proof for the continuous case. The proof for the discrete case is similar. The proof of the theorem relies on Lemma 4, which appears in Appendix A, and states that, for $a, b \in \Re$, $b > a$, $b > 0$, if $c \geq 0$ then $\frac{a+c}{b+c} \geq \frac{a}{b}$.

Let x be any point in the region $S_{h(X_k)}$, where $S_{h(X_k)} = \{x \in S : h(x) > h(X_k)\}$. Then $f(x) \geq h(x)$, where $h(x) = f^* - K||x - x^*||$. Because the original objective function $f(x)$ is a Lipschitz function with Lipschitz constant K, we have

$$f^* - f(x) \leq K||x - x^*||,$$

which is equivalent to

$$f(x) \geq f^* - K||x - x^*||.$$

By the definition of $h(x)$ and the Lipschitz condition, for any $x \in S_{h(X_k)}$, one has $f(x) \geq h(x)$, which implies that

$$S_{h(X_k)} \subseteq S_{f(X_k)}.$$

Next, by applying Lemma 4 and using the facts that $S_{h(X_k)} \subseteq S_{f(X_k)}$ and $f(x) \geq h(x)$ for any $x \in S_{h(X_k)}$, we show that $\pi_{f,T_k}(S_{f(X_k)}) \geq \pi_{h,T_k}(S_{h(X_k)})$.

According to the definition of the Boltzmann distribution, we have,

$$\pi_{f,T_k}(S_{f(X_k)}) = \frac{\int_{S_{f(X_k)}} e^{f(x)/T_k} \mathrm{d}x}{\int_S e^{f(x)/T_k} \mathrm{d}x} = \frac{\int_{S_{f(X_k)}} e^{f(x)/T_k} \mathrm{d}x}{\int_{S \setminus S_{f(X_k)}} e^{f(x)/T_k} \mathrm{d}x + \int_{S_{f(X_k)}} e^{f(x)/T_k} \mathrm{d}x}$$

and applying the fact that $S_{h(X_k)} \subseteq S_{f(X_k)}$, we have,

$$= \frac{\int_{S_{h(X_k)}} e^{f(x)/T_k} \mathrm{d}x + \int_{S_{f(X_k)} \setminus S_{h(X_k)}} e^{f(x)/T_k} \mathrm{d}x}{\int_{S \setminus S_{f(X_k)}} e^{f(x)/T_k} \mathrm{d}x + \int_{S_{h(X_k)}} e^{f(x)/T_k} \mathrm{d}x + \int_{S_{f(X_k)} \setminus S_{h(X_k)}} e^{f(x)/T_k} \mathrm{d}x}.$$

We apply Lemma 4, $(a+c)/(b+c) \geq a/b$ for $b > a > 0$ and $c > 0$, and use

$$a = \int\limits_{S_{h(X_k)}} e^{f(x)/T_k} \mathrm{d}x + \int\limits_{S_{f(X_k)} \setminus S_{h(X_k)}} e^{f(X_k)/T_k} \mathrm{d}x$$

$$b = \int\limits_{S \setminus S_{f(X_k)}} e^{f(x)/T_k} \mathrm{d}x + \int\limits_{S_{h(X_k)}} e^{f(x)/T_k} \mathrm{d}x + \int\limits_{S_{f(X_k)} \setminus S_{h(X_k)}} e^{f(X_k)/T_k} \mathrm{d}x$$

$$c = \int\limits_{S_{f(X_k)} \setminus S_{h(X_k)}} e^{f(x)/T_k} \mathrm{d}x - \int\limits_{S_{f(X_k)} \setminus S_{h(X_k)}} e^{f(X_k)/T_k} \mathrm{d}x.$$

Note that $c > 0$ because for all $x \in S_{f(X_k)}, f(x) \geq f(X_k)$. This yields,

$$\pi_{f,T_k}(S_{f(X_k)}) \geq \frac{\int_{S_{h(X_k)}} e^{f(x)/T_k} \mathrm{d}x + \int_{S_{f(X_k)} \setminus S_{h(X_k)}} e^{f(X_k)/T_k} \mathrm{d}x}{\int_{S \setminus S_{f(X_k)}} e^{f(x)/T_k} \mathrm{d}x + \int_{S_{h(X_k)}} e^{f(x)/T_k} \mathrm{d}x + \int_{S_{f(X_k)} \setminus S_{h(X_k)}} e^{f(X_k)/T_k} \mathrm{d}x}$$
$$\geq \frac{\int_{S_{h(X_k)}} e^{f(x)/T_k} \mathrm{d}x}{\int_{S \setminus S_{f(X_k)}} e^{f(x)/T_k} \mathrm{d}x + \int_{S_{h(X_k)}} e^{f(x)/T_k} \mathrm{d}x + \int_{S_{f(X_k)} \setminus S_{h(X_k)}} e^{f(X_k)/T_k} \mathrm{d}x}.$$

We again apply Lemma 4 by setting

$$a = \int\limits_{S_{h(X_k)}} e^{h(x)/T_k} \mathrm{d}x$$

$$b = \int\limits_{S \setminus S_{f(X_k)}} e^{f(x)/T_k} \mathrm{d}x + \int\limits_{S_{h(X_k)}} e^{h(x)/T_k} \mathrm{d}x + \int\limits_{S_{f(X_k)} \setminus S_{h(X_k)}} e^{f(X_k)/T_k} \mathrm{d}x$$

$$c = \int\limits_{S_{h(X_k)}} e^{f(x)/T_k} \mathrm{d}x - \int\limits_{S_{h(X_k)}} e^{h(x)/T_k} \mathrm{d}x,$$

where $c > 0$ since $f(x) \geq h(x)$ for all $x \in S_{h(X_k)}$. This yields,

$$\pi_{f,T_k}(S_{f(X_k)}) \geq \frac{\int_{S_{h(X_k)}} e^{h(x)/T_k} \mathrm{d}x}{\int_{S \setminus S_{f(X_k)}} e^{f(x)/T_k} \mathrm{d}x + \int_{S_{h(X_k)}} e^{h(x)/T_k} \mathrm{d}x + \int_{S_{f(X_k)} \setminus S_{h(X_k)}} e^{f(X_k)/T_k} \mathrm{d}x}$$

and because for all $x \in S \setminus S_{f(X_k)}, f(x) \leq f(X_k)$, we have,

$$\begin{aligned} &\geq \frac{\int_{S_{h(X_k)}} e^{h(x)/T_k} \mathrm{d}x}{\int_{S \setminus S_{f(X_k)}} e^{f(X_k)/T_k} \mathrm{d}x + \int_{S_{h(X_k)}} e^{h(x)/T_k} \mathrm{d}x + \int_{S_{f(X_k)} \setminus S_{h(X_k)}} e^{f(X_k)/T_k} \mathrm{d}x} \\ &= \frac{\int_{S_{h(X_k)}} e^{h(x)/T_k} \mathrm{d}x}{\int_{S \setminus S_{h(X_k)}} e^{f(X_k)/T_k} \mathrm{d}x + \int_{S_{h(X_k)}} e^{h(x)/T_k} \mathrm{d}x} \\ &= \pi_{h,T_k}(S_{h(X_k)}). \end{aligned}$$

Therefore, $\pi_{f,T_k}(S_{f(X_k)}) \geq \pi_{h,T_k}(S_{h(X_k)})$, i.e., $h(x)$ is a function "worse" than the original function $f(x)$. □

3.2 Calculation of the adaptive search cooling schedule

Given the kth record values for (P1) and (P2), we next develop a method to calculate T_k and $\tilde{T}_k$, respectively such that

$$\pi_{h,T_k}(S_{h(X_k)}) \geq 1 - \alpha \quad \text{and} \tag{7}$$

$$\tilde{\pi}_{\tilde{h},\tilde{T}_k}(\tilde{S}_{\tilde{h}(\tilde{X}_k)}) \geq 1 - \alpha \tag{8}$$

for $0 \leq \alpha < 1$. Note that if inequalities (7) and (8) are satisfied with respect to T_k and $\tilde{T}_k$, then according to Theorem 1, the desired results

$$\pi_{f,T_k}(S_{f(X_k)}) \geq 1 - \alpha \quad \text{and} \quad \tilde{\pi}_{\tilde{f},\tilde{T}_k}(\tilde{S}_{\tilde{f}(\tilde{X}_k)}) \geq 1 - \alpha$$

are also satisfied.

For both continuous and discrete cases, the probability of sampling the improving region of the worst case function not only depends on the current record value but also depends on the location of the global optimal point. For example, in the continuous case, if x^* is an interior point and $(f^* - f(X_k))/K$ is less than the shortest distance from x^* to the boundary of the feasible set S, then the improving region $S_{h(X_k)}$ is a full n-dimensional ball with center at x^*. But if x^* is located on the boundary of the set S, then no matter how small the value of $(f^* - f(X_k))/K$, $S_{h(X_k)}$ is always a part of the n-dimensional ball. In general, the location of the global point is unknown. The following theory allows a general location of the optimal point, however a tighter bound is possible if some information on the location of the optimal point is known.

Theorem 2 *Consider the program* $(P1)$ *with the convex feasible region* S *and the program* $(P2)$ *with the feasible region* $\tilde{S}$ *being the collection of integer points contained in an* n*-dimensional hyperrectangle. Suppose* X_k, $f(X_k)$ *and* T_k *are given for current iteration* k *for* $(P1)$ *and* $\tilde{X}_k$, $\tilde{f}(\tilde{X}_k)$ *and* $\tilde{T}_k$ *are given for current iteration* k *for* $(P2)$.

Define the continuous sets B_{θ_k}, B_ρ, $\tilde{B}_{\tilde{\theta}_k}$, and $\tilde{B}_{\tilde{\rho}}$ as follows

$$\begin{aligned}
B_{\theta_k} &= \{x \in \Re^n : \|x - x^*\| \le (f^* - f(X_k))/K = \theta_k\},\\
B_\rho &= \{x \in \Re^n : \|x - x^*\| \le \rho\},\\
\tilde{B}_{\tilde{\theta}_k} &= \{\tilde{x} \in \Re^n : \max_{i=1,\ldots,n} |\tilde{x}_i - \tilde{x}_i^*| \le \lfloor(\tilde{f}^* - \tilde{f}(\tilde{X}_k))/\tilde{K}\rfloor + 0.5 = \tilde{\theta}_k\},\\
\tilde{B}_{\tilde{\rho}} &= \{\tilde{x} \in \Re^n : \max_{i=1,\ldots,n} |\tilde{x}_i - \tilde{x}_i^*| \le \tilde{\rho} + 1\}.
\end{aligned}$$

Then

$$\pi_{h,T_k}(S_{h(X_k)}) \ge \frac{\int_{B_{\theta_k}} e^{h(x)/T_k}\,\mathrm{d}x}{\int_{B_\rho} e^{h(x)/T_k}\,\mathrm{d}x} \quad \text{and} \quad \tilde{\pi}_{\tilde{h},\tilde{T}_k}(\tilde{S}_{\tilde{h}(\tilde{X}_k)}) \ge \frac{\sum_{\tilde{x}\in(\tilde{B}_{\tilde{\theta}_k}\cap Z^n)} e^{\tilde{h}(\tilde{x})/\tilde{T}_k}}{\sum_{\tilde{x}\in(\tilde{B}_{\tilde{\rho}}\cap Z^n)} e^{\tilde{h}(\tilde{x})/\tilde{T}_k}}.$$

Proof See Appendix B. □

According to Theorem 2, the probability of sampling the improving region of the worst case function corresponding for the continuous program (P1) is bounded by the probability of sampling the n-dimensional ball B_{θ_k} with center x^* over the larger ball B_ρ. And for the discrete program (P2), the probability of sampling the improving region of $\tilde{h}$ is bounded by the probability of sampling the discrete points contained in the n-dimensional hypercube $\tilde{B}_{\tilde{\theta}_k}$ with center $\tilde{x}^*$ over the discrete domain contained in the hypercube $\tilde{B}_{\tilde{\rho}}$. The locations of the optimal points of both continuous and discrete programs do not influence the shapes of B_{θ_k} and $\tilde{B}_{\tilde{\theta}_k}$, which suggests a way of calculating a lower bound on the probability of sampling the improving region without the knowledge of the optimal point.

We next discuss a method to calculate T_k and $\tilde{T}_k$, respectively such that

$$\frac{\int_{B_{\theta_k}} e^{h(x)/T_k}\,\mathrm{d}x}{\int_{B_\rho} e^{h(x)/T_k}\,\mathrm{d}x} \ge 1 - \alpha \quad \text{and} \tag{9}$$

$$\frac{\sum_{\tilde{x}\in(\tilde{B}_{\tilde{\theta}_k}\cap Z^n)} e^{\tilde{h}(\tilde{x})/\tilde{T}_k}}{\sum_{\tilde{x}\in(\tilde{B}_{\tilde{\rho}}\cap Z^n)} e^{\tilde{h}(\tilde{x})/\tilde{T}_k}} \ge 1 - \alpha \tag{10}$$

for $0 \le \alpha < 1$. Note that if (9) and (10) are satisfied with respect to T_k and $\tilde{T}_k$, then (7) and (8) are also satisfied by Theorem 2.

We first consider solving for T_k for the continuous problem. The following theorem states that solving for the temperature T_k for the continuous problem to satisfy (9) is equivalent to solving (11) in Theorem 3 for z_k and using z_k to determine T_k.

Theorem 3 *Consider the continuous problem defined by* (P1) *with optimal objective function value f^*. Let $M_k = \left(\frac{K}{f^*-f(X_k)}\right)^n \cdot (\frac{1-\alpha}{\alpha}) \cdot \frac{v_n(B_\rho \backslash B_{\theta_k})}{2\pi^{n/2}/\Gamma(\frac{n}{2})}$. Then solving for z_k that satisfies*

$$p_{M_k}(z_k) = (n-1)! - \sum_{i=0}^{n-1} \frac{(n-1)!}{i!} z_k^i e^{-z_k} - M_k z_k^n e^{-z_k} \ge 0 \tag{11}$$

and setting $T_k = (f^* - f(X_k))/z_k$*, provides* T_k *that satisfies the inequality* $\frac{\int_{B_{\theta_k}} e^{h(x)/T_k}\,\mathrm{d}x}{\int_{B_\rho} e^{h(x)/T_k}\,\mathrm{d}x} \geq 1-\alpha$.

