Journal of Global Optimization (2023) 87:31–55
https://doi.org/10.1007/s10898-023-01307-7

Hesitant adaptive search with estimation and quantile adaptive search for global optimization with noise

Zelda B. Zabinsky[1] · **David D. Linz**[1]

Received: 11 May 2022 / Accepted: 13 June 2023 / Published online: 30 June 2023

Abstract
Adaptive random search approaches have been shown to be effective for global optimization problems, where under certain conditions, the expected performance time increases only linearly with dimension. However, previous analyses assume that the objective function can be observed directly. We consider the case where the objective function must be estimated, often using a noisy function, as in simulation. We present a finite-time analysis of algorithm performance that combines estimation with a sampling distribution. We present a framework called Hesitant Adaptive Search with Estimation, and derive an upper bound on function evaluations that is cubic in dimension, under certain conditions. We extend the framework to Quantile Adaptive Search with Estimation, which focuses sampling points from a series of nested quantile level sets. The analyses suggest that computational effort is better expended on sampling improving points than refining estimates of objective function values during the progress of an adaptive search algorithm.

1 Introduction

Adaptive random search algorithms for global optimization are often characterized by how they generate points within a feasible region, i.e., their sampling distribution. Many algorithms attempt to iteratively improve the sampling distribution to focus on promising regions, based on observations [3, 5, 8, 16–20, 22, 27]. An additional complication is the need to estimate the objective function value as the algorithm progresses.

This paper provides a finite-time analysis of a class of adaptive random search algorithms applied to problems that require *estimation* to account for noise, that is, problems where the objective function cannot be evaluated directly but must be estimated. We describe an extension to Hesitant Adaptive Search, we call Hesitant Adaptive Search with Estimation (HAS-E), and embed a confidence interval on the estimate of the current objective function value into the algorithm. The analysis relates the number of replications used in the estimation of the objective function to the overall performance of the algorithm.

✉ Zelda B. Zabinsky
zelda@uw.edu

[1] Department of Industrial and Systems Engineering, University of Washington, Seattle, WA 98191-2650, USA

In contrast to asymptotic convergence, a finite-time analysis of performance is available for a class of adaptive random search algorithms when the objective function can be evaluated directly through an oracle (i.e., a black-box function). Specifically, prior research has derived a finite-time analysis for Pure Adaptive Search (PAS), Hesitant Adaptive Search (HAS), Backtracking Adaptive Search (BAS), and Annealing Adaptive Search (AAS) [2, 4, 21, 23, 25–28, 30, 31].

These algorithms provide a framework for analysis, and are not intended to be implemented directly. However, the analyses shed light on the role of the sampling distribution and probability of generating improving points on performance. Under certain conditions, the expected number of function evaluations required to sample below a specified objective function value increases only linearly in dimension when optimizing a function without noise. We address the question of how estimation of a noisy function impacts the performance.

We first analyze the performance of HAS-E, where the sampling distribution focuses on nested level sets while allowing for 'hesitation.' We also introduce a new adaptive random search algorithm, called Quantile Adaptive Search with Estimation (QAS-E), which samples over the entire domain but parametrically modifies the sampling distribution based on a sequence of quantiles. The motivation for the analysis of QAS-E is to provide finite-time analyses that can be adapted for use in adaptive random search algorithms that use quantiles in the adaptive mechanics [6, 7, 9, 29].

The main result of this paper is in Theorem 3, which provides an upper bound on the expected number of function evaluations (including replications) required to first obtain a value within a target ϵ of the global minimum. This is used to show, in Corollary 1, that under certain conditions the expected number of function evaluations (including replications) to obtain a value less than ϵ above the minimum is bounded by a cubic function of the domain dimension. We then use the analysis of HAS-E to derive analogous bounds for QAS-E, which resembles quantile-based algorithms in practice.

2 Preliminaries

Consider an optimization problem,

$$\min_{x \in S} f(x) \tag{P}$$

where S is a closed and bounded subset of $\mathbb{R}^n$, $x \in S \subset \mathbb{R}^n$, and $f : \mathbb{R}^n \to \mathbb{R}$. We assume a unique minimum, and denote the minimum value and the optimal point in the domain, respectively, as:

$$y_* = \min_{x \in S} f(x) \text{ and } x_* = \arg\min_{x \in S} f(x). \tag{1}$$

Similarly, denote the maximum value and a maximal point as:

$$y^* = \max_{x \in S} f(x) \text{ and } x^* = \arg\max_{x \in S} f(x). \tag{2}$$

We only use y^* in the analysis, and any value associated with an upper confidence bound on y^* may be used.

Furthermore, we define the diameter d of S as the greatest distance between any two points in S.

We are particularly interested in the situation where the objective function cannot be evaluated directly but is expressed as

$$f(x) = E[g(x, \chi)] \tag{3}$$

where $g(x, \chi)$ is a "noisy" function of $x \in S$ and a random variable χ. Often, $g(x, \chi)$ is evaluated using a discrete-event simulation and is replicated a number of times at each point x, taking the sample mean as an estimate of $f(x)$ [1, 13]. In this paper, we seek to relate the number of replications on an iteration to the overall performance of an adaptive random search algorithm.

The finite-time analysis of Pure Adaptive Search establishes that, under certain conditions, the expected number of iterations (i.e., function evaluations) until PAS achieves a value close to the minimum increases only linearly in terms of the dimension of the domain [27, 30, 31]. However, the requirement that PAS improves at every iteration makes it difficult to implement practically. Hesitant Adaptive Search generalizes PAS by relaxing this requirement and allowing hesitation, thereby extending the class of algorithms it represents [4, 26]. We summarize HAS and a key result.

HAS is defined by a sampling distribution ζ with support on S and a bettering probability $b(y)$, $0 < b(y) \leq 1$, defined for $y_* < y \leq y^*$. On any iteration with objective function value y, HAS generates an improving point with probability $b(y)$ by drawing from the normalized restriction of ζ on the improving level set. The probability of hesitation is $1 - b(y)$, where the current point does not change. HAS is defined as follows.

Hesitant Adaptive Search (HAS), cf. [4]

- *Step 0* Sample X_0 in S according to the probability distribution ζ on S. Set $\bar{Y}_0 = f(X_0)$. Set $k = 0$.
- *Step 1* Generate X_{k+1} from the normalized restriction of ζ on the improving set $S_k = \{x \in S : f(x) < \bar{Y}_k\}$ with probability $b(\bar{Y}_k)$, and set $\bar{Y}_{k+1} = f(X_{k+1})$. Otherwise, set $X_{k+1} = X_k$ and $\bar{Y}_{k+1} = \bar{Y}_k$.
- *Step 2* If a stopping criterion is met, stop. Otherwise, increment k and return to Step 1.

Note that $\bar{Y}_k$ is non-increasing, and is decreasing on any iteration with probability $b(\bar{Y}_k)$.

The finite-time analysis of HAS in [4] provides a closed form expression for the expected number of iterations until reaching a specified $\epsilon > 0$ above the minimum function value, denoted $E[N(y_* + \epsilon)]$, as,

$$E[N(y_* + \epsilon)] = 1 + \int_{y_*+\epsilon}^{\infty} \frac{d\rho(t)}{b(t) \cdot p(t)} \tag{4}$$

where $\rho(y) = \zeta(f^{-1}([-\infty, y]))$, and $p(y) = \rho((-\infty, y])$. A complete characterization of HAS for problems with mixed continuous-integer variables is in [26].

The HAS analysis provides insight into the relationship between the bettering probability and performance. However, HAS is still difficult to implement because it is impractical to draw from the normalized restriction of ζ on the improving level set. Another way to define an adaptive random search algorithm is to always sample from the entire set S, and iteratively update a parameter controlling the sampling distribution.

Annealing Adaptive Search is an abstraction of simulation annealing, and it always samples from a Boltzmann distribution on the entire set S. The temperature parameter for the Bolzmann distribution is iteratively decreased to control the update of the sampling distribution. The analysis in [23, 24] establishes stochastic dominance between AAS and a special case of HAS, making use of the finite-time analysis of HAS. The analysis of AAS was used to derive an analytical cooling schedule for simulated annealing algorithms [24].

The analyses of PAS, HAS, and AAS provide insight into the performance of adaptive random search algorithms, however, they assume the objective function $f(x)$ can be evaluated exactly. As random search algorithms are being applied broadly to functions that require estimation, a major question is how estimation impacts performance.

3 Hesitant adaptive random search with estimation (HAS-E)

Given that the value of $f(x)$, for $x \in S$, cannot be directly observed, we consider estimating the value by performing a certain number of independent replications and taking the sample mean. Suppose $g(x, \chi_r)$ is evaluated at a point x for R replications, $r = 1, \ldots, R$. The sample mean estimate (dropping the x for notational convenience) is,

$$\hat{y}^{est} = \frac{\sum_{r=1}^{R} g(x, \chi_r)}{R}. \tag{5}$$

We assume that $\hat{y}^{est} \sim N(f(x), \frac{\sigma}{\sqrt{R}})$, where $\sigma^2 = Var(g(x, \chi))$, and σ is known. A standard probability bound is given by

$$P\left(f(x) - \frac{\sigma \cdot z_{\alpha/2}}{\sqrt{R}} \leq \hat{y}^{est} \leq f(x) + \frac{\sigma \cdot z_{\alpha/2}}{\sqrt{R}}\right) = 1 - \alpha \tag{6}$$

for $0 \leq \alpha \leq 1$ and where $z_{\alpha/2}$ is the standard normal value at $\alpha/2$.

We are interested in an upper bound of the estimate and let $\hat{y}^{high}$ be the upper confidence interval value, given by

$$\hat{y}^{high} = \hat{y}^{est} + \frac{\sigma \cdot z_{\alpha/2}}{\sqrt{R}}. \tag{7}$$

Since we want to know how far $\hat{y}^{high}$ is from the true value $f(x)$, we note that $\hat{y}^{high} \sim N\left(f(x) + \frac{\sigma \cdot z_{\alpha/2}}{\sqrt{R}}, \frac{\sigma}{\sqrt{R}}\right)$ and, also from (6), we have,

$$P\left(f(x) \leq \hat{y}^{high} \leq f(x) + 2\frac{\sigma \cdot z_{\alpha/2}}{\sqrt{R}}\right) = 1 - \alpha. \tag{8}$$

HAS-E uses the estimate $\hat{y}^{high}$ to focus sampling on regions that are likely to be improving. In contrast to HAS that samples in the improving level set with a bettering probability, HAS-E samples in the level set associated with $\hat{y}^{high}$ with a bettering probability.

On the kth iteration of HAS-E, the sampled point $x_k \in S$ is evaluated with R_k independent replications of $g(x_k, \chi_r)$ for $r = 1, \ldots, R_k$, and then the estimate $\hat{y}_k^{est}$ is calculated as in (5) and the upper bound $\hat{y}_k^{high}$ as in (7). We let $y_k = f(x_k)$ be the true objective function value at x_k, and the improving level set S_{y_k} be

$$S_{y_k} = \{x \in S : f(x) < y_k\}. \tag{9}$$

Similarly, we let

$$S_{\hat{y}_k^{high}} = \{x \in S : f(x) < \hat{y}_k^{high}\} \tag{10}$$

be the level set associated with the upper confidence interval bound and note that

$$P\left(S_{y_k} \subset S_{\hat{y}_k^{high}}\right) \geq 1 - \alpha$$

from (8).