Proof

$$\frac{\int_{B_{\theta_k}} e^{h(x)/T_k}\,\mathrm{d}x}{\int_{B_\rho} e^{h(x)/T_k}\,\mathrm{d}x} = \frac{\int_{B_{\theta_k}} e^{h(x)/T_k}\,\mathrm{d}x}{\int_{B_\rho \setminus B_{\theta_k}} e^{h(x)/T_k}\,\mathrm{d}x + \int_{B_{\theta_k}} e^{h(x)/T_k}\,\mathrm{d}x}$$

and by definition, $h(x) = f^* - K\|x - x^*\|$ for all $x \in B_{\theta_k}$, and $h(x) = f(X_k)$ for all $x \in B_\rho \setminus B_{\theta_k}$, thus,

$$= \frac{\int_{B_{\theta_k}} e^{(f^* - K\|x-x^*\|)/T_k}\,\mathrm{d}x}{\int_{B_\rho \setminus B_{\theta_k}} e^{f(X_k)/T_k}\,\mathrm{d}x + \int_{B_{\theta_k}} e^{(f^* - K\|x-x^*\|)/T_k}\,\mathrm{d}x}$$

$$= \frac{e^{f^*/T_k} \cdot \int_{B_{\theta_k}} e^{-K\|x-x^*\|/T_k}\,\mathrm{d}x}{e^{f(X_k)/T_k} \cdot v_n(B_\rho \setminus B_{\theta_k}) + e^{f^*/T_k} \cdot \int_{B_{\theta_k}} e^{-K\|x-x^*\|/T_k}\,\mathrm{d}x}.$$

By substituting $u = x - x^*$ and changing the integration from rectangular coordinates to polar coordinates, with $G = \{u \in \Re^n : \|u\| \leq (f^* - f(X_k))/K = \theta\}$, we obtain

$$\frac{\int_{B_{\theta_k}} e^{h(x)/T_k}\,\mathrm{d}x}{\int_{B_\rho} e^{h(x)/T_k}\,\mathrm{d}x} = \frac{e^{f^*/T_k} \cdot \int_G e^{-K\|u\|/T_k}\,\mathrm{d}u}{e^{f(X_k)/T_k} \cdot v_n(B_\rho \setminus B_{\theta_k}) + e^{f^*/T_k} \cdot \int_G e^{-K\|u\|/T_k}\,\mathrm{d}u}$$

$$= \frac{(2\pi^{n/2}/\,\Gamma(n/2)) \cdot e^{f^*/T_k} \cdot \int_0^\theta r^{n-1} e^{-Kr/T_k}\,\mathrm{d}r}{v_n(B_\rho \setminus B_{\theta_k}) \cdot e^{f(X_k)/T_k} + \frac{2\pi^{n/2}}{\Gamma(n/2)} \cdot e^{f^*/T_k} \cdot \int_0^\theta r^{n-1} e^{-Kr/T_k}\,\mathrm{d}r}$$

Therefore, the inequality $\frac{\int_{B_{\theta_k}} e^{h(x)/T_k}\,\mathrm{d}x}{\int_{B_\rho} e^{h(x)/T_k}\,\mathrm{d}x} \geq 1 - \alpha$ is equivalent to

$$\frac{(2\pi^{n/2}/\,\Gamma(n/2)) \cdot e^{f^*/T_k} \cdot \int_0^\theta r^{n-1} e^{-Kr/T_k}\,\mathrm{d}r}{v_n(B_\rho \setminus B_{\theta_k}) \cdot e^{f(X_k)/T_k} + \frac{2\pi^{n/2}}{\Gamma(n/2)} \cdot e^{f^*/T_k} \cdot \int_0^\theta r^{n-1} e^{-Kr/T_k}\,\mathrm{d}r} \geq 1 - \alpha. \tag{12}$$

Carrying out the integration, inequality (12) is,

$$(n-1)! - \sum_{i=0}^{n-1} \frac{(n-1)!}{i!} \left(\frac{f^* - f(X_k)}{T_k}\right)^i e^{(f(X_k) - f^*)/T_k}$$
$$- \left(\frac{K}{T_k}\right)^n \left(\frac{1-\alpha}{\alpha}\right) \frac{v_n(B_\rho \setminus B_{\theta_k})}{2\pi^{n/2}/\,\Gamma(\frac{n}{2})} e^{(f(X_k) - f^*)/T_k} \geq 0.$$

Substituting $z_k = (f^* - f(X_k))/T_k$ and $M_k = \left(\frac{K}{f^* - f(X_k)}\right)^n \cdot \left(\frac{1-\alpha}{\alpha}\right) \cdot \frac{v_n(B_\rho \setminus B_{\theta_k})}{2\pi^{n/2}/\,\Gamma(\frac{n}{2})}$ into the above equation, we have

$$(n-1)! - \sum_{i=0}^{n-1} \frac{(n-1)!}{i!} z_k^i e^{-z_k} - M_k z_k^n e^{-z_k} \geq 0.$$

Therefore solving for T_k in inequality $\frac{\int_{B_{\theta_k}} e^{h(x)/T_k}\,\mathrm{d}x}{\int_{B_\rho} e^{h(x)/T_k}\,\mathrm{d}x} \geq 1 - \alpha$ is equivalent to solving for z_k in (11) and setting $T_k = (f^* - f(X_k))/z_k$. □

Theorem 3 transfers the problem of solving for T_k in inequality $\frac{\int_{B_{\theta_k}} e^{h(x)/T_k}\,\mathrm{d}x}{\int_{B_\rho} e^{h(x)/T_k}\,\mathrm{d}x} \geq 1-\alpha$ to the problem of solving for z_k in Eq. 11. Later in this section, we discuss characteristics of $p_{M_k}(z_k)$. Next, in Theorem 4, we derive an analogous expression to solve for the temperature $\tilde{T}_k$ in the discrete problem. For the discrete problem, we consider two cases; when there is a single discrete point contained in the region $\tilde{B}_{\tilde{\theta}_k}$, and when there are several discrete points contained in the region. In the first case, $\tilde{T}_k$ is calculated directly. In the latter case, $\tilde{T}_k$ is calculated using $p_{\tilde{M}_k}(\tilde{z}_k)$. The proof of Theorem 4 relies on two lemmas, Lemmas 4 and 5, which appear in the appendices.

Theorem 4 *Consider the discrete problem defined by* $(P2)$ *with optimal objective function value* $\tilde{f}^*$. *If* $\frac{\tilde{f}^*-\tilde{f}(\tilde{X}_k)}{\tilde{K}} < 1$, *then* $\tilde{T}_k \leq \frac{\tilde{f}^*-\tilde{f}(\tilde{X}_k)}{\ln[\frac{1-\alpha}{\alpha}(|\tilde{B}_{\tilde{\rho}}\cap Z^n|-1)]}$ *solves the inequality* $\frac{\sum_{\tilde{x}\in(\tilde{B}_{\tilde{\theta}_k}\cap Z^n)} e^{\tilde{h}(\tilde{x})/\tilde{T}_k}}{\sum_{\tilde{x}\in(\tilde{B}_{\tilde{\rho}}\cap Z^n)} e^{\tilde{h}(\tilde{x})/\tilde{T}_k}} \geq 1-\alpha$. *Moreover, if* $\frac{\tilde{f}^*-\tilde{f}(\tilde{X}_k)}{\tilde{K}} \geq 1$, *setting* $\tilde{M}_k = (\frac{\tilde{K}}{\tilde{f}^*-\tilde{f}(\tilde{X}_k)})^n \cdot (\frac{1-\alpha}{\alpha}) \cdot \frac{|\tilde{B}_{\tilde{\rho}}\cap Z^n|-(2\tilde{\delta}_k)^n}{2^n\cdot n}$, *with* $\hat{f}^* = \tilde{f}^* - 0.5\tilde{K}$ *and* $|\tilde{B}_{\tilde{\rho}} \cap Z^n|$ *representing the number of discrete points contained in the hypercube* $\tilde{B}_{\tilde{\rho}}$ *and* $\tilde{\delta}_k = \frac{\hat{f}^*-\tilde{f}(\tilde{X}_k)}{\tilde{K}}$, *we have that solving for* $\tilde{z}_k$ *satisfying*

$$p_{\tilde{M}_k}(\tilde{z}_k) = (n-1)! - \sum_{i=0}^{n-1} \frac{(n-1)!}{i!}\tilde{z}_k^i e^{-\tilde{z}_k} - \tilde{M}_k \tilde{z}_k^n e^{-\tilde{z}_k} \geq 0 \qquad (13)$$

and setting $\tilde{T}_k = (\hat{f}^* - \tilde{f}(\tilde{X}_k))/\tilde{z}_k$ *provides a* $\tilde{T}_k$ *that satisfies the inequality* $\frac{\sum_{\tilde{x}\in(\tilde{B}_{\tilde{\theta}_k}\cap Z^n)} e^{\tilde{h}(\tilde{x})/\tilde{T}_k}}{\sum_{\tilde{x}\in(\tilde{B}_{\tilde{\rho}}\cap Z^n)} e^{\tilde{h}(\tilde{x})/\tilde{T}_k}} \geq 1-\alpha$.

Proof If $\frac{\tilde{f}^*-\tilde{f}(\tilde{X}_k)}{\tilde{K}} < 1$, then there is only one lattice point, $\tilde{x}^*$, contained in the hypercube $\tilde{B}_{\tilde{\theta}_k}$. Hence

$$\frac{\sum_{\tilde{x}\in(\tilde{B}_{\tilde{\theta}_k}\cap Z^n)} e^{\tilde{h}(\tilde{x})/\tilde{T}_k}}{\sum_{\tilde{x}\in(\tilde{B}_{\tilde{\rho}}\cap Z^n)} e^{\tilde{h}(\tilde{x})/\tilde{T}_k}} = \frac{e^{\tilde{f}^*/\tilde{T}_k}}{e^{\tilde{f}^*/\tilde{T}_k} + (|\tilde{B}_{\tilde{\rho}} \cap Z^n| - 1)e^{\tilde{f}(\tilde{X}_k)/\tilde{T}_k}} \geq 1-\alpha,$$

which is equivalent to

$$\tilde{T}_k \leq \frac{\tilde{f}^* - \tilde{f}(\tilde{X}_k)}{\ln[\frac{1-\alpha}{\alpha}(|\tilde{B}_{\tilde{\rho}} \cap Z^n| - 1)]}.$$

Therefore, if $\frac{\tilde{f}^*-\tilde{f}(\tilde{X}_k)}{\tilde{K}} < 1$, $\tilde{T}_k \leq \frac{\tilde{f}^*-\tilde{f}(\tilde{X}_k)}{\ln[\frac{1-\alpha}{\alpha}(|\tilde{B}_{\tilde{\rho}}\cap Z^n|-1)]}$ solves $\frac{\sum_{\tilde{x}\in(\tilde{B}_{\tilde{\theta}_k}\cap Z^n)} e^{\tilde{h}(\tilde{x})/\tilde{T}_k}}{\sum_{\tilde{x}\in(\tilde{B}_{\tilde{\rho}}\cap Z^n)} e^{\tilde{h}(\tilde{x})/\tilde{T}_k}} \geq 1-\alpha$.

Now consider the case $\frac{\tilde{f}^*-\tilde{f}(\tilde{X}_k)}{\tilde{K}} \geq 1$, which implies

$$\frac{\sum_{\tilde{x}\in(\tilde{B}_{\tilde{\theta}_k}\cap Z^n)} e^{\tilde{h}(\tilde{x})/\tilde{T}_k}}{\sum_{\tilde{x}\in(\tilde{B}_{\tilde{\rho}}\cap Z^n)} e^{\tilde{h}(\tilde{x})/\tilde{T}_k}} = \frac{\sum_{\tilde{x}\in(\tilde{B}_{\tilde{\theta}_k}\cap Z^n)} e^{\tilde{h}(\tilde{x})/T_k}}{\sum_{\tilde{x}\in(\tilde{B}_{\tilde{\theta}_k}\cap Z^n)} e^{\tilde{h}(\tilde{x})/\tilde{T}_k} + \sum_{\tilde{x}\in(\tilde{B}_{\tilde{\rho}}\cap Z^n)\setminus(\tilde{B}_{\tilde{\theta}_k}\cap Z^n)} e^{\tilde{f}(\tilde{X}_k)/\tilde{T}_k}}.$$

We let $m_k = \lfloor \frac{\tilde{f}^* - \tilde{f}(\tilde{X}_k)}{\tilde{K}} \rfloor$, and since $(\tilde{B}_{\tilde{\theta}_k} \cap Z^n) = \{\tilde{x} \in Z^n : \max_{i=1,\ldots,n}(|\tilde{x}_i - \tilde{x}_i^*|) \leq m_k\}$, this implies that the number of lattice points contained in the set $\tilde{B}_{\tilde{\theta}_k}$ is $(2m_k + 1)^n$. Consequently, we have

$$\frac{\sum_{\tilde{x}\in(\tilde{B}_{\tilde{\theta}_k}\cap Z^n)} e^{\tilde{h}(\tilde{x})/\tilde{T}_k}}{\sum_{\tilde{x}\in(\tilde{B}_{\tilde{\rho}}\cap Z^n)} e^{\tilde{h}(\tilde{x})/\tilde{T}_k}} = \frac{\sum_{\tilde{x}\in(\tilde{B}_{\tilde{\theta}_k}\cap Z^n)} e^{\tilde{h}(\tilde{x})/\tilde{T}_k}}{\sum_{\tilde{x}\in(\tilde{B}_{\tilde{\theta}_k}\cap Z^n)} e^{\tilde{h}(\tilde{x})/\tilde{T}_k} + \left(|\tilde{B}_{\tilde{\rho}} \cap Z^n| - (2m_k+1)^n\right) e^{\tilde{f}(\tilde{X}_k)/\tilde{T}_k}}$$

and applying Lemmas 4 and 5 yields

$$\geq \frac{\int_0^{\tilde{\delta}_k} e^{(\hat{f}^* - u\tilde{K})/\tilde{T}_k} \cdot 2n \cdot (2u)^{n-1} \mathrm{d}u}{\int_0^{\tilde{\delta}_k} e^{(\hat{f}^* - u\tilde{K})/\tilde{T}_k} \cdot 2n \cdot (2u)^{n-1} \mathrm{d}u + \left(|\tilde{B}_{\tilde{\rho}} \cap Z^n| - (2m_k+1)^n\right) e^{\tilde{f}(\tilde{X}_k)/\tilde{T}_k}}$$

and using the fact that $\tilde{\delta}_k \leq m_k + 0.5$, one has

$$\geq \frac{\int_0^{\tilde{\delta}_k} e^{(\hat{f}^* - u\tilde{K})/\tilde{T}_k} \cdot 2n \cdot (2u)^{n-1} \mathrm{d}u}{\int_0^{\tilde{\delta}_k} e^{(\hat{f}^* - u\tilde{K})/\tilde{T}_k} \cdot 2n \cdot (2u)^{n-1} du + \left(|\tilde{B}_{\tilde{\rho}} \cap Z^n| - (2\tilde{\delta}_k)^n\right) e^{\tilde{f}(\tilde{X}_k)/\tilde{T}_k}}.$$