We are interested in sampling a point in a target level set $S_{y_*+\epsilon}$ for some $\epsilon > 0$. The general approach of HAS-E is to sample from the normalized restriction of ζ on $S_{\bar{y}_k^{high}}$ with bettering probability $b\left(\bar{y}_k^{high}\right)$, and hesitate (remain at the same point) with probability $1 - b\left(\bar{y}_k^{high}\right)$, where $\bar{y}_k^{high}$ is the estimated upper confidence bound on the kth iteration. Later, for analysis purposes, we let γ be a minimum bettering probability such that $b(y) \geq \gamma$ for $y_* + \epsilon \leq y \leq y^*$. See Fig. 1 for an illustration of three iterations of HAS-E.

HAS-E requires input parameters α and σ, along with a sampling distribution ζ with support on the entire domain S, and the bettering probability $b(y)$. The input parameter α is used to determine how conservative the user wants to be, as in (8). HAS-E also requires a sequence of the number of replications on iteration k, i.e., $\{R_k, k = 0, 1, \ldots\}$, which is discussed later.

As with HAS, the HAS-E algorithm is a framework for analysis, and not intended to be implemented directly. However, the framework allows us to analyze the algorithm's performance on any iteration k.

Hesitant Adaptive Search with Estimation (HAS-E)

- *Step 0* Sample X_0 in S according to the probability distribution ζ on S. Conduct R_0 independent replications of the function at the initial selected point, i.e., $g(X_0, \chi_r)$ for $r = 1, \ldots, R_0$. Estimate the value $\hat{y}_0^{high}$ as in (7) and set $\bar{y}_0^{high} = \hat{y}_0^{high}$. Set $\bar{Y}_0 = f(X_0)$. Set $k = 0$.
- *Step 1* Generate X_{k+1} from the normalized restriction of ζ on the set $S_{\bar{y}_k^{high}}$ with bettering probability $b(\bar{y}_k^{high})$, and estimate $\hat{y}_{k+1}^{high}$ as in (7) with R_k independent replications of $g(X_{k+1}, \chi_r)$ for $r = 1, \ldots, R_k$. Otherwise (with probability $1-b(\bar{y}_k^{high})$), set $X_{k+1} = X_k$ and $\hat{y}_{k+1}^{high} = \hat{y}_k^{high}$. Then update

$$\bar{Y}_{k+1} = \begin{cases} f(X_{k+1}) & \text{if } f(X_{k+1}) < \bar{Y}_k \\ \bar{Y}_k & \text{otherwise} \end{cases}$$

  and its associated upper confidence bound estimate,

$$\bar{y}_{k+1}^{high} = \begin{cases} \hat{y}_{k+1}^{high} & \text{if } f(X_{k+1}) < \bar{Y}_k \\ \bar{y}_k^{high} & \text{otherwise.} \end{cases}$$

- *Step 2* If a stopping criterion is met, stop. Otherwise, increment k and return to Step 1.

Note that $\bar{Y}_k$ is non-increasing, however it is possible for $\bar{y}_k^{high}$ to increase and decrease.

The analysis begins with Theorem 1, characterizing the number of replications chosen on each iteration. Theorem 1 provides the number of replications needed to achieve the degree of confidence the user desires (α) given problem characteristics. We next prove in Theorem 2 that under certain assumptions about the replications, HAS-E stochastically dominates a special case of HAS without estimation. This allows us to provide (in Theorem 3) upper bounds on the expected number of HAS-E iterations and the expected number of function evaluations, including replications, to achieve an optimal solution with function value below a specified threshold. Although HAS-E is not directly implementable, the statement in Theorem 3 implies that, on average, the best observed point after the specified number of function evaluations, will have a true function value less than $y_* + \epsilon$. Finally, Corollary 1 provides bounds on the expected number of HAS-E iterations and expected number of function evaluations,

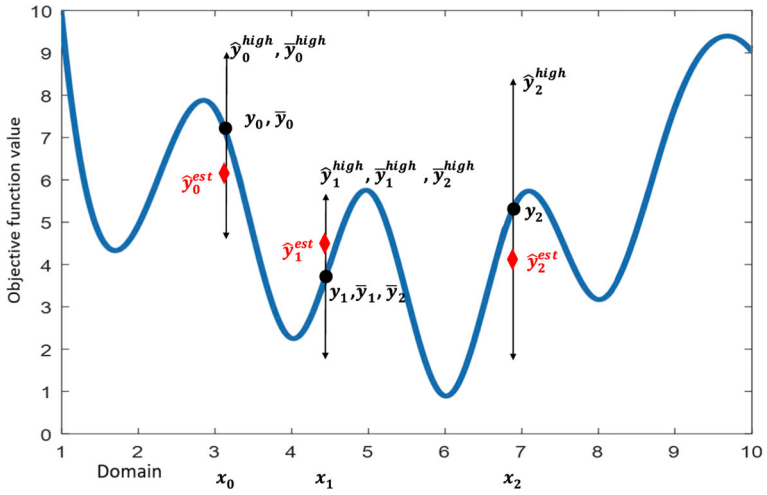


Fig. 1 An illustration of three HAS-E iterations ($k = 0, 1, 2$) on a one-dimensional problem. The sampled points are labeled x_0, x_1, and x_2, with true values y_0, y_1, and y_2. The estimated values, $\hat{y}_k^{est}$, and the upper confidence values, $\hat{y}_k^{high}$, are shown for $k = 0, 1, 2$. The best true values, $\bar{y}_0$, $\bar{y}_1$, and $\bar{y}_2$, and their associated upper confidence values $\bar{y}_0^{high}$, $\bar{y}_1^{high}$, and $\bar{y}_2^{high}$ are also illustrated

that, under certain assumptions, increase linearly in dimension and cubic in dimension, respectively.

For purposes of our analysis, we consider a lower bound on the ratio of volumes of the true level set to the level set associated with the upper confidence interval bound $\hat{y}_k^{high}$, i.e., $\nu(S_{y_k})/\nu(S_{\hat{y}_k^{high}})$, where $\nu(\cdot)$ is the n-dimensional volume of a set. A trivial lower bound for this quantity can be based on the desired accuracy value ϵ, with $\epsilon > 0$. For y_k and $\hat{y}_k^{high}$ such that $y_* + \epsilon < y_k \leq \hat{y}_k^{high} \leq y^*$, the following lower bound on the ratio

$$\frac{\nu(S_{y_k})}{\nu(S_{\hat{y}_k^{high}})} \geq \frac{\nu(S_{y_*+\epsilon})}{\nu(S)}$$

holds.

However, this trivial lower bound may be very small, so we consider another lower bound, and relate it to the number of replications used in the estimation. We let q denote a lower bound on the ratio, with $0 < q < 1$, such that

$$\frac{\nu(S_{y_k})}{\nu(S_{\hat{y}_k^{high}})} \geq q$$

for $y_* + \epsilon < y_k \leq \hat{y}_k^{high} \leq y^*$. Figure 2 illustrates the level sets S_{y_k} and $S_{\hat{y}_k^{high}}$. Notice that the ratio of the volumes $\nu(S_{y_k})/\nu\left(S_{\hat{y}_k^{high}}\right)$ becomes close to 1 as the distance between the values $\hat{y}_k^{high}$ and y_k decreases, which typically occurs as the number of replications increases. Theorem 1 provides a bound on the number of replications needed such that the ratio of volumes can be bounded below for a selected value q.

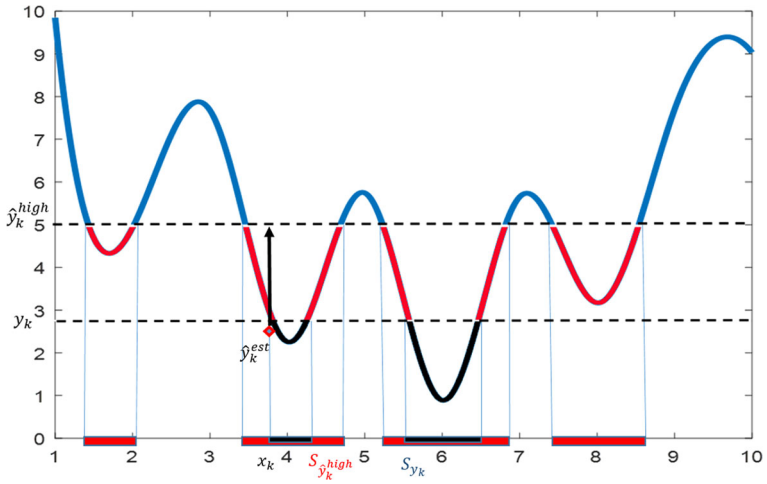


Fig. 2 An illustration of the values y_k and $\hat{y}_k^{high}$ along with their corresponding level sets S_{y_k} and $S_{\hat{y}_k^{high}}$ for a one-dimensional problem. The level sets are shown highlighted on the horizontal axis. The ratio between the volumes of the level sets, $\nu(S_{y_k})/\nu\left(S_{\hat{y}_k^{high}}\right)$, increases and approaches one as the difference between $\hat{y}_k^{high}$ and y_k decreases and $\hat{y}_k^{high}$ approaches y_k (which happens with a large number of replications)

For a function $f(x)$ on domain S, we define a quantity $\mathcal{K}_q$, for a given $0 < q < 1$, which can be viewed as the maximum ratio of the change in objective function to the diameter of S

$$\mathcal{K}_q = \frac{\kappa_q}{d} \tag{11}$$

where κ_q is the maximum value such that $\nu(S_z)/\nu(S_{z+\kappa_q}) > q$ for any z, $y_* < z < y^*$. The quantity $\mathcal{K}_q$ depends on characteristics of the problem.

Furthermore, we define $\mathcal{B}_y$ as the largest ball centered at x_* that can be inscribed inside a level set S_y for $y_* < y < y^*$, and let r_y be its radius. Using these two defined concepts, we now relate the number of replications to a selected q.