Therefore, if $\frac{\tilde{f}^* - \tilde{f}(\tilde{X}_k)}{\tilde{K}} \geq 1$, $\tilde{T}_k$ can be determined by solving the inequality

$$\frac{\sum_{\tilde{x}\in(\tilde{B}_{\tilde{\theta}_k}\cap Z^n)} e^{\tilde{h}(\tilde{x})/\tilde{T}_k}}{\sum_{\tilde{x}\in(\tilde{B}_{\tilde{\rho}}\cap Z^n)} e^{\tilde{h}(\tilde{x})/\tilde{T}_k}}$$

$$\geq \frac{\int_0^{\tilde{\delta}_k} e^{(\hat{f}^* - u\tilde{K})/\tilde{T}_k} \cdot 2n \cdot (2u)^{n-1} \mathrm{d}u}{\int_0^{\tilde{\delta}_k} e^{(\hat{f}^* - u\tilde{K})/\tilde{T}_k} \cdot 2n \cdot (2u)^{n-1} \mathrm{d}u + \left(|\tilde{B}_{\tilde{\rho}} \cap Z^n| - (2\tilde{\delta}_k)^n\right) e^{\tilde{f}(\tilde{X}_k)/\tilde{T}_k}} \geq 1 - \alpha.$$

Following the similar method used for proving Theorem 3 and the fact that

$$\int_0^{\tilde{\delta}_k} e^{(\hat{f}^* - u\tilde{K})/\tilde{T}_k} \cdot 2n \cdot (2u)^{n-1} \mathrm{d}u$$

$$= 2^n n \cdot \left(\frac{\tilde{T}_k}{\tilde{K}}\right)^n \cdot \left[(n-1)! - e^{-\tilde{K}\tilde{\delta}_k/\tilde{T}_k} \sum_{i=0}^{n-1} \frac{(n-1)!}{i!} \left(\tilde{K}\tilde{\delta}_k/\tilde{T}_k\right)\right] \cdot e^{\hat{f}^*/\tilde{T}_k},$$

it is easy to prove that solving for $\tilde{T}_k$ satisfying the second inequality above is equivalent to solving for $\tilde{z}_k$ in the inequality (13) and setting $\tilde{T}_k = (\hat{f}^* - \tilde{f}(\tilde{X}_k))/\tilde{z}_k$. □

Theorems 3 and 4 provide a way to solve for T_k and $\tilde{T}_k$ for the continuous problem and the discrete problem, respectively. Both theorems are related to the function

$$p_{\mathcal{M}}(z) = (n-1)! - \sum_{i=0}^{n-1} \frac{(n-1)!}{i!} z^i e^{-z} - \mathcal{M} z^n e^{-z}$$

with $\mathcal{M} = M_k$ for the continuous problem and $\mathcal{M} = \tilde{M}_k$ for the discrete problem. Next we discuss the characteristics of the $p_{\mathcal{M}}(z)$ function.

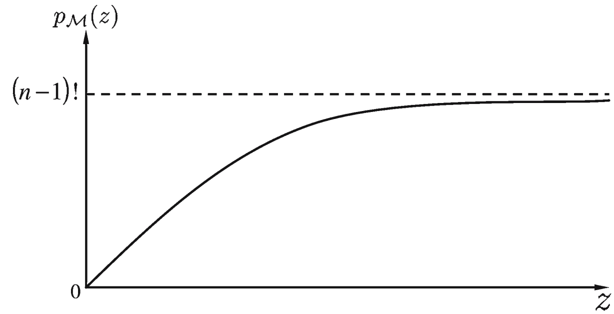
Lemma 1 *Consider the function* $p_{\mathcal{M}}(z) = (n-1)! - \sum_{i=0}^{n-1} \frac{(n-1)!}{i!} z^i e^{-z} - \mathcal{M} z^n e^{-z}$ *with* $z \in \Re^+$. *Then*

(1) *if* $n\mathcal{M} \le 1$, $p_{\mathcal{M}}(z) > 0$ *for all* $z > 0$;
(2) *if* $n\mathcal{M} > 1$, *the equation* $p_{\mathcal{M}}(z) = 0$ *has an unique solution* z^* *such that* $z^* > \frac{n\mathcal{M}-1}{\mathcal{M}}$, *and for all* $z \ge z^*$ *one has* $p_{\mathcal{M}}(z) \ge 0$

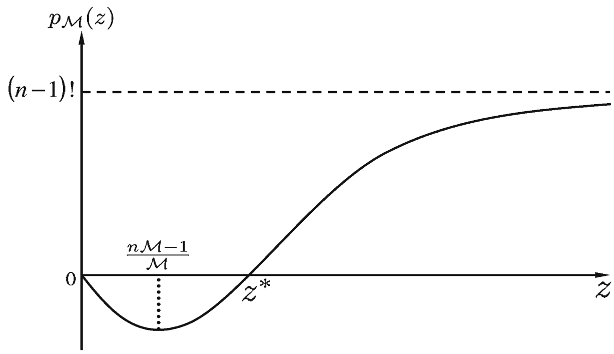
Proof The lemma is proved by analyzing the shape of the function $p_{\mathcal{M}}(z)$ corresponding to the parameters n and $\mathcal{M}$. The detailed proof is given in Appendix D. □

According to Lemma 1, finding z such that $p_{\mathcal{M}}(z) \ge 0$ can be separated into two cases based on the parameters n and $\mathcal{M}$. These two cases are shown in Fig. 2. If $n\mathcal{M} \le 1$, any positive value of z satisfies $p_{\mathcal{M}}(z) \ge 0$; if $n\mathcal{M} > 1$, any $z \ge z^*$ will satisfy $p_{\mathcal{M}}(z) \ge 0$, where z^* solves equation $p_{\mathcal{M}}(z) = 0$. Lemma 1 guarantees the uniqueness and existence of z^*, which implies that z^* can be calculated numerically by any simple line search algorithm, e.g. bisection. An important issue for those line search algorithms are the choice of initial points, i.e. the upper and lower bounds of z^*. Lemma 1 provides a lower bound, $\frac{n\mathcal{M}-1}{\mathcal{M}} < z^*$. Next, in Lemmas 2 and 3, we present upper bounds on z^* corresponding to the continuous and discrete problems, respectively.

Fig. 2 Two cases of the graph of the function $p_{\mathcal{M}}(z)$



(a) the graph of the function $p_{\mathcal{M}}(z)$ for $n\mathcal{M} \le 1$.



(b) the graph of the function $p_{\mathcal{M}}(z)$ for $n\mathcal{M} > 1$

Lemma 2 *Consider the continuous problem* $(P1)$ *and the function*

$$p_{M_k}(z_k) = (n-1)! - \sum_{i=0}^{n-1} \frac{(n-1)!}{i!} z_k^i e^{-z_k} - M_k z_k^n e^{-z_k}$$

where z_k *and* M_k *are defined as in Theorem 3. If* $nM_k > 1$, *let* z_k^* *be the solution to* $p_{M_k}(z_k) = 0$, *then*

$$z_k^* \leq \frac{L_k}{(1-\varepsilon)}, \tag{14}$$

where $0 < \varepsilon < 1$, $L_k = \ln\left(\frac{1-\alpha}{\alpha} \cdot \frac{v_n(B_\rho)-v_n(B_{\hat{\varepsilon}_k})}{v_n(B_{\hat{\varepsilon}_k})}\right)$, $0 \leq \alpha < 1$, $B_{\hat{\varepsilon}_k}$ *is an n-dimensional ball with radius* $\hat{\varepsilon}_k = \varepsilon(f^* - f(X_k))/K$ *and* B_ρ *is an n-dimensional ball with radius* ρ *as defined in Theorem 2.*

Proof See Appendix E. □

Now we turn to an upper bound on $\tilde{z}_k^*$ for the discrete problem.

Lemma 3 *Consider the discrete problem* $(P2)$ *and the function*

$$p_{\tilde{M}_k}(\tilde{z}_k) = (n-1)! - \sum_{i=0}^{n-1} \frac{(n-1)!}{i!} \tilde{z}_k^i e^{-\tilde{z}_k} - \tilde{M}_k \tilde{z}_k^n e^{-\tilde{z}_k},$$

where $\tilde{z}_k$ *and* $\tilde{M}_k$ *are defined as in Theorem 4. If* $\frac{\tilde{f}^*-\tilde{f}(\tilde{X}_k)}{\tilde{K}} \geq 1$ *and* $n\tilde{M}_k > 1$, *letting* $\tilde{z}_k^*$ *be the solution to* $p_{\tilde{M}_k}(\tilde{z}_k) = 0$, *we have that*

$$\tilde{z}_k^* \leq \frac{\tilde{L}_k}{(1-\varepsilon)}, \tag{15}$$

where $0 < \varepsilon < 1$, $\tilde{L}_k = \ln\left(\frac{1-\alpha}{\alpha} \cdot \frac{|\tilde{B}_{\tilde{\rho}} \cap Z^n| - (2\varepsilon\tilde{\delta}_k)^n}{(2\varepsilon\tilde{\delta}_k)^n}\right)$, $0 \leq \alpha < 1$, $\tilde{\delta}_k$ *and* $\hat{f}^*$ *are defined as in Theorem 4,* $\tilde{B}_{\tilde{\rho}}$ *is defined as in Theorem 2, and* $|\tilde{B}_{\tilde{\rho}} \cap Z^n|$ *denotes the number of integer points contained in* $\tilde{B}_{\tilde{\rho}}$.

Proof See Appendix E. □

3.3 Cooling schedule strategy

We are now ready to summarize the cooling schedule strategy for the adaptive search algorithm. This cooling schedule strategy ensures that the probability of sampling an improving point is not less than $1-\alpha$. Consequently the expected number of record points and sample points is linear in the dimension of the problem. This linearity result assumes that points are sampled according to a Boltzmann distribution. Since generating points according to a Boltzmann distribution is in general more difficult when the temperature is small, we endeavor to keep T_k as high as possible in our cooling schedule strategy.

3.3.1 Cooling Schedule Strategy for (P1)

- Keep $T_k = \infty$ until $nM_k > 1$, where M_k is defined in Theorem 3.
- When $nM_k > 1$, either
  - (i) use a numerical method to calculate z_k^* by solving $p_{M_k}(z_k) = 0$ as in Theorem 3 Eq. 11 and set $T_k = (f^* - f(X_k))/z_k^*$, or
  - (ii) use the upper bound on z_k^* in Lemma 2 Eq. 14 to calculate $T_k = \frac{(1-\epsilon)(f^*-f(X_k))}{L_k}$, where L_k is defined in Lemma 2.

3.3.2 Cooling Schedule Strategy for (P2)

- Keep $\tilde{T}_k = \infty$ until $n\tilde{M}_k > 1$, where $\tilde{M}_k$ is defined in Theorem 4.
- When $n\tilde{M}_k > 1$ and $\tilde{f}^* - \tilde{f}(\tilde{X}_k)/\tilde{K} < 1$, set $\tilde{T}_k = \frac{\tilde{f}^*-\tilde{f}(\tilde{X}_k)}{\ln((|(\tilde{B}_{\tilde{\rho}} \cap Z^n)|-1)(1-\alpha)/\alpha)}$.
- When $n\tilde{M}_k > 1$ and $\tilde{f}^* - \tilde{f}(\tilde{X}_k)/\tilde{K} \geq 1$, either
  - (i) use a numerical method to calculate $\tilde{z}_k^*$ by solving $p_{\tilde{M}_k}(\tilde{z}_k) = 0$ as in Theorem 4 Eq. 13 and set $\tilde{T}_k = (\hat{f}^* - \tilde{f}(\tilde{X}_k))/\tilde{z}_k^*$, where $\hat{f}^* = \tilde{f}^* - 0.5\tilde{K}$ as defined in Theorem 4, or
  - (ii) use the upper bound on $\tilde{z}_k^*$ in Lemma 3 Eq. 15 to calculate $\tilde{T}_k = \frac{(1-\epsilon)(\hat{f}^*-\tilde{f}(\tilde{X}_k))}{\tilde{L}_k}$, where $\tilde{L}_k$ is defined in Lemma 3.

To implement the cooling schedule strategies, we need to know several pieces of information: dimension n, the current record value $f(X_k)$ or $\tilde{f}(\tilde{X}_k)$, the Lipschitz constant K for (P1) or $\tilde{K}$ for (P2), the optimal value f^*, and the n-dimensional volume of the feasible set. For a specific problem, we certainly know the dimension. At the kth iteration, we also know the record value. The constant K or $\tilde{K}$ is assumed to be known, but an estimate that is an upper bound may also be used [16]. In general, the maximum function value f^* is unknown. However any lower bound to f^* that exceeds $f(X_k)$ may be used at the sacrifice of a cooler temperature than needed. In the following computational study, we estimate f^* using order statistics as in [18]. Similarly, any upper bound on the volume of the feasible region may be used while maintaining the linearity properties of the resulting cooling schedule.

4 Numerical results

We have performed a computational study comparing an adaptive cooling schedule with several other cooling schedules on continuous and discrete test problems from the literature. The other cooling schedules include: a cooling schedule proposed by Bohachevsky et al. [5], an exponential cooling schedule [11], and a logarithmic cooling schedule [10]. Several parameters and implementations are considered and are summarized below.
Adaptive cooling schedule (*summarized in Sect. 3.3*)

$$f^* \text{ known}, \alpha = 0.01, 0.05, 0.9, \qquad T_k = \frac{f^*-f(X_k)}{z_k},$$

$$\hat{f} \text{ estimate}, \alpha = 0.01, 0.05, 0.9, \qquad T_k = \frac{\hat{f}-f(X_k)}{z_k}.$$

Fixed beta cooling schedule [5]

$$f^* \text{ known}, \beta = 0.01, 1, 100, \qquad T_k = \beta(f^* - f(X_k)),$$
$$\hat{f} \text{ estimate}, \beta = 0.01, 1, 100, \qquad T_k = \beta(\hat{f} - f(X_k)).$$

Exponential cooling schedule [11]

$$T_k = T_0\gamma^k,$$
$$T_0 = 0.01, 1, 100 \text{ and } \gamma = 0.01, 0.99.$$

Logarithmic cooling schedule [10]

$$T_k = \frac{T_0}{\ln(k+1)},$$
$$T_0 = 0.01, 1, 100.$$

The adaptive cooling schedule developed in this paper uses the information of f^*. In the computational study, we tested two cases, one is to give a known global optimum, and the other is to estimate and update $\hat{f}_k$ during the process of running the algorithm. The estimate $\hat{f}_k$ for f^* is based on order statistics and was used in [18],

$$\hat{f}_k = f(X_k) + \frac{f(X_k) - f(X_{k-1})}{(1-q)^{-n/2} - 1}, \tag{16}$$

where $f(X_k)$ and $f(X_{k-1})$ are the first and second best points found so far, n is the dimension of the test problem and we chose $q = 0.1$.

The adaptive cooling schedule is summarized in Sect. 3.3, where z_k is calculated every time an improving point is found. The problem specific parameters (including an upper bound on the Lipschitz constant) used in this calculation are summarized with the test problems given in Appendix F. The other parameter specific to this cooling schedule is α, where $1 - \alpha$ is the desired probability of sampling an improving point. Three values of α, 0.01, 0.05, and 0.9, were used in the computational study to represent a spread of values.

The cooling schedule suggested by Bohachevsky et al. [5] has a form $T_k = \beta(f^* - f(X_k))$ where the parameter value of β is chosen at the onset and held constant throughout the algorithm. The temperature is updated every time an improving point is found. This fixed beta cooling schedule was also run when the global optimum f^* is known and when it is estimated with $\hat{f}_k$ as in Eq. 16. Typically the parameter value β is determined by trial and error, and we used three values of β, 0.01, 1, and 100, where the smallest value was determined by trial and error to provide good performance. The two other values were chosen to represent a spread.

The exponential cooling schedule [11] of the form $T_k = T_0\gamma^k$ was used and the temperature was updated every iteration (in contrast to the previous cooling schedules that updated temperature when an improving point was found). Values of T_0, 0.01, 1, and 100, and γ, 0.01, 0.99, were chosen by trial and error, and to represent a spread of values.

The logarithmic cooling schedule [10] of the form $T_k = \frac{T_0}{\ln(k+1)}$ was also implemented where temperature was updated every iteration, and three values of T_0, 0.01, 1, and 100, were used.

The candidate point generator used in the simulated annealing algorithm is from the family of Hit-and-Run methods with continuous [22] and discrete [21] versions.

Four test problems, described in Appendix F, were used, two in a continuous form and two in a discrete form. The sinusoidal function was introduced in [24], used in [2], and was motivated by an engineering design problem of composite laminate structures.

For each cooling schedule, 100 runs with random starting points were performed on each test problem. The computational results are shown in the following tables and graphs. Tables 1–4 show the performance of the 21 different cooling schedules with parameters for solving test problems 1–4, where in each cell the upper integer indicates the number of successful runs out of 100 that achieved a value no worse than the associated y-value, and the lower number indicates the average number of function evaluations of the successful runs to achieve the y-value. The bold line indicates the parameter value with the best performance. Figures 3–6 graphically illustrate

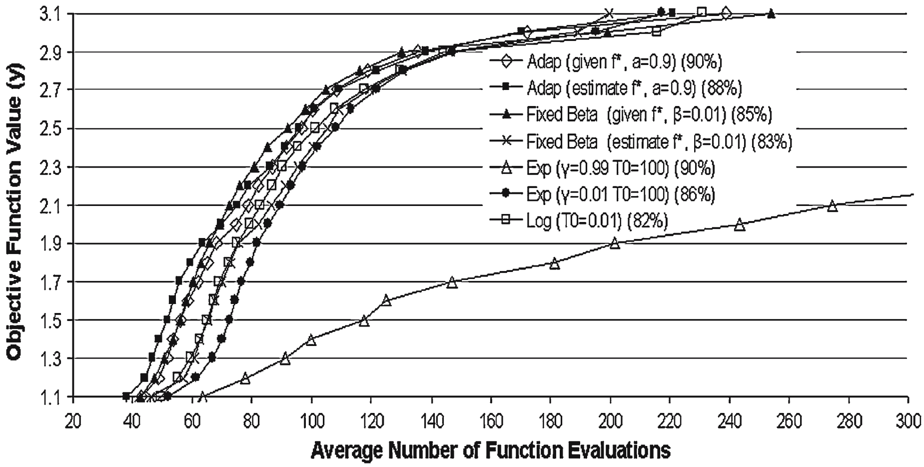


Fig. 3 Average number of function evaluations versus function value for successful runs for the continuous six-dimensional Hartmann6 test problem

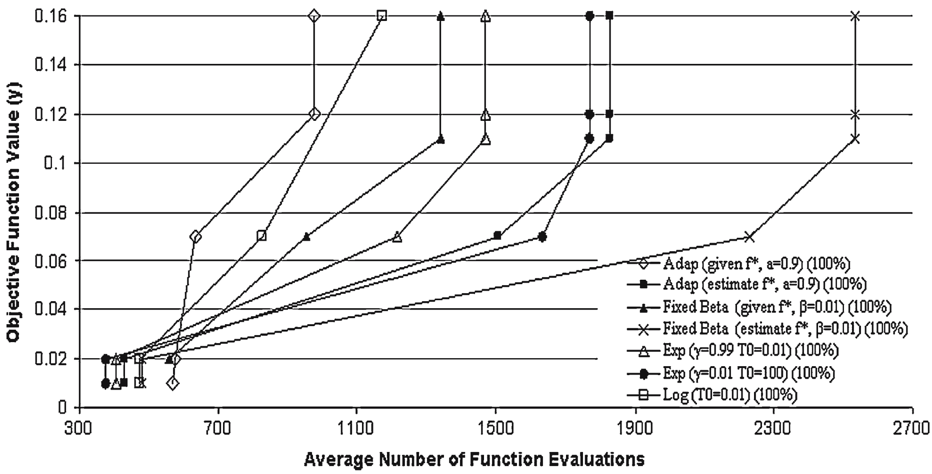


Fig. 4 Average number of function evaluations versus function value for successful runs for the discrete six-dimensional Hartmann6 test problem

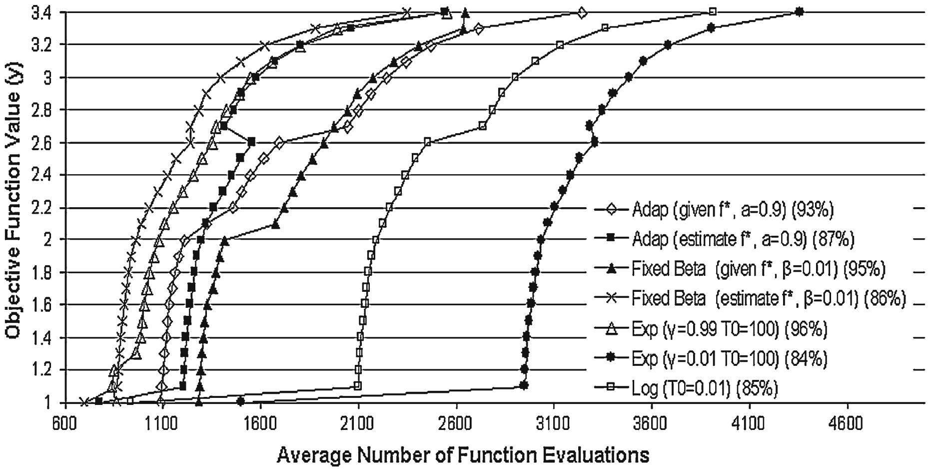


Fig. 5 Average number of function evaluations versus function value for successful runs for the continuous ten-dimensional sinusoidal test problem

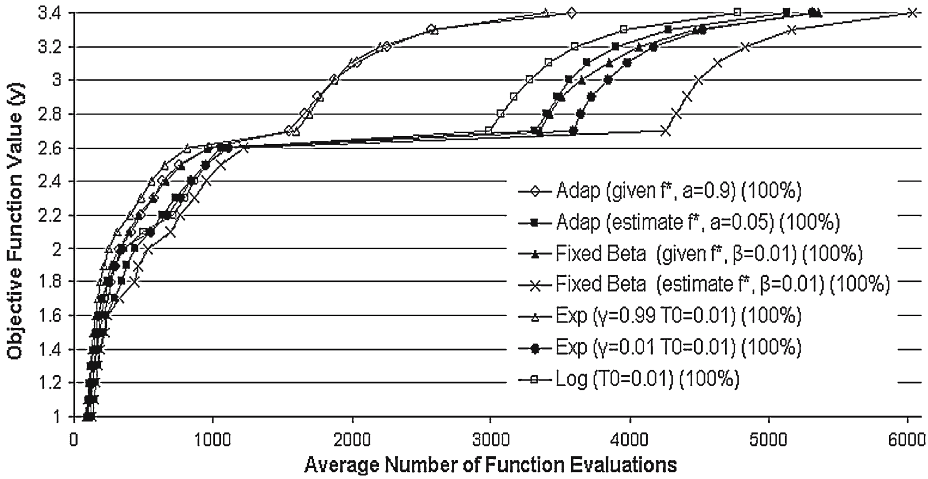


Fig. 6 Average number of function evaluations versus function value for successful runs for the discrete ten-dimensional sinusoidal test problem

related results from Tables 1–4. The four graphs show the performance of each cooling schedule with its best parameter value for solving each test problem.

While all of the cooling schedules converge to zero temperature, the graphs highlight the different progress of the algorithms. For example, in Fig. 6, several cooling schedules have long plateaus which are indicative of the algorithm getting stuck and not quickly finding an improving point. This may be caused by the temperature going to zero too quickly, and then relying on the candidate point generator (Hit-and-Run) to eventually find an improving point that is close to global optimum.

There is still a gap between the theoretical analysis and the algorithm in practice, because the Hit-and-Run generator does not sample according to a true Boltzmann distribution. However the theory does motivate the adaptive cooling schedule. The

Table 1 Number of successful runs out of 100 and average number of function evaluations for successful runs for the continuous six-dimensional Hartmann test problem

y	2.0	2.2	2.4	2.6	2.8	3.0	3.2
Adap (given f*)	100	100	100	100	100	98	65
$\alpha = 0.01$	77.5	84.2	93.8	103.5	121.9	186.9	236.0
Adap (given f*)	100	100	100	100	100	98	65
$\alpha = 0.05$	75.5	82.2	92.0	101.3	119.4	184.8	232.8
Adap (given f*)	100	100	100	100	100	96	65
$\alpha = 0.9$	74.7	82.0	91.7	100.5	118.6	172.1	230.3
Adap (estimate f*)	99	99	99	99	99	95	67
$\alpha = 0.01$	77.2	86.6	95.5	104.5	123.9	164.8	238.1
Adap (estimate f*)	99	99	99	99	99	95	67
$\alpha = 0.05$	77.3	86.5	95.9	104.8	125.0	164.6	240.9
Adap (estimate f*)	99	99	99	99	99	95	68
$\alpha = 0.9$	69.5	78.7	90.9	100.9	121.9	170.6	240.8
Fixed beta (given f*)	100	100	100	100	100	98	60
$\beta = 0.01$	69.0	75.8	85.1	97.8	115.9	199.0	259.7
Fixed beta (given f*)	100	100	98	98	95	85	31
$\beta = 1$	179.4	241.9	329.0	399.0	488.7	635.2	884.2
Fixed beta (given f*)	99	95	89	71	38	14	0
$\beta = 100$	285.0	388.6	508.3	590.5	667.2	593.8	–
Fixed beta (estimate f*)	99	99	99	99	99	96	66
$\beta = 0.01$	82.0	91.2	100.3	109.2	130.3	189.3	253.3
Fixed beta (estimate f*)	100	100	100	98	96	86	27
$\beta = 1$	118.2	149.3	188.5	272.6	375.3	520.1	536.4
Fixed beta (estimate f*)	100	96	92	72	40	16	0
$\beta = 100$	287.2	371.4	479.5	573.3	647.9	754.2	–
Exp ($\gamma = 0.99$)	100	100	100	100	100	96	62
$T_0 = 0.01$	68.2	75.0	83.8	96.8	113.5	189.0	229.2
Exp ($\gamma = 0.99$)	100	100	100	100	100	96	59
$T_0 = 1$	101.7	113.4	130.8	145.2	176.7	254.0	406.0
Exp ($\gamma = 0.99$)	100	100	100	100	100	96	68
$T_0 = 100$	243.5	324.0	378.6	468.0	556.4	675.2	813.6
Exp ($\gamma = 0.01$)	100	100	100	100	100	95	61
$T_0 = 0.01$	86.3	93.4	101.9	113.5	134.4	207.9	274.8
Exp ($\gamma = 0.01$)	100	100	100	100	100	95	62
$T_0 = 1$	88.6	95.7	103.8	115.6	135.0	204.2	269.3
Exp ($\gamma = 0.01$)	100	100	100	100	100	96	63
$T_0 = 100$	85.4	93.2	101.8	113.3	130.8	195.3	249.0
Log	100	100	100	100	100	98	60
$T_0 = 0.01$	79.4	86.64	95.31	107.94	129.46	215.85	251.03
Log	100	100	100	100	100	100	60
$T_0 = 1$	72.1	82.8	90.0	99.0	128.8	212.8	597.0
Log	99	96	95	73	45	8	0
$T_0 = 100$	272.8	360.7	501.5	603.0	735.6	976.9	–

numerical results indicate that in most of the cases, the adaptive cooling schedule with its best choice of parameter was first or second best compared to the others, and its performance is not very sensitive to the choice of α. In the theory, the choice of α impacts the temperature and consequently the number of function evaluations, but our adaptive cooling schedule is derived from a conservative worst case function, h or $\tilde{h}$, and bound on the Lipschitz constant, K or $\tilde{K}$. In practice, we speculate that the impact of α on performance is overshadowed by our conservative worst case function and loose Lipschitz bounds.