Theorem 1 *Consider problem (P) and the kth iteration of HAS-E with $x_k \in S$, $y_k = f(x_k)$, $y_* + \epsilon < y_k \leq y^*$ for $\epsilon > 0$, and with $\hat{y}_k^{high}$ estimated with R replications, as in (7). Also, suppose that $y_k \leq \hat{y}_k^{high} \leq y_k + \frac{2\cdot\sigma\cdot z_{\alpha/2}}{\sqrt{R}}$ (which occurs with probability $(1-\alpha)$). For any given value $0 < q < 1$ and associated $\mathcal{K}_q$, if*

$$R \geq \left(\frac{\sqrt[n]{q} \cdot 2 \cdot \sigma \cdot z_{\alpha/2}}{(1 - \sqrt[n]{q}) \cdot r_{y_*+\epsilon} \cdot \mathcal{K}_q} \right)^2 \tag{12}$$

then

$$\frac{\nu(S_{y_k})}{\nu(S_{\hat{y}_k^{high}})} \geq q. \tag{13}$$

Proof See the proof in Appendix A. □

Theorem 1 provides an expression for the number of replications needed to achieve a lower bound of q. When the value of q is close to one, indicating a relatively tight upper

confidence bound, more replications are needed than the number of replications when q has a small value. The difference becomes more pronounced as the dimension of the problem increases. Setting the number of replications equal to the expression in (12), we use Lemma 1 to show that the number of replications is on the order of n^2,

$$\left(\frac{\sqrt[n]{q}\cdot 2\cdot\sigma\cdot z_{\alpha/2}}{(1-\sqrt[n]{q})\cdot r_{y_*+\epsilon}\cdot\mathcal{K}_q}\right)^2 \le \left(\frac{q}{1-q}-(n-1)\frac{ln(q)}{(1-q)^2}\right)^2\left(\frac{2\cdot\sigma\cdot z_{\alpha/2}}{r_{y_*+\epsilon}\cdot\mathcal{K}_q}\right)^2. \quad (14)$$

This suggests that the number of replications should increase quadratically in dimension to achieve a tight estimate. We ask the question whether a tight estimate is worthwhile for overall performance.

We next develop an upper bound on the expected number of iterations of HAS-E to achieve a function value of $y_* + \epsilon$ or better.

The analysis of HAS-E proceeds in Theorem 2 by showing that HAS-E stochastically dominates a special case of the HAS algorithm, that we call HAS1. Then, using HAS1, Theorem 3 provides an upper bound on the expected number of iterations of HAS-E and expected number of function evaluations (including replications) to achieve $y_* + \epsilon$ or better.

The special case HAS1 has a uniform sampling distribution, i.e., $\zeta^{HAS1} \sim$ Uniform, and the bettering probability is chosen to be constant for all y,

$$b^{HAS1}(y) = \gamma\cdot(1-\alpha)\cdot q \quad (15)$$

where $0 < \gamma \le 1$, $0 < \alpha < 1$, and $0 < q < 1$.

Let $\bar{Y}_k^{HASE}$ be the best sampled value on the kth iteration of the HAS-E algorithm. Let $\bar{Y}_k^{HAS1}$ be the best sampled value of HAS1 on the kth iteration.

For the performance analysis of HAS-E in Theorems 2, 3, and Corollary 1, we make the following assumptions.

Assumption 1 (i) The sampling distribution ζ dominates the uniform distribution on S, that is,

$$P\left(\bar{Y}_0^{HASE}\le y\right) \ge P\left(\bar{Y}_0^{HAS1}\le y\right) \text{ for } y_* < y \le y^*.$$

(ii) The bettering probability in HAS-E is bounded below by a positive constant, that is, for some positive γ, $0 < \gamma \le 1$,

$$b(y) \ge \gamma \text{ for } y_* < y \le y^*.$$

Theorem 2 proves stochastic dominance of HAS-E over HAS1.

Theorem 2 *Given the conditions in Assumption 1 and setting $R_k = R$ for all k as,*

$$R = \left(\frac{\sqrt[n]{q}\cdot 2\cdot\sigma\cdot z_{\alpha/2}}{(1-\sqrt[n]{q})\cdot r_{y_*+\epsilon}\cdot\mathcal{K}_q}\right)^2 \quad (16)$$

then $\bar{Y}_k^{HASE}$ stochastically dominates $\bar{Y}_k^{HAS1}$, that is,

$$P(\bar{Y}_k^{HASE}\le y) \ge P(\bar{Y}_k^{HAS1}\le y) \text{ for } k = 0, 1, \ldots$$

where $y_* < y \le y^*$.

Proof The proof is provided in Appendix A. □

Since $\bar{Y}_k^{HASE}$ stochastically dominates $\bar{Y}_k^{HAS1}$, and the finite-time performance of HAS1 is captured in [4, 26], we can bound the finite-time behavior of HAS-E.

We are particularly interested in the expected behavior. We next derive an upper bound on the expected number of HAS-E iterations to achieve a sample point within a target level set $S_{y_*+\epsilon}$ for $\epsilon > 0$, denoted $E[N_I^{HASE}(y_* + \epsilon)]$. The proof relies on the stochastic dominance of HAS-E over HAS1 in Theorem 2, and uses an upper bound on HAS1 iterations, as in (4). Theorem 3 also expresses the expected number of function evaluations, including replications, needed to achieve a sample point within the target level set $S_{y_*+\epsilon}$, denoted $E[N_R^{HASE}(y_*+\epsilon)]$.

Theorem 3 *An upper bound on the expected number of HAS-E iterations until reaching a value of $y_* + \epsilon$ or better, for $\epsilon > 0$, is given by,*

$$E[N_I^{HASE}(y_* + \epsilon)] \leq 1 + \left(\frac{1}{\gamma \cdot (1-\alpha) \cdot q}\right) ln\left(\frac{\nu(S)}{\nu(S_{y_*+\epsilon})}\right) \tag{17}$$

and an upper bound on the expected number of HAS-E function evaluations including replications is

$$E[N_R^{HASE}(y_* + \epsilon)] \leq \left(\frac{\sqrt[n]{q} \cdot 2 \cdot \sigma \cdot z_{\alpha/2}}{(1 - \sqrt[n]{q}) \cdot r_{y_*+\epsilon} \cdot \mathcal{K}_q}\right)^2 E[N_I^{HASE}(y_* + \epsilon)]. \tag{18}$$