Table 2 Number of successful runs out of 100 and average number of function evaluations for successful runs for the discrete six-dimensional Hartmann test problem

y	0.02	0.04	0.06	0.08	0.10	0.12	0.14	0.16
Adap (given f*)	100	100	100	100	100	100	100	100
$\alpha = 0.01$	563.2	613.1	613.1	1080.7	1080.7	1080.7	1081.0	1081.0
Adap (given f*)	100	100	100	100	100	100	100	100
$\alpha = 0.05$	579.9	645.9	645.9	1095.1	1095.1	1095.1	1095.5	1095.5
Adap (given f*)	100	100	100	100	100	100	100	100
$\alpha = 0.9$	576.7	634.3	634.3	977.2	977.2	977.2	977.5	977.5
Adap (estimate f*)	100	100	100	100	100	100	100	100
$\alpha = 0.01$	422.7	1861.9	1861.9	2194.7	2194.7	2196.1	2196.4	2196.4
Adap (estimate f*)	100	100	100	100	100	100	100	100
$\alpha = 0.05$	430.6	1995.1	1995.1	2347.9	2347.9	2349.3	2349.6	2349.6
Adap (estimate f*)	100	100	100	100	100	100	100	100
$\alpha = 0.9$	427.6	1505.6	1505.6	1824.9	1824.9	1826.6	1827.5	1827.5
Fixed beta (given f*)	100	100	100	100	100	100	100	100
$\beta = 0.01$	557.3	953.7	953.7	1340.9	1340.9	1341.3	1341.4	1341.4
Fixed beta (given f*)	100	100	100	90	90	90	90	90
$\beta = 1$	511.7	553.9	553.9	8710.4	8710.4	8710.4	8890.9	8890.9
Fixed beta (given f*)	100	100	100	88	88	88	86	86
$\beta = 100$	532.0	586.2	586.2	10131.4	10131.4	10131.4	10512.9	10512.9
Fixed beta (estimate f*)	100	100	100	100	100	100	100	100
$\beta = 0.01$	479.1	2229.3	2229.3	2532.8	2532.8	2533.5	2533.9	2533.9
Fixed beta (estimate f*)	100	100	100	97	97	97	97	97
$\beta = 1$	584.0	623.4	623.4	7571.8	7571.8	7571.8	7909.4	7909.4
Fixed beta (estimate f*)	100	100	100	91	91	91	89	89
$\beta = 100$	517.5	574.0	574.0	10284.4	10284.4	10395.3	10745.5	10745.5
Exp ($\gamma = 0.99$)	100	100	100	100	100	100	100	100
$T_0 = 0.01$	405.3	1214.7	1214.7	1468.2	1468.2	1469.1	1469.1	1469.1
Exp ($\gamma = 0.99$)	100	100	100	100	100	100	100	100
$T_0 = 1$	422.6	792.3	792.3	1675.1	1675.1	1676.2	1676.5	1676.5
Exp ($\gamma = 0.99$)	100	100	100	100	100	100	100	100
$T_0 = 100$	507.2	633.0	633.0	1975.3	1975.3	1975.6	1976.1	1976.1
Exp ($\gamma = 0.01$)	100	100	100	100	100	100	100	100
$T_0 = 0.01$	409.9	1708.7	1708.7	1862.2	1862.2	1863.9	1864.6	1864.6
Exp ($\gamma = 0.01$)	100	100	100	100	100	100	100	100
$T_0 = 1$	397.2	1754.0	1754.0	1932.7	1932.7	1935.8	1935.9	1935.9
Exp ($\gamma = 0.01$)	100	100	100	100	100	100	100	100
$T_0 = 100$	376.0	1633.7	1633.7	1767.3	1767.3	1769.5	1770.0	1770.0
Log	100	100	100	100	100	100	100	100
$T_0 = 0.01$	474.46	828.04	828.04	1172	1172	1172	1172	1172
Log	100	100	100	97	97	97	95	95
$T_0 = 1$	526.5	570.4	570.4	8879.5	8879.5	8963.0	9276.8	9276.8
Log	100	100	100	87	87	87	85	85
$T_0 = 100$	532.0	586.2	586.2	10357.6	10357.6	10357.6	10723.4	10723.4

The performance of the fixed beta cooling schedule and the logarithmic cooling schedule highly depends on the choice of parameter (among three parameter values, only one of them performs well and the other two have among the four worst performance). And in most of the cases the performance of the fixed beta cooling schedule and the logarithmic cooling schedule with their best choice of parameters (among the three tested values) is worse than that of our adaptive cooling schedule with the best choice of α. The performance of the exponential cooling schedule is very close to the performance of the adaptive cooling schedule. In most of the cases, the performance

Table 3 Number of successful runs out of 100 and average number of function evaluations for successful runs for the continuous ten-dimensional sinusoidal test problem

y	1.0	1.5	2.0	2.5	3.0	3.465
Adap (given f*)	98	98	98	98	91	91
$\alpha = 0.01$	921.7	953.6	1027.3	1249.1	1733.0	3921.8
Adap (given f*)	100	100	100	100	92	92
$\alpha = 0.05$	1206.3	1237.9	1312.7	1532.4	1874.9	4100.5
Adap (given f*)	99	99	99	99	93	93
$\alpha = 0.9$	1089.9	1123.6	1208.0	1614.6	2245.2	4438.5
Adap (estimate f*)	100	91	91	91	87	87
$\alpha = 0.01$	775.1	1228.5	1297.2	1496.1	1569.7	3858.6
Adap (estimate f*)	100	91	91	91	87	87
$\alpha = 0.05$	775.2	1228.9	1297.3	1498.6	1575.2	3858.3
Adap (estimate f*)	100	91	91	91	87	87
$\alpha = 0.9$	775.4	1229.1	1297.4	1496.6	1576.3	3810.4
Fixed beta (given f*)	100	100	100	100	96	95
$\beta = 0.01$	1281.8	1315.0	1413.5	1865.1	2171.1	3875.1
Fixed beta (given f*)	100	71	2	0	0	0
$\beta = 1$	2951.3	18540.4	18918.0	–	–	–
Fixed beta (given f*)	100	61	3	0	0	0
$\beta = 100$	3345.3	20012.7	14022.3	–	–	–
Fixed beta (estimate f*)	100	91	91	91	86	86
$\beta = 0.01$	695.7	893.4	960.5	1166.2	1396.6	3691.6
Fixed beta (estimate f*)	100	99	82	43	14	1
$\beta = 1$	1016.0	2314.1	10498.8	15589.3	16582.1	10178.0
Fixed beta (estimate f*)	100	69	2	0	0	0
$\beta = 100$	3339.4	19613.0	14795.0	–	–	–
Exp ($\gamma = 0.99$)	100	89	89	89	83	83
$T_0 = 0.01$	1061.4	1847.1	1911.0	2130.4	2664.6	4866.9
Exp ($\gamma = 0.99$)	100	97	97	97	93	93
$T_0 = 1$	440.1	670.8	754.1	985.2	1231.5	3260.9
Exp ($\gamma = 0.99$)	100	99	99	99	96	96
$T_0 = 100$	864.6	995.8	1078.3	1298.9	1549.8	3544.3
Exp ($\gamma = 0.01$)	100	87	87	87	82	82
$T_0 = 0.01$	828.6	1078.8	1139.9	1339.8	1676.5	3968.4
Exp ($\gamma = 0.01$)	100	88	88	88	83	83
$T_0 = 1$	932.7	1334.1	1400.8	1610.4	1750.9	4021.9
Exp ($\gamma = 0.01$)	100	90	90	90	84	84
$T_0 = 100$	1504.8	2975.2	3039.1	3231.2	3488.3	5671.4
Log	95	89	89	89	85	85
$T_0 = 0.01$	934.65	2126.3	2190	2392.8	2907.4	5157.8
Log	100	100	100	100	100	0
$T_0 = 1$	380.8	579.0	958.7	4593.9	11367.0	–
Log	100	64	2.0	0	0	0
$T_0 = 100$	3148.0	19023.1	23262.0	–	–	–

of adaptive cooling schedule with known f^* is better than the performance using $\hat{f}_k$ estimated by Eq. 16. For fixed beta cooling schedule, the performance with known f^* for solving two test problems out of four is better than the performance using $\hat{f}_k$. We expect performance given f^* to be better than performance using an estimate because the cooling schedule can take advantage of better information. A big advantage of our robust adaptive cooling schedule is eliminating the need for trial and error to determine parameter values.

Table 4 Number of successful runs out of 100 and average number of function evaluations for successful runs for the discrete ten-dimensional sinusoidal test problem

y	1.0	1.5	2.0	2.5	3.0	3.5
Adap (given f*)	100	100	100	100	100	99
$\alpha = 0.01$	115.7	183.6	318.7	774.4	2742.3	19620.0
Adap (given f*)	100	100	100	100	100	100
$\alpha = 0.05$	113.3	173.3	319.2	748.8	2143.8	19224.9
Adap (given f*)	100	100	100	100	100	100
$\alpha = 0.9$	116.5	179.4	330.7	758.8	1874.6	19168.4
Adap (estimate f*)	100	100	100	100	100	98
$\alpha = 0.01$	126.6	211.1	441.6	942.5	3570.8	20583.9
Adap (estimate f*)	100	100	100	100	100	98
$\alpha = 0.05$	126.6	211.1	441.6	942.5	3568.7	20506.9
Adap (estimate f*)	100	100	100	100	100	98
$\alpha = 0.9$	127.2	210.8	440.1	931.4	3651.2	20614.6
Fixed beta (given f*)	100	100	100	100	100	97
$\beta = 0.01$	97.0	157.3	324.3	769.8	3653.7	18951.3
Fixed beta (given f*)	100	95	2	0	0	0
$\beta = 1$	2373.1	17530.7	9485.0	–	–	–
Fixed beta (given f*)	100	71	2	0	0	0
$\beta = 100$	2315.3	17855.8	38299.5	–	–	–
Fixed beta (estimate f*)	100	100	100	100	100	99
$\beta = 0.01$	126.7	220.3	529.9	1058.2	4495.2	21273.5
Fixed beta (estimate f*)	100	100	91	51	17	0
$\beta = 1$	409.9	2337.5	9870.0	15599.2	23929.1	–
Fixed beta (estimate f*)	100	82	3	0	0	0
$\beta = 100$	2591.9	16956.9	14490.0	–	–	–
Exp ($\gamma = 0.99$)	100	100	100	100	100	100
$T_0 = 0.01$	89.4	141.1	252.9	649.3	1873.8	18070.1
Exp ($\gamma = 0.99$)	100	100	100	100	100	99
$T_0 = 1$	282.6	362.1	503.9	1018.6	2684.4	18907.2
Exp ($\gamma = 0.99$)	100	100	100	100	99	99
$T_0 = 100$	623.0	777.5	923.6	1300.0	2755.5	19854.8
Exp ($\gamma = 0.01$)	100	100	100	100	100	100
$T_0 = 0.01$	94.5	155.0	342.4	944.5	3842.8	18963.1
Exp ($\gamma = 0.01$)	100	100	100	100	99	99
$T_0 = 1$	92.9	162.6	493.0	925.7	2017.6	16976.1
Exp ($\gamma = 0.01$)	100	100	100	100	98	96
$T_0 = 100$	97.4	193.0	411.9	1024.4	4122.0	18207.5
Log	100	100	100	100	100	99
$T_0 = 0.01$	104.05	171.37	361.37	954.48	3285.4	20734
Log	100	100	100	100	100	0
$T_0 = 1$	350.0	512.9	893.9	5455.3	14410.8	–
Log	100	86	3	0	0	0
$T_0 = 100$	2764.1	18294.6	37474.7	–	–	–

5 Summary and conclusion

Simulated annealing is a class of sequential probabilistic search techniques for solving global optimization problems. The performance of simulated annealing is highly dependent on the choice of cooling schedule employed. In this paper, we consider an algorithm called AS which is designed to model an idealized version of simulated annealing by assuming points can be sampled exactly according to a sequence of Boltzmann distributions. The paper focuses on analytically deriving cooling schedule strategies for AS applied to a general class of optimization problems over continuous and discrete domains. By choosing the temperature to guarantee an improvement with probability no less than $1-\alpha$, the cooling schedule maintains the linear in complexity property for the expected number of sample points required by AS to solve this class of global optimization problems. Some numerical results demonstrate an effective implementation of the theoretical adaptive cooling schedule.

Acknowledgements This research was supported in part by the National Science Foundation under grants DMI-9820744, DMI-9820878, DMI-0244286 and DMI-0244291; the Department of Defense Research & Engineering (DDR&E) Multidisciplinary University Research Initiative (MURI) on "Low Energy Electronics Design for Mobile Platforms" and managed by the Army Research Office (ARO) under grant ARO DAAH04-96-1-0377; and by the Department of Defense Augmentation Awards for Science and Engineering Research Training (ASSERT) on "Optimization Algorithms for Low Power Mobile Platforms" and managed by the Army Research Office under grant ARO DAAG55-98-1-0155.

Appendix A

The following lemma is important for proving the results of Theorems 1, 4 and Lemma 3.

Lemma 4 *For all $a, b \in \Re$ such that $b > a$ and $b > 0$, if $c \geq 0$, then*

$$\frac{a+c}{b+c} \geq \frac{a}{b}.$$

Appendix B

Proof of Theorem 2

Proof We first prove the result for continuous case, i.e.,

$$\pi_{h,T_k}(S_{h(X_k)}) \geq \frac{\int_{B_{\theta_k}} e^{h(x)/T_k}\,\mathrm{d}x}{\int_{B_\rho} e^{h(x)/T_k}\,\mathrm{d}x}.$$

Recall that

$$\pi_{h,T_k}(S_{h(X_k)}) = \frac{\int_{S_{h(X_k)}} e^{h(x)/T_k}\,\mathrm{d}x}{\int_S e^{h(x)/T_k}\,\mathrm{d}x}.$$

Therefore proving Theorem 2 is equivalent to proving

$$\frac{\int_{S_{h(X_k)}} e^{h(x)/T_k}\,\mathrm{d}x}{\int_{B_{\theta_k}} e^{h(x)/T_k}\,\mathrm{d}x} \geq \frac{\int_S e^{h(x)/T_k}\,\mathrm{d}x}{\int_{B_\rho} e^{h(x)/T_k}\,\mathrm{d}x}. \tag{17}$$

First we will prove that

$$\frac{v_n\left(B_\rho \setminus B_{\theta_k}\right)}{v_n\left(S \setminus S_{h(X_k)}\right)} \geq \frac{v_n\left(B_{\theta_k}\right)}{v_n\left(S_{h(X_k)}\right)}, \tag{18}$$

where $v_n(\cdot)$ is the n-dimensional volume of a set. The inequality (18) will be used to prove (17). According to the argument given in [25, p 335], we have,

$$\frac{v_n\left(S_{h(X_k)}\right)}{v_n\left(S\right)} \geq \frac{v_n\left(B_{\theta_k}\right)}{v_n\left(B_\rho\right)},$$

which is equivalent to

$$\frac{v_n\left(B_\rho\right)}{v_n\left(B_{\theta_k}\right)}-1 \geq \frac{v_n\left(S\right)}{v_n\left(S_{h(X_k)}\right)}-1$$

and considering the fact that $B_{\theta_k} \subseteq B_\rho$ and $S_{h(X_k)} \subseteq S$, we have

$$\frac{v_n\left(B_\rho \setminus B_{\theta_k}\right)}{v_n\left(B_{\theta_k}\right)} \geq \frac{v_n\left(S \setminus S_{h(X_k)}\right)}{v_n\left(S_{h(X_k)}\right)},$$

hence,

$$\frac{v_n\left(B_\rho \setminus B_{\theta_k}\right)}{v_n\left(S \setminus S_{h(X_k)}\right)} \geq \frac{v_n\left(B_{\theta_k}\right)}{v_n\left(S_{h(X_k)}\right)}. \tag{19}$$

Now consider the left-hand side of (17). We have,

$$\begin{aligned}\frac{\int_{S_{h(X_k)}} e^{h(x)/T_k}\,\mathrm{d}x}{\int_{B_{\theta_k}} e^{h(x)/T_k}\,\mathrm{d}x} &= \frac{\left(\int_{S_{h(X_k)}} e^{h(x)/T_k}\,\mathrm{d}x\right)\left(\int_{S_{h(X_k)}} e^{h(x)/T_k}\,\mathrm{d}x+\int_{S\setminus S_{h(X_k)}} e^{f(X_k)/T_k}\,\mathrm{d}x\right)}{\left(\int_{B_{\theta_k}} e^{h(x)/T_k}\,\mathrm{d}x\right)\left(\int_{S_{h(X_k)}} e^{h(x)/T_k}\,\mathrm{d}x+\int_{S\setminus S_{h(X_k)}} e^{f(X_k)/T_k}\,\mathrm{d}x\right)}\\ &= \frac{\int_{S_{h(X_k)}} e^{h(x)/T_k}\,\mathrm{d}x+\int_{S\setminus S_{h(X_k)}} e^{f(X_k)/T_k}\,\mathrm{d}x}{\int_{B_{\theta_k}} e^{h(x)/T_k}\,\mathrm{d}x+\frac{\int_{B_{\theta_k}} e^{h(x)/T_k}\,\mathrm{d}x}{\int_{S_{h(X_k)}} e^{h(x)/T_k}\,\mathrm{d}x}v_n\left(S \setminus S_{h(X_k)}\right)e^{f(X_k)/T_k}}.\end{aligned}$$

According to the fact that $S_{h(X_k)} \subseteq B_{\theta_k}$ and applying Lemma 4.1 given in [23] with $T_k < T_\infty$, we have $\frac{\int_{S_{h(X_k)}} e^{h(x)/T_k}\,\mathrm{d}x}{\int_{B_{\theta_k}} e^{h(x)/T_k}\,\mathrm{d}x} \geq \frac{v_n(S_{h(X_k)})}{v_n(B_{\theta_k})}$. Therefore,

$$\frac{\int_{S_{h(X_k)}} e^{h(x)/T_k}\,\mathrm{d}x}{\int_{B_{\theta_k}} e^{h(x)/T_k}\,\mathrm{d}x} \geq \frac{\int_{S_{h(X_k)}} e^{h(x)/T_k}\,\mathrm{d}x+\int_{S\setminus S_{h(X_k)}} e^{f(X_k)/T_k}\,\mathrm{d}x}{\int_{B_{\theta_k}} e^{h(x)/T_k}\,\mathrm{d}x+\frac{v_n\left(B_{\theta_k}\right)}{v_n\left(S_{h(X_k)}\right)}v_n\left(S \setminus S_{h(X_k)}\right)e^{f(X_k)/T_k}}$$

and because of the fact that $\frac{v_n\left(B_\rho \setminus B_{\theta_k}\right)}{v_n\left(S\setminus S_{h(X_k)}\right)} \geq \frac{v_n\left(B_{\theta_k}\right)}{v_n\left(S_{h(X_k)}\right)}$ in inequality (19), we have

$$\begin{aligned}\frac{\int_{S_{h(X_k)}} e^{h(x)/T_k}\,\mathrm{d}x}{\int_{B_{\theta_k}} e^{h(x)/T_k}\,dx} &\geq \frac{\int_{S_{h(X_k)}} e^{h(x)/T_k}\,\mathrm{d}x+\int_{S\setminus S_{h(X_k)}} e^{f(X_k)/T_k}\,\mathrm{d}x}{\int_{B_{\theta_k}} e^{h(x)/T_k}\,\mathrm{d}x+\frac{v_n\left(B_\rho \setminus B_{\theta_k}\right)}{v_n\left(S\setminus S_{h(X_k)}\right)}v_n\left(S \setminus S_{h(X_k)}\right)e^{f(X_k)/T_k}}\\ &= \frac{\int_{S_{h(X_k)}} e^{h(x)/T_k}\,\mathrm{d}x+\int_{S\setminus S_{h(X_k)}} e^{f(X_k)/T_k}\,\mathrm{d}x}{\int_{B_{\theta_k}} e^{h(x)/T_k}\,\mathrm{d}x+v_n\left(B_\rho \setminus B_{\theta_k}\right)e^{f(X_k)/T_k}}\\ &= \frac{\int_{S_{h(X_k)}} e^{h(x)/T_k}\,\mathrm{d}x+\int_{S\setminus S_{h(X_k)}} e^{f(X_k)/T_k}\,\mathrm{d}x}{\int_{B_{\theta_k}} e^{h(x)/T_k}\,\mathrm{d}x+\int_{B_\rho\setminus B_{\theta_k}} e^{f(X_k)/T_k}\,\mathrm{d}x}\\ &= \frac{\int_S e^{h(x)/T_k}\,\mathrm{d}x}{\int_{B_\rho} e^{h(x)/T_k}\,\mathrm{d}x}.\end{aligned}$$

Hence

$$\pi_{h,T_k}(S_{h(X_k)}) = \frac{\int_{S_{h(X_k)}} e^{h(x)/T_k}\,dx}{\int_S e^{h(x)/T_k}\,dx} \geq \frac{\int_{B_{\theta_k}} e^{h(x)/T_k}\,dx}{\int_{B_\rho} e^{h(x)/T_k}\,dx}.$$

Next, we prove the result for discrete case, i.e.,

$$\tilde{\pi}_{\tilde{h},\tilde{T}_k}(\tilde{S}_{\tilde{h}(\tilde{X}_k)}) \geq \frac{\sum_{\tilde{x}\in(\tilde{B}_{\tilde{\theta}_k}\cap Z^n)} e^{\tilde{h}(\tilde{x})/\tilde{T}_k}}{\sum_{\tilde{x}\in(\tilde{B}_{\tilde{\rho}}\cap Z^n)} e^{\tilde{h}(\tilde{x})/\tilde{T}_k}}. \tag{20}$$

To prove the result, we first prove that

$$\frac{|\tilde{S}_{\tilde{h}(\tilde{X}_k)}|}{|\tilde{S}|} \geq \frac{|\tilde{B}_{\tilde{\theta}_k} \cap Z^n|}{|\tilde{B}_{\tilde{\rho}} \cap Z^n|}. \tag{21}$$

And then given the inequality (21), inequality (20) can be proved following the same procedure used for the continuous case.

Let $\tilde{B}$ be a continuous set composed of the union of hypercubes surrounding all integer points in $\tilde{S}$, and let $\tilde{B}_{\tilde{h}(\tilde{X}_k)}$ be the continuous set composed of the union of hypercubes surrounding all integer points in $\tilde{S}_{\tilde{h}(\tilde{X}_k)}$, i.e.,

$$\tilde{B} = \{x \in \Re^n : \text{Round}(x) \in \tilde{S}\}$$
$$\tilde{B}_{\tilde{h}(\tilde{X}_k)} = \{x \in \Re^n : \text{Round}(x) \in \tilde{S}_{\tilde{h}(\tilde{X}_k)}\}.$$

Since we assume that $\tilde{S}$ is a discrete region over a hyperrectangle, according to the definition of $\tilde{h}$, the improving set $\tilde{S}_{\tilde{h}(\tilde{X}_k)}$ is also a discrete hyperrectangular region. Therefore,

$$\frac{|\tilde{S}_{\tilde{h}(\tilde{X}_k)}|}{|\tilde{S}|} = \frac{v_n(\tilde{B}_{\tilde{h}(\tilde{X}_k)})}{v_n(\tilde{B})}.$$

Now consider the four continuous sets, $\tilde{B}$, $\tilde{B}_{\tilde{h}(\tilde{X}_k)}$, $\tilde{B}_{\tilde{\theta}_k}$, and $\tilde{B}_{\tilde{\rho}}$. We know that $\tilde{B}$ is a convex set, $\tilde{B}_{\tilde{h}(\tilde{X}_k)} \in \tilde{B}$, $\tilde{B}_{\tilde{\theta}_k}$ and $\tilde{B}_{\tilde{\rho}}$ are hyperrectangles with center $\tilde{x}^*$, and $\tilde{B}_{\tilde{h}(\tilde{X}_k)} \subseteq \tilde{B}_{\tilde{\theta}_k}$, $\tilde{B} \subseteq \tilde{B}_{\tilde{\rho}}$, and $\tilde{B}_{\tilde{\theta}_k} \subseteq \tilde{B}_{\tilde{\rho}}$. According to the argument given in [25, p 335], we have,

$$\frac{v_n(\tilde{B}_{\tilde{h}(\tilde{X}_k)})}{v_n(\tilde{B})} \geq \frac{v_n(\tilde{B}_{\tilde{\theta}_k})}{v_n(\tilde{B}_{\tilde{\rho}})}$$

and

$$\frac{v_n(\tilde{B}_{\tilde{\theta}_k})}{v_n(\tilde{B}_{\tilde{\rho}})} = \frac{|\tilde{B}_{\tilde{\theta}_k} \cap Z^n|}{|\tilde{B}_{\tilde{\rho}} \cap Z^n|},$$

because $\tilde{B}_{\tilde{\theta}_k}$ and $\tilde{B}_{\tilde{\rho}}$ are defined such that the volume of each set equals the number of discrete points contained in the set. Therefore,

$$\frac{|\tilde{S}_{\tilde{h}(\tilde{X}_k)}|}{|\tilde{S}|} \geq \frac{|\tilde{B}_{\tilde{\theta}_k} \cap Z^n|}{|\tilde{B}_{\tilde{\rho}} \cap Z^n|}.$$

We have now proved Eq. 21. Following the same procedure used for the continuous case we could further prove the desired result for the discrete case in Theorem 2. □

Appendix C

This appendix contains a lemma used in proof of Theorem 4. Based on the definition of $\tilde{h}(x)$, the following lemma holds.

Lemma 5 *Given the current record value $\tilde{X}_k$, one has,*

$$\text{if } \lfloor \frac{\tilde{f}^* - f(\tilde{X}_k)}{\tilde{K}} \rfloor < 1, \quad \sum_{\tilde{x} \in (\tilde{B}_{\tilde{\theta}_k} \cap Z^n)} e^{\tilde{h}(\tilde{x})/\tilde{T}_k} = e^{\tilde{f}^*/\tilde{T}_k},$$

$$\text{otherwise,} \quad \sum_{\tilde{x} \in (\tilde{B}_{\tilde{\theta}_k} \cap Z^n)} e^{\tilde{h}(\tilde{x})/\tilde{T}_k} \geq \int_0^{\tilde{\delta}_k} e^{(\hat{f}^* - u\tilde{K})/\tilde{T}_k} 2n \cdot (2u)^{n-1} \mathrm{d}u,$$

where $\hat{f}^* = \tilde{f}^* - 0.5\tilde{K}$ and $\tilde{\delta}_k = \frac{\hat{f}^* - f(\tilde{X}_k)}{\tilde{K}}$.

Proof As defined in Theorem 2, the set $(\tilde{B}_{\tilde{\theta}_k} \cap Z^n)$ is an n-dimensional hypercube with center at $\tilde{x}^*$ and the length of each side is $2(\tilde{f}^* - \tilde{f}(\tilde{X}_k))/\tilde{K}$, i.e.,

$$\tilde{B}_{\tilde{\theta}_k} \cap Z^n = \{\tilde{x} \in Z^n : \max_{i=1,\ldots,n} (|\tilde{x}_i - \tilde{x}_i^*|) \leq (\tilde{f}^* - \tilde{f}(\tilde{X}_k))/\tilde{K}\} \tag{22}$$

and because $\tilde{x}^*$ must also be integer,

$$\tilde{B}_{\tilde{\theta}_k} \cap Z^n = \{\tilde{x} \in Z^n : \max_{i=1,\ldots,n} (|\tilde{x}_i - \tilde{x}_i^*|) \leq \lfloor \tilde{f}^* - \tilde{f}(\tilde{X}_k))/\tilde{K} \rfloor\}.$$

Obviously, if $\lfloor \tilde{f}^* - \tilde{f}(\tilde{X}_k))/\tilde{K} \rfloor < 1$, there is only one integer point in $\tilde{B}_{\tilde{\theta}_k}$. Hence one has

$$\sum_{\tilde{x} \in \tilde{B}_{\tilde{\theta}_k} \cap Z^n} e^{\tilde{h}(\tilde{x})/\tilde{T}_k} = e^{\tilde{f}^*/\tilde{T}_k}.$$

For $\lfloor \tilde{f}^* - \tilde{f}(\tilde{X}_k))/\tilde{K} \rfloor \geq 1$, let $m = \lfloor \tilde{f}^* - \tilde{f}(\tilde{X}_k))/\tilde{K} \rfloor$, and for a given $j \in \{1, \ldots, m\}$, define S_j, a subset of $\tilde{B}_{\tilde{\theta}_k} \cap Z^n$, as follows

$$S_j = \{\tilde{x} \in Z^n : \max_{i=1,\ldots,n} (|\tilde{x}_i - \tilde{x}_i^*|) = j\},$$

i.e., S_j is a collection of lattice points contained on the surface of the box $[\tilde{x}_1^* - j, \tilde{x}_1^* + j] \times \cdots \times [\tilde{x}_n^* - j, \tilde{x}_n^* + j]$. According to the definition of $\tilde{h}$, one has $\tilde{h}(\tilde{x}) = \tilde{f}^* - j\tilde{K}$ for all $\tilde{x} \in S_j$. Therefore

$$\sum_{\tilde{x} \in \tilde{B}_{\tilde{\theta}_k} \cap Z^n} e^{\tilde{h}(\tilde{x})/\tilde{T}_k} = e^{\tilde{f}^*/\tilde{T}_k} + \sum_{j=1}^{m} \left(|S_j| e^{(\tilde{f}^* - j\tilde{K})/\tilde{T}_k} \right), \tag{23}$$

where $|S_j|$ represents the number of lattice points contained in the set S_j.