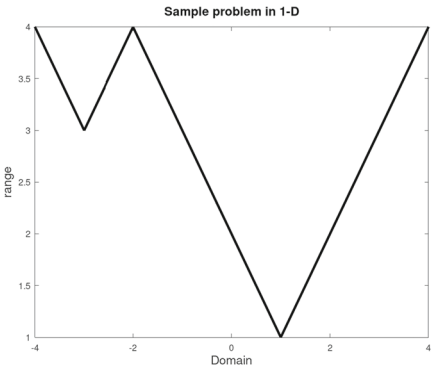
Proof See the proof in Appendix A. □

The expressions in (17) and (18) provide insight into the value of replications. Focusing on the impact of q in (17), we see that the expected number of iterations decreases as q increases, indicating some benefit to a large value of q with a relatively tight upper confidence bound. However, the expected number of replications in (18) indicates that the large number of replications needed for a large value of q overshadows the benefit. This can be interpreted as a tradeoff between sampling from a larger than needed level set (with loose upper confidence bound and fewer replications) and sampling from a more accurate estimate of the current level set (with tight upper confidence bound and more replications). The expected number of replications also reflects the desired tightness of the upper confidence bound, through $z_{\alpha/2}$. This leads us to consider algorithms that use few replications as long as the estimation approaches the true function value as the algorithm approaches the global minimum.

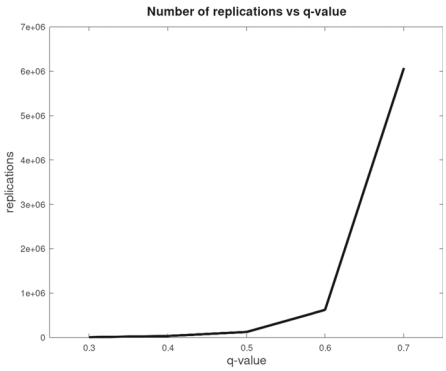
To illustrate these results, we present a one-dimensional sample problem, $f(x)$, in Fig. 3(a). Details are provided in Appendix C. We plot the number of replications R, expected number of iterations $E[N_I^{HASE}(y_*+\epsilon)]$, and expected number of function evaluations $E[N_R^{HASE}(y_* + \epsilon)]$ for a range of q values for the sample problem. Figure 3(b) illustrates that the number of replications increases with q, whereas Fig. 3(c) illustrates that the number of iterations decreases with q. Combined, the total number of function evaluations including replications increases with q, as shown in Fig. 3(d). This illustrates that fewer replications with a loose upper confidence bound is effective for overall performance.

We also explore the impact of dimension n on function evaluations in Corollary 1. The upper bound on the number of replications in (14) indicate that the number of replications is quadratic in dimension. The bound on expected number of iterations in (19) is linear in dimension, for problems satisfying the conditions in Corollary 1. Together, the expected number of total function evaluations including replications in Corollary 1 is cubic in dimension. An interpretation is that sampling on the level set associated with the upper confidence bound as opposed to the true level set increases the number of function evaluations quadratically in dimension as opposed to the number needed if the function were able to be evaluated exactly. Also, the dimension n magnifies the difference comparing values of q, reinforcing the intuition that fewer replications are better.

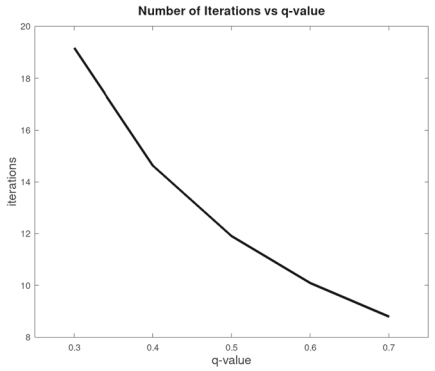
The following corollary couples this with a bound on $\nu(S)/\nu(S_{y_*+\epsilon})$ in terms of dimension n for a class of problems satisfying a Lipschitz constant.



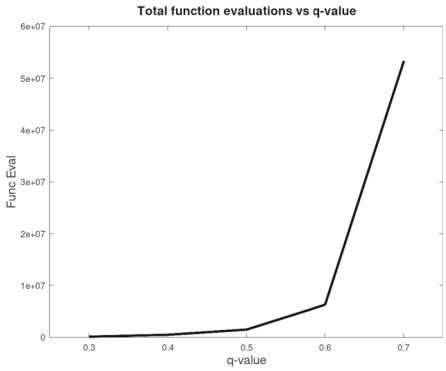
(a) One-dimensional sample problem.



(b) Replications R as in (16).



(c) Expected number of iterations, as in (17).



(d) Expected number of total function evaluations, as in (18).

Fig. 3 A one-dimensional sample problem $f(x)$, illustrating that the number of replications R, as in (16), increase as q increases, while the expected number of iterations $E[N_I^{HASE}(y_* + \epsilon)]$, as in (17), decrease as q increases. The bound on expected number of total function evaluations including replications, $E[N_R^{HASE}(y_* + \epsilon)]$, as in (18), increase as q increases