To calculate $|S_j|$, we first define the rounding region of lattice point. Let the box $B_{\tilde{x}} = [\tilde{x}_1 - 0.5, \tilde{x}_1 + 0.5] \times \cdots \times [\tilde{x}_n - 0.5, \tilde{x}_n + 0.5]$ denote the rounding region of the lattice point $\tilde{x}$, i.e., $B_{\tilde{x}}$ is an n-dimensional hypercube of unit length with $\tilde{x}$ being the center of the hypercube. Since $v_n(B_{\tilde{x}}) = 1$, one has

$$|S_j| = \sum_{\tilde{x} \in S_j} v_n(B_{\tilde{x}}) = v_n(\cup_{\tilde{x} \in S_j} B_{\tilde{x}}),$$
$$\cup_{\tilde{x} \in S_j} B_{\tilde{x}} = \{\tilde{x} \in \Re^n : j - 0.5 \le \max_{i=1,\dots,n} (|\tilde{x}_i - \tilde{x}_i^*|) \le j + 0.5\}.$$

Now considering two boxes $B_L = [\tilde{x}_1^* - j + 0.5, \tilde{x}_1^* + j - 0.5] \times \cdots \times [\tilde{x}_n^* - j + 0.5, \tilde{x}_n^* + j - 0.5]$ and $B_U = [\tilde{x}_1^* - j - 0.5, \tilde{x}_1^* + j + 0.5] \times \cdots \times [\tilde{x}_n^* - j - 0.5, \tilde{x}_n^* + j + 0.5]$, one has

$$v_n(\cup_{\tilde{x} \in S_j} B_{\tilde{x}}) = v_n(B_U \setminus B_L) = v_n(B_U) - v_n(B_L) = (2j+1)^n - (2j-1)^n,$$

where the second equality follows from the fact $B_L \subset B_U$. Hence

$$|S_j| = (2j+1)^n - (2j-1)^n. \tag{24}$$

Thus, by applying Eq. 24 to Eq. 23, we get that

$$\sum_{\tilde{x} \in \tilde{B}_{\tilde{\theta}_k} \cap Z^n} e^{\tilde{h}(\tilde{x})/\tilde{T}_k} = e^{\tilde{f}^*/\tilde{T}_k} + \sum_{j=1}^{m} \left((2j+1)^n - (2j-1)^n\right) e^{(\tilde{f}^* - j\tilde{K})/\tilde{T}_k}.$$

Considering that

$$\int_{j-0.5}^{j+0.5} e^{(\tilde{f}^* - j\tilde{K})/\tilde{T}_k} \cdot 2n \cdot (2u)^{n-1} \mathrm{d}u = e^{(\tilde{f}^* - j\tilde{K})/\tilde{T}_k} \int_{j-0.5}^{j+0.5} d(2u)^n$$
$$= \left((2j+1)^n - (2j-1)^n\right) e^{(\tilde{f}^* - j\tilde{K})/\tilde{T}_k}$$

and because $\lfloor \tilde{f}^* - \tilde{f}(\tilde{X}_k)/\tilde{K} \rfloor \ge 1$, one has

$$\sum_{\tilde{x} \in \tilde{B}_{\tilde{\theta}_k} \cap Z^n} e^{\tilde{h}(\tilde{x})/\tilde{T}_k} = e^{\tilde{f}^*/\tilde{T}_k} + \sum_{j=1}^{m} \left(\int_{j-0.5}^{j+0.5} e^{(\tilde{f}^* - j\tilde{K})/\tilde{T}_k} \cdot 2n \cdot (2u)^{n-1} \mathrm{d}u \right)$$

and because $\tilde{f}^* - j\tilde{K} \ge \tilde{f}^* - (u + 0.5)\tilde{K} = \hat{f}^* - u\tilde{K}$ for all $u \in [j - 0.5, j + 0.5]$ and where $\hat{f}^* = \tilde{f}^* - 0.5\tilde{K}$, one has

$$\ge e^{\tilde{f}^*/\tilde{T}_k} + \sum_{j=1}^{m} \left(\int_{j-0.5}^{j+0.5} e^{(\hat{f}^* - u\tilde{K})/\tilde{T}_k} \cdot 2n \cdot (2u)^{n-1} \mathrm{d}u \right)$$
$$= \int_{0}^{0.5} e^{\tilde{f}^*/\tilde{T}_k} \cdot 2n \cdot (2u)^{n-1} \mathrm{d}u + \int_{0.5}^{m+0.5} e^{(\hat{f}^* - u\tilde{K})/\tilde{T}_k} \cdot 2n \cdot (2u)^{n-1} \mathrm{d}u$$

and apply the fact that $\tilde{f}^* > \tilde{f}^* - (u + 0.5)\tilde{K} = \hat{f}^* - u\tilde{K}$ for all $u \in [0, 0.5]$, one has

$$> \int_{0}^{0.5} e^{(\hat{f}^* - u\tilde{K})/\tilde{T}_k} \cdot 2n \cdot (2u)^{n-1} \mathrm{d}u + \int_{0.5}^{m+0.5} e^{(\hat{f}^* - u\tilde{K})/\tilde{T}_k} \cdot 2n \cdot (2u)^{n-1} \mathrm{d}u$$
$$= \int_{0}^{m+0.5} e^{(\hat{f}^* - u\tilde{K})/\tilde{T}_k} \cdot 2n \cdot (2u)^{n-1} \mathrm{d}u$$

and since $m+0.5 = \lfloor \frac{\tilde{f}^*-\tilde{f}(\tilde{X}_k)}{\tilde{K}} \rfloor + 0.5 \geq \frac{\tilde{f}^*-\tilde{f}(\tilde{X}_k)}{\tilde{K}} - 1 + 0.5$, one has

$$\geq \int_0^{\frac{\tilde{f}^*-\tilde{f}(\tilde{X}_k)}{\tilde{K}}-0.5} e^{(\hat{f}^*-u\tilde{K})/\tilde{T}_k} \cdot 2n \cdot (2u)^{n-1} \mathrm{d}u$$

and setting $\tilde{\delta}_k = \frac{\hat{f}^*-\tilde{f}(\tilde{X}_k)}{\tilde{K}} = \frac{\tilde{f}^*-\tilde{f}(\tilde{X}_k)}{\tilde{K}} - 0.5$, yields the result,

$$\geq \int_0^{\tilde{\delta}_k} e^{(\hat{f}^*-u\tilde{K})/\tilde{T}_k} \cdot 2n \cdot (2u)^{n-1} \mathrm{d}u.$$

□

Appendix D

Proof of Lemma 1

Proof We prove the lemma by analyzing the shape of the function $p_{\mathcal{M}}(z)$. According to the definition, z is a one-dimensional positive variable. It is easy to prove that $\lim_{z\to 0} p_{\mathcal{M}}(z) = 0$. Next let us look at the limit of $p_{\mathcal{M}}(z)$ as z goes to infinity,

$$\lim_{z\to\infty} p_{\mathcal{M}}(z) = (n-1)! - \lim_{z\to\infty} \sum_{i=0}^{n-1} \frac{(n-1)!}{i!} \frac{z^i}{e^z} - \lim_{z\to\infty} \mathcal{M} \frac{z^n}{e^z}.$$

Because $\frac{z^i}{e^z} \to 0$ as $z \to \infty$, for all i, we can see that

$$\lim_{z\to\infty} p_{\mathcal{M}}(z) = (n-1)!.$$

Now consider the derivative of $p_{\mathcal{M}}(z)$,

$$p'_{\mathcal{M}}(z) = \frac{dp(z)}{dz} = (1-n\mathcal{M})z^{n-1}e^{-z} + \mathcal{M}z^n e^{-z}.$$

If $n\mathcal{M} \leq 1$, for z positive, one has $p'_{\mathcal{M}}(z) > 0$. In this case, we have,

(1) $\lim_{z\to 0} p_{\mathcal{M}}(z) = 0$,
(2) $\lim_{z\to\infty} p_{\mathcal{M}}(z) = (n-1)!$
(3) $p'_{\mathcal{M}}(z) > 0$,

which implies that the function $p_{\mathcal{M}}(z)$ is monotonically increasing and concave as depicted in Fig. (1a). Therefore the value of $p_{\mathcal{M}}(z)$ is positive for $z > 0$.

On the other hand, if $n\mathcal{M} > 1$, i.e., $n\mathcal{M} - 1 > 0$, we have,

$$p'_{\mathcal{M}}(z) \begin{cases} < 0 & \text{if } 0 < z < (n\mathcal{M}-1)/\mathcal{M}, \\ = 0 & \text{if } z = (n\mathcal{M}-1)/\mathcal{M}, \\ > 0 & \text{if } z > (n\mathcal{M}-1)/\mathcal{M}. \end{cases}$$

According to the property of $p'_{\mathcal{M}}(z)$ for the case $n\mathcal{M} > 1$, and the fact that $\lim_{z\to 0} p_{\mathcal{M}}(z) = 0$ and $\lim_{z\to\infty} p_{\mathcal{M}}(z) = (n-1)! > 0$, one can draw the function shape of $p_{\mathcal{M}}(z)$, as shown in Fig. (1b). Therefore, for $n\mathcal{M} > 1$, the equation $p_{\mathcal{M}}(z) = 0$ with $z > 0$ has a unique solution z^*.

The lower bound of z^*, i.e. $z^* \geq \frac{n\mathcal{M}-1}{\mathcal{M}}$, is immediate from the fact that $\lim_{z\to 0} p_{\mathcal{M}}(z) = 0$, $\lim_{z\to\infty} p_{\mathcal{M}}(z) = (n-1)! > 0$ and the property of $p'_{\mathcal{M}}(z) = 0$ at $z = (n\mathcal{M}-1)/\mathcal{M}$. □

Appendix E

Proof of Lemma 2

Proof According to Theorem 3, solving for T_k^* in $\frac{\int_{B_{\theta_k}} e^{h(x)/T_k}\,\mathrm{d}x}{\int_{B_\rho} e^{h(x)/T_k}\,\mathrm{d}x} = 1-\alpha$ is equivalent to solving for z_k^* in $p_{M_k}(z_k^*) = 0$ and setting $T_k^* = (f^* - f(X_k))/z_k^*$. Therefore, to prove that $L_k/(1-\varepsilon)$ is an upper bound on z_k^*, we prove an associated lower bound on T_k^*.

To establish the lower bound on T_k^*, for $0 < \varepsilon < 1$, we define a function $g\colon\ B_\rho \to \Re$ by

$$g(x) = \begin{cases} f^* - \varepsilon(f^* - f(X_k)), & \text{if } x \in B_{\hat{\varepsilon}_k}, \\ f(X_k), & \text{if } x \in B_\rho \setminus B_{\hat{\varepsilon}_k}, \end{cases}$$

where $B_{\hat{\varepsilon}_k} = \{x \in B_\rho : \|x - x^*\| < \hat{\varepsilon}_k = \varepsilon(f^* - f(X_k))/K\}$. Then $h(x) \geq g(x)$ on B_{θ_k}, but $h(x) = g(x)$ on $B_\rho \setminus B_{\theta_k}$. Also $B_{\hat{\varepsilon}_k} \subseteq B_{\theta_k}$. Following the same arguments in the proof of Theorem 1, one can show that

$$\frac{\int_{B_{\theta_k}} e^{h(x)/T_k}\,\mathrm{d}x}{\int_{B_\rho} e^{h(x)/T_k}\,\mathrm{d}x} \geq \frac{\int_{B_{\hat{\varepsilon}_k}} e^{g(x)/T_k}\,\mathrm{d}x}{\int_{B_\rho} e^{g(x)/T_k}\,\mathrm{d}x}.$$

Now consider $\frac{\int_{B_{\hat{\varepsilon}_k}} e^{g(x)/T_k}\,\mathrm{d}x}{\int_{B_\rho} e^{g(x)/T_k}\,\mathrm{d}x}$, we have

$$\begin{aligned}\frac{\int_{B_{\hat{\varepsilon}_k}} e^{g(x)/T_k}\,\mathrm{d}x}{\int_{B_\rho} e^{g(x)/T_k}\,\mathrm{d}x} &= \frac{\int_{B_{\hat{\varepsilon}_k}} e^{(f^*-\varepsilon(f^*-f(X_k)))/T_k}\mathrm{d}x}{\int_{B_\rho \setminus B_{\hat{\varepsilon}_k}} e^{f(X_k)/T_k}\mathrm{d}x + \int_{B_{\hat{\varepsilon}_k}} e^{(f^*-\varepsilon(f^*-f(X_k)))/T_k}\mathrm{d}x} \\ &= \frac{v_n(B_{\hat{\varepsilon}_k}) \cdot e^{(f^*-\varepsilon(f^*-f(X_k)))/T_k}}{(v_n(B_\rho) - v_n(B_{\hat{\varepsilon}_k})) \cdot e^{f(X_k)/T_k} + v_n(B_{\hat{\varepsilon}_k}) \cdot e^{(f^*-\varepsilon(f^*-f(X_k)))/T_k}}.\end{aligned}$$

After algebraic manipulations, it can be seen that if we let

$$T^l = \frac{(1-\varepsilon)(f^* - f(X_k))}{L_k},$$

where

$$L_k = \ln\left(\frac{1-\alpha}{\alpha} \cdot \frac{v_n(B_\rho) - v_n(B_{\hat{\varepsilon}_k})}{v_n(B_{\hat{\varepsilon}_k})}\right).$$

then T^l satisfies

$$\frac{\int_{B_{\theta_k}} e^{h(x)/T^l}\,\mathrm{d}x}{\int_{B_\rho} e^{h(x)/T^l}\,\mathrm{d}x} \geq \frac{\int_{B_{\hat{\varepsilon}_k}} e^{g(x)/T^l}\,\mathrm{d}x}{\int_{B_\rho} e^{g(x)/T^l}\,\mathrm{d}x} = 1-\alpha.$$

Since T_k^* is the solution to $\frac{\int_{B_{\theta_k}} e^{h(x)/T_k}\,\mathrm{d}x}{\int_{B_\rho} e^{h(x)/T_k}\,\mathrm{d}x} = 1-\alpha$, and we have

$$\frac{\int_{B_{\theta_k}} e^{h(x)/T^l}\,\mathrm{d}x}{\int_{B_\rho} e^{h(x)/T^l}\,\mathrm{d}x} \geq \frac{\int_{B_{\theta_k}} e^{h(x)/T_k^*}\,\mathrm{d}x}{\int_{B_\rho} e^{h(x)/T_k^*}\,\mathrm{d}x},$$

we know

$$T^l \le T_k^*$$

(also see [19, Proposition 3.9]). Thus T^l is a lower bound on T_k^*, and equivalently, $\frac{L_k}{1-\varepsilon}$ is an upper bound on z_k^*. □

Proof of Lemma 3

Proof According to Theorem 4, solving for $\tilde{z}_k^*$ in $p_{\tilde{M}_k}(\tilde{z}_k) = 0$ is equivalent to solving for $\tilde{T}_k^*$ in the equation

$$\frac{\int_0^{\tilde{\delta}_k} e^{(\hat{f}^* - u\tilde{K})/\tilde{T}_k} \cdot 2n \cdot (2u)^{n-1} \mathrm{d}u}{\int_0^{\tilde{\delta}_k} e^{(\hat{f}^* - u\tilde{K})/\tilde{T}_k} \cdot 2n \cdot (2u)^{n-1} \mathrm{d}u + \left(|\tilde{B}_{\tilde{\rho}} \cap Z^n| - (2\tilde{\delta}_k)^n\right) e^{\tilde{f}(\tilde{X}_k)/\tilde{T}_k}} = 1 - \alpha \quad (25)$$

and setting $\tilde{z}_k^* = (\tilde{f}^* - \tilde{f}(\tilde{X}_k))/\tilde{T}_k^*$, where $\tilde{\delta}_k = \frac{\hat{f}^* - \tilde{f}(\tilde{X}_k)}{\tilde{K}}$. Therefore, to prove that $\tilde{L}_k/(1-\varepsilon)$ is an upper bound on $\tilde{z}_k^*$, we can equivalently prove an associated lower bound on $\tilde{T}_k^*$.

To establish the lower bound on $\tilde{T}_k^*$, we first define a function $\hat{h}(x)$ and sets $\hat{S}$ and $\tilde{G}$ so we can write Eq. 25 as

$$\frac{\int_{\tilde{G}} e^{\hat{h}(x)/\tilde{T}_k} \mathrm{d}x}{\int_{\hat{S}} e^{\hat{h}(x)/\tilde{T}_k} \mathrm{d}x} = \frac{\int_0^{\tilde{\delta}_k} e^{(\hat{f}^* - u\tilde{K})/\tilde{T}_k} \cdot 2n \cdot (2u)^{n-1} \mathrm{d}u}{\int_0^{\tilde{\delta}_k} e^{(\hat{f}^* - u\tilde{K})/\tilde{T}_k} \cdot 2n \cdot (2u)^{n-1} \mathrm{d}u + \left(|\tilde{B}_{\tilde{\rho}} \cap Z^n| - (2\tilde{\delta}_k)^n\right) e^{\tilde{f}(\tilde{X}_k)/\tilde{T}_k}}$$
$$= 1 - \alpha$$

where $\hat{S} = \{x \in \Re^n : \mathrm{Round}(x) \in (\tilde{B}_{\tilde{\rho}} \cap Z^n)\}$, $\tilde{G} = \{x \in \hat{S} : \max_{i=1,\ldots,n}(|x_i - \tilde{x}_i^*|) \le (\hat{f}^* - \tilde{f}(\tilde{X}_k))/\tilde{K}\}$ and

$$\hat{h}(x) = \begin{cases} \hat{f}^* - \tilde{K} \max_{i=1,\ldots,n}(|x_i - \tilde{x}_i^*|), & \text{if } x \in \tilde{G}, \\ \tilde{f}(\tilde{X}_k), & \text{if } x \in \hat{S} \setminus \tilde{G}. \end{cases}$$

For $0 < \varepsilon < 1$, we now define a set $\tilde{G}_{\hat{\varepsilon}}$ and a function $\tilde{g}(x)$ as

$$\tilde{G}_{\hat{\varepsilon}} = \{x \in \hat{S} : \max_{i=1,\ldots,n} (|x_i - \tilde{x}_i^*|) \le \hat{\varepsilon} = \varepsilon(\hat{f}^* - \tilde{f}(\tilde{X}_k))/\tilde{K}\}$$

$$\tilde{g}(x) = \begin{cases} \hat{f}^* - \varepsilon(\hat{f}^* - \tilde{f}(\tilde{X}_k)), & \text{if } x \in \tilde{G}_{\hat{\varepsilon}}, \\ \tilde{f}(\tilde{X}_k), & \text{if } x \in \hat{S} \setminus \tilde{G}_{\hat{\varepsilon}}, \end{cases}$$

so that

$$\frac{\int_{\tilde{G}} e^{\hat{h}(x)/\tilde{T}_k} \mathrm{d}x}{\int_{\hat{S}} e^{\hat{h}(x)/\tilde{T}_k} \mathrm{d}x} \ge \frac{\int_{\tilde{G}_{\hat{\varepsilon}}} e^{\tilde{g}(x)/\tilde{T}_k} \mathrm{d}x}{\int_{\hat{S}} e^{\tilde{g}(x)/\tilde{T}_k} \mathrm{d}x}.$$

This is true because $\tilde{G}_{\hat{\varepsilon}} \subseteq \tilde{G}$ and $\hat{h}(x) \geq \tilde{g}(x)$ for $x \in \tilde{G}$, and $\hat{h}(x) = \tilde{g}(x)$ for $x \in \hat{S} \setminus \tilde{G}$. Now we have

$$\frac{\int_{\tilde{G}_{\hat{\varepsilon}}} e^{\tilde{g}(x)/\tilde{T}_k}\, \mathrm{d}x}{\int_{\hat{S}} e^{\tilde{g}(x)/\tilde{T}_k}\, \mathrm{d}x} = \frac{\int_{\tilde{G}_{\hat{\varepsilon}}} e^{(\hat{f}^* - \varepsilon(\hat{f}^* - \tilde{f}(\tilde{X}_k)))/\tilde{T}_k} \mathrm{d}x}{\int_{\hat{S}\setminus \tilde{G}_{\hat{\varepsilon}}} e^{\tilde{f}(\tilde{X}_k)/\tilde{T}_k} \mathrm{d}x + \int_{\tilde{G}_{\hat{\varepsilon}}} e^{(\hat{f}^* - \varepsilon(\hat{f}^* - \tilde{f}(\tilde{X}_k)))/\tilde{T}_k} \mathrm{d}x}$$
$$= \frac{v_n(\tilde{G}_{\hat{\varepsilon}}) \cdot e^{(\hat{f}^* - \varepsilon(\hat{f}^* - \tilde{f}(\tilde{X}_k)))/\tilde{T}_k}}{(v_n(\hat{S}) - v_n(\tilde{G}_{\hat{\varepsilon}})) \cdot e^{\tilde{f}(\tilde{X}_k)/\tilde{T}_k} + v_n(\tilde{G}_{\hat{\varepsilon}}) \cdot e^{(\hat{f}^* - \varepsilon(\hat{f}^* - \tilde{f}(\tilde{X}_k)))/\tilde{T}_k}},$$

where $v_n(\tilde{G}_{\hat{\varepsilon}}) = (2\varepsilon \cdot \tilde{\delta}_k)^n$ and $v_n(\hat{S}) = |\tilde{B}_{\tilde{\rho}} \cap Z^n| = |\tilde{S}|$. After algebraic manipulations, it can be seen that if we let

$$\tilde{T}^l = \frac{(1-\varepsilon)(\hat{f}^* - \tilde{f}(\tilde{X}_k))}{\tilde{L}_k},$$

where

$$\tilde{L}_k = \ln\left(\frac{1-\alpha}{\alpha} \cdot \frac{|\tilde{S}| - (2\varepsilon\tilde{\delta}_k)^n}{(2\varepsilon\tilde{\delta}_k)^n}\right),$$

then $\tilde{T}^l$ satisfies

$$\frac{\int_{\tilde{G}} e^{\hat{h}(x)/\tilde{T}^l}\, \mathrm{d}x}{\int_{\hat{S}} e^{\hat{h}(x)/\tilde{T}^l}\, \mathrm{d}x} \geq \frac{\int_{\tilde{G}_{\hat{\varepsilon}}} e^{\tilde{g}(x)/\tilde{T}^l}\, \mathrm{d}x}{\int_{\hat{S}} e^{\tilde{g}(x)/\tilde{T}^l}\, \mathrm{d}x} = 1 - \alpha.$$

Since $\tilde{T}_k^*$ is the solution to $\frac{\int_{\tilde{G}} e^{\hat{h}(x)/\tilde{T}_k}\, \mathrm{d}x}{\int_{\hat{S}} e^{\hat{h}(x)/\tilde{T}_k}\, \mathrm{d}x} = 1 - \alpha$, and we have

$$\frac{\int_{\tilde{G}} e^{\hat{h}(x)/\tilde{T}^l}\, \mathrm{d}x}{\int_{\hat{S}} e^{\hat{h}(x)/\tilde{T}^l}\, \mathrm{d}x} \geq \frac{\int_{\tilde{G}} e^{\hat{h}(x)/\tilde{T}_k^*}\, \mathrm{d}x}{\int_{\hat{S}} e^{\hat{h}(x)/\tilde{T}_k^*}\, \mathrm{d}x},$$

we know

$$\tilde{T}^l \leq \tilde{T}_k^*$$

(also see [19, Proposition 3.9]). Thus $\tilde{T}^l$ is a lower bound on $\tilde{T}_k^*$, and equivalently, $\frac{\tilde{L}_k}{1-\varepsilon}$ is an upper bound on $\tilde{z}_k^*$. □

Appendix F

F.1 Test problem 1: six-dimensional Hartmann problem over a continuous domain [19]

$$f(x) = \sum_{i=1}^{4} c_i \exp\left(-\sum_{j=1}^{6} a_{ij} \left(\frac{1}{j} x_j - p_{ij}\right)^2\right),$$

$$\text{s.t} \quad 0 \leq x_j \leq j, \qquad \text{for } j = 1, \ldots, 6,$$

where $c_1 = 1, c_2 = 1.2, c_3 = 3, c_4 = 3.2$,

$$\begin{pmatrix} a_{11}=10 & a_{12}=3 & a_{13}=17 & a_{14}=3.5 & a_{15}=1.7 & a_{16}=8 \\ a_{21}=0.05 & a_{22}=10 & a_{23}=17 & a_{24}=0.1 & a_{25}=8 & a_{16}=14 \\ a_{31}=3 & a_{32}=3.5 & a_{33}=1.7 & a_{34}=10 & a_{35}=17 & a_{36}=8 \\ a_{41}=17 & a_{42}=8 & a_{43}=0.05 & a_{44}=10 & a_{45}=0.1 & a_{46}=14 \end{pmatrix},$$

$$\begin{pmatrix} p_{11}=0.1312 & p_{12}=0.1696 & p_{13}=0.5569 & p_{14}=0.0124 & p_{15}=0.8283 & p_{16}=0.5886 \\ p_{21}=0.2329 & p_{22}=0.4135 & p_{23}=0.8307 & p_{24}=0.3736 & p_{25}=0.1004 & p_{16}=0.9991 \\ p_{31}=0.2348 & p_{32}=0.1451 & p_{33}=0.3522 & p_{34}=0.2883 & p_{35}=0.3047 & p_{36}=0.6650 \\ p_{41}=0.4047 & p_{42}=0.8828 & p_{43}=0.8732 & p_{44}=0.5743 & p_{45}=0.1091 & p_{46}=0.0381 \end{pmatrix}.$$

The global optimum $f^* = 3.32$ was determined numerically, and an estimated upper bound of the Lipschitz constant is $K \leq 1320.52$. The stopping rule is: either reach maximal number of function evaluations (1,500) or find a function value f such that $(f^* - f)/f* <= 0.01$.

F.2 Test problem 2: six-dimensional Hartmann problem over a discrete domain (modified directly from the continuous problem)

$$f(x) = \sum_{i=1}^{4} c_i \exp\left(-\sum_{j=1}^{6} a_{ij}\left(\frac{1}{10j}x_j - p_{ij}\right)^2\right),$$

$$\text{s.t } 0 \leq x_j \leq 10j \quad \text{for } j = 1, \ldots, 6,$$

where c_i, a_{ij}, and p_{ij} are the same as in the continuous problem. The global optimum $f^* = 0.166$ was determined numerically, and an estimated upper bound of the Lipschitz constant is $\tilde{K} \leq 1320.52$. The stopping rule is: either reach the maximal number of function evaluations (30,000) or find f^*.

F.3 Test problems 3 and 4: ten-dimensional Sinusoidal function over continuous and discrete domains [24]

$$\text{Minimize } f(x) = -2.5\prod_{i=1}^{10}\sin(x_i) - \prod_{i=1}^{10}\sin(5(x_i))$$

$$\begin{aligned} &\text{s.t (over a continuous domain)} \quad 0 \leq x_i \leq 180 \quad \text{for } i = 1, \ldots, 10, \\ &\text{or s.t (over a discrete domain)} \quad 0 \leq x_i \leq 180 \quad \text{and } x_i \text{ integer valued} \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{for } i = 1, \ldots, 10. \end{aligned}$$

The global optimum is $f^* = 3.5$ for both the continuous and discrete problems, an estimated upper bound of the Lipschitz constant for the continuous problem is $K \leq 23.72$, and an estimated upper bound of the Lipschitz constant for the discrete problem is $\tilde{K} \leq 7$. The stopping rule for the continuous problem is: either reach the maximal number of function evaluations (50,000) or find a function value f such that $(f^* - f)/f^* \leq 0.01$. The stopping rule for the discrete problem is: either reach the maximal number of function evaluations (50,000) or find f^*.

References

1. Aarts, E.H.L., Van Laarhoven, P.J.M.: Simulated annealing: an introduction. Statistica Neerlandica **43**, 31–52 (1989)
2. Ali, M.M., Khompatraporn, C., Zabinsky, Z.B.: A numerical evaluation of several global optimization algorithms on selected benchmark test problems. J. Glob. Optim. **31**(4), 635–672 (2005)
3. Alrefaei, M.H., Andradóttir, S.: A simulated annealing algorithm with constant temperature for discrete stochastic optimization. Manag. Sci. **45**(5), 748–764 (1999)
4. Bélisle, C.J.P.: Convergence theorems for a class of simulated annealing algorithms on R^n. J. Appl. Probab. **29**, 885–895 (1992)
5. Bohachevsky, I.O., Johnson, M.E., Stein, M.L.: Generalized simulated annealing for function optimization. Technometrics **28**, 209–217 (1986)
6. Cohn, H., Fielding, M.: Simulated annealing: searching for an optimal temperature schedule. SIAM J. Optim. **9**(3), 779–802 (1999)
7. Corana, A., Marchesi, M., Martini, C., Ridella, S.: Minimizing multimodal functions of continuous variables with the "simulated annealing" algorithm. ACM T. Math. Software **13**, 262–280 (1987)
8. Dekker, A., Aarts, E.H.L.: Global optimization and simulated annealing. Math. Program. **50**, 367–393 (1991)
9. Fielding, M.: Simulated annealing with an optimal fixed temperature. SIAM J. Optim. **11**(2), 289–307 (2000)
10. Hajek, B.: Cooling schedules for optimal annealing. Math. Oper. Res. **13**, 311–329 (1988)
11. Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P.: Optimization by simulated annealing. Science **220**, 671–680 (1983)
12. Locatelli, M.: Simulated annealing algorithms for continuous global optimization: convergence conditions. J. Optim. Theory Appl. **104**, 121–133 (2000)
13. Locatelli, M.: Convergence of a simulated annealing algorithm for continuous global optimization. J. Glob. Optim. **18**, 219–234 (2000)
14. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E.: Equations of state calculations by fast computing machines. J. Chem. Phys. **21**, 1087–1092 (1953)
15. Pincus, M.: A closed form solution for certain programming problems. Oper. Res. **16**, 690–694 (1968)
16. Pintér, J.D.: Global Optimization in Action (Continuous and Lipschitz Optimization: Algorithms, Implementations and Applications). Kluwer Academic Publishers, Dordrecht/Boston/London (1996)
17. Rubinstein, R.Y.: Simulation and the Monte Carlo Method. Wiley, New York (1981)
18. Romeijn, H.E., Smith, R.L.: Simulated annealing for constrained global optimization. J. Glob. Optim. **5**, 101–126 (1994)
19. Romeijn, H.E., Smith, R.L.: Simulated annealing and adaptive search in global optimization. Probab. Eng. Inform. Sci. **8**, 571–590 (1994)
20. Shen, Y., Zabinsky, Z.B., Smith, R.L.: Annealing adaptive search with a Markov chain Monte Carlo sampler for global optimization. Technical Report, Industrial Engineering, University of Washington, Seattle, WA (2005)
21. Shen, Y.: Annealing adaptive search with hit-and-run sampling methods for global optimization. *PhD dissertation, University of Washington* (2005)
22. Smith, R.L.: Efficient Monte Carlo procedures for generating points uniformly distributed over bounded region. Oper. Res. **32**, 1296–1308 (1984)
23. Zabinsky, Z.B.: Stochastic Adaptive Search for Global Optimization. Kluwer Academic Publishers, Boston/Dordrecht/London (2003)
24. Zabinsky, Z.B., Graesser, D.L., Tuttle, M.E., Kim, G.I.: Global Optimization of Composite Laminate Using Improving Hit-and-Run. In: Floudas, C.A., Pardalos, P.M. (eds.) Recent Advances in Global Optimization. Princeton University Press, Princeton, NJ (1992)
25. Zabinsky, Z.B., Smith, R.L.: Pure adaptive search in global optimization. Math. Program. **53**, 323–338 (1992)