Corollary 1 *When S in (P) is a convex feasible region in n dimensions with a diameter d and $f(x)$ satisfies the Lipschitz condition with Lipschitz constant at most $\mathcal{L}$, then the expected number of iterations for HAS-E to reach a value $y_* + \epsilon$, $\epsilon > 0$, is bounded by,*

$$E[N_I^{HASE}(y_* + \epsilon)] \leq 1 + \left(\frac{n}{\gamma \cdot (1-\alpha) \cdot q}\right) ln\left(\frac{\mathcal{L} \cdot d}{\epsilon}\right) \tag{19}$$

and the expected number of function evaluations (including replications) to achieve a value of $y_* + \epsilon$ or better is upper-bounded by a cubic function of domain dimension,

$$\begin{aligned} &E\left[N_R^{HASE}(y_* + \epsilon)\right] \\ &\leq \left(\left(\frac{q}{1-q} - (n-1)\frac{ln(q)}{(1-q)^2}\right)\left(\frac{2 \cdot \sigma \cdot z_{\alpha/2}}{r_{y_*+\epsilon} \cdot \mathcal{K}_q}\right)\right)^2 \left(1 + \left(\frac{n}{\gamma \cdot (1-\alpha) \cdot q}\right) ln\left(\frac{\mathcal{L} \cdot d}{\epsilon}\right)\right) \\ &\sim O\left(n^3\right). \end{aligned}$$

Proof The expression in (17), combined with the bounds on $\left(\nu(S)/\nu(S_{y_*+\epsilon})\right)$ in [26, 27, 30], produce the linear number of iterations. Coupling this with the upper bound on replications that is quadratic in dimension, as in (14), provides an upper bound on total function evaluations that is cubic in dimension. □

This corollary provides a bound on the expected number of function evaluations including replications needed to obtain a value of $y_* + \epsilon$ or less that is a *polynomial* function of the dimension, holding all other parameters constant. This result generally extends the finite-time results of the HAS framework for problems with estimation. In the next section, we examine a framework based on an adaptive search framework that samples from a series of nested quantile level sets.

4 Quantile adaptive search with estimation (QAS-E)

We now define a Quantile Adaptive Search with Estimation (QAS-E) which conceptualizes an optimization algorithm that samples according to a probability distribution parameterized by quantile. QAS-E utilizes a series of sampling distributions defined by density function ζ_k associated with quantile δ_k on iteration k.

The motivation for incorporating a quantile as a parameter in the sampling distribution is to provide a finite-time analysis to aid in the development of algorithms that use quantiles in their adaptive mechanics [6, 7, 9, 22, 29]. QAS-E differs from HAS-E in its sampling distribution; instead of sampling X_{k+1} from the normalized restriction of ζ on $S_{\tilde{y}_k^{high}}$ as in HAS-E, QAS-E always samples from S however, the distribution ζ_k depends on a quantile parameter δ_k. The intuition is that it is relatively easy to sample a point in a level set associated with a high quantile, but it is challenging to sample a point from a level set associated with a low quantile. Instead of attempting to hit a target level set associated with a low quantile on the first iteration, QAS-E allows the quantile to be reduced iteratively, thereby modifying the parameterized sampling distribution. The hope is that small changes in the reduction of the quantile parameter will aid in implementation.

We draw an analogy to the use of a temperature parameter in the Boltzmann distribution. When the temperature is high, it is relatively easy to sample from the Boltzmann distribution, but when the temperature is low, it is difficult to efficiently generate a sample point. The idea is that it is computationally easier to approximate a Boltzmann distribution with a small change in temperature, gradually reducing the temperature. The analysis of Annealing Adaptive Search provided insight that led to the development of an adaptive cooling schedule for simulated annealing [23, 24].

The following analysis of QAS-E provides insight into the computational potential for algorithms that focus on sampling from level sets with quantile estimators. A general challenge with implementation is selecting the δ_k-quantile values and associated sampling distributions ζ_k for which an adaptive algorithm has desirable performance. We parameterize the sampling distribution by a quantile value, denoted $\zeta_k(\delta_k)$. This is analogous to the way the Boltzmann distribution is parameterized by temperature. In the following analysis, the Boltzmann distribution is a possible family for QAS-E.

There is a relationship between a quantile value δ and the associated objective function value y_δ. For a quantile value, $0 < \delta < 1$, let the associated level set be denoted

$$S_\delta = \{x \in S : f(x) < y_\delta\} \tag{20}$$

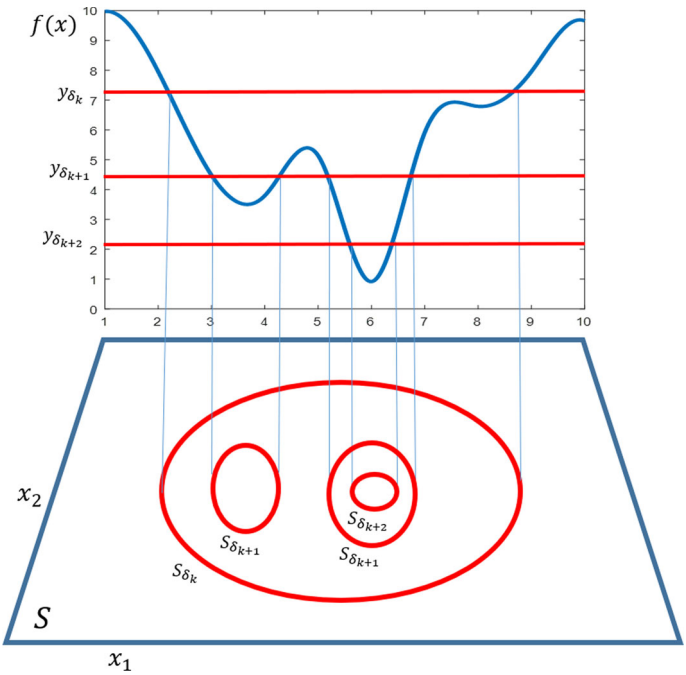


Fig. 4 An illustration of three nested quantile level sets with $\delta_k > \delta_{k+1} > \delta_{k+2}$. Quantile Adaptive Search seeks to sample from the level set S_{δ_k} on the kth iteration

where y_δ is the δ-quantile of the domain S, or explicitly,

$$y_\delta = \underset{y_* < y \le y^*}{\arg\min}\ P(f(X) \le y) \ge \delta \tag{21}$$

when X is uniformly sampled on S.
When sampling according to the probability distribution $\zeta_k(\delta_k)$ on S, we relate the probability of landing inside of a level set S_δ to the probability of achieving an objective function value of y_δ or better through an integral, as

$$P\left(Y_k \le y_\delta\right) = P(X_k \in S_\delta) = \int_{S_\delta} \zeta_k(\delta_k)(x) \cdot dx$$

where X_k is drawn from $\zeta_k(\delta_k)$ and $Y_k = f(X_k)$.

To illustrate the general form of QAS-E, see Fig. 4 with three level sets associated with decreasing quantile levels $\delta_{k+2} < \delta_{k+1} < \delta_k$ so that $S_{\delta_{k+2}} \subset S_{\delta_{k+1}} \subset S_{\delta_k}$. The sampling distribution $\zeta_k(\delta_k)$ is chosen to maintain some minimum probability of sampling within the associated level set S_{δ_k} at each iteration k. Therefore, the iterative selection of a quantile δ_k can be seen as a mechanism for focusing the sampling distribution on nested quantile level sets.

QAS-E requires input parameters α and σ, as in HAS-E, and a sequence of parameterized sampling distributions $\{\zeta_k(\delta_k), k = 0, 1, \ldots\}$. It also requires a sequence of the number of replications on iteration k, i.e., $\{R_k, k = 0, 1, \ldots\}$. The analysis of QAS-E requires several conditions on the sampling distributions stated in Assumption 2, and discussed later. We

formally write the algorithm based on the selection of parameterized sampling distributions $\zeta_k(\delta_k)$.

Quantile Adaptive Search with Estimation (QAS-E)

- *Step 0* Sample X_0 in S according to the probability distribution $\zeta_0(\delta_0)$ on S. Conduct R_0 independent replications of the function at the initial selected point, i.e., $g(X_0, \chi_r)$ for $r = 1, \ldots, R_0$. Estimate the value $\hat{y}_0^{high}$ as in (7) and set $\bar{y}_0^{high} = \hat{y}_0^{high}$. Set $\bar{Y}_0 = f(X_0)$. Set $k = 0$.
- *Step 1* Update the parameter quantile δ_{k+1} and its sampling probability distribution $\zeta_{k+1}(\delta_{k+1})$. Generate X_{k+1} from the probability distribution $\zeta_{k+1}(\delta_{k+1})$ on S. Perform R_k independent replications of $g(X_{k+1}, \chi_r)$ for $r = 1, \ldots, R_k$ (if $X_{k+1} \neq X_k$) and estimate $\hat{y}_{k+1}^{high}$ as in (7). Then update

$$\bar{Y}_{k+1} = \begin{cases} f(X_{k+1}) & \text{if } f(X_{k+1}) < \bar{Y}_k \\ \bar{Y}_k & \text{otherwise} \end{cases}$$

  and its associated upper confidence bound estimate,

$$\bar{y}_{k+1}^{high} = \begin{cases} \hat{y}_{k+1}^{high} & \text{if } f(X_{k+1}) < \bar{Y}_k \\ \bar{y}_k^{high} & \text{otherwise.} \end{cases}$$

- *Step 2* If a stopping criterion is met, stop. Otherwise, increment k and return to Step 1.

Note that when there is no noise in the objective function, then no replications are needed and $\hat{y}_k^{high}$ can be replaced with the true function value y_k in the algorithm.

The QAS-E algorithm iteratively samples from a sequence of distributions parameterized by a quantile value. The intent is for the distributions to increase the chances of generating improving sets, much in the same way that the Boltzmann distribution with a temperature parameter increases its focus on improving level sets. At each iteration, the upper confidence bound estimate $\bar{y}_k^{high}$ has an associated quantile value (denoted $\bar{\delta}_k^{high}$) through (20) and (21), that may help inform the choice of quantile parameter. The sampling distribution $\zeta_k(\delta_k)$ is parameterized by quantile value to aid in adaptively varying δ_k. Exploratory numerical results in [14] adapted δ_k ad hoc, with implicit consequences on the sampling distribution. The numerical results suggest that adapting δ_k improves performance, however there was previously no theory to guide the adaptation to achieve desired performance. The relationship between quantile as a parameter of the sampling distribution in QAS-E may guide the implementation of adaptive random search methods in an analogous way that AAS aided in developing a cooling schedule for simulated annealing.

For the performance analysis of QAS-E, we make the following assumptions regarding the sampling distribution with quantile parameter, $\zeta_k(\delta_k)$ on iteration k. The conditions in Assumption 2 ensure that each sampling distribution does no worse than the previous one at generating improving points. Assumptions 2(*i*)-(*ii*) are similar to Assumptions 1(*i*)-(*ii*) for HAS-E.

Assumption 2 (i) *The sampling distribution $\zeta_k(\delta_k)$ dominates the uniform distribution, that is,*

$$P\left(\bar{Y}_k^{QASE} \leq y\right) \geq \left(\bar{Y}_0^{HAS2} \leq y\right) \tag{22}$$

for any iteration k and $y_* < y \leq y^*$. This requirement forces each sampling distribution to be more focused on improvement than the uniform distribution. In effect this excludes

distributions that are not able to sample from nested level sets better than uniform, perhaps due to local behavior.

(ii) The probability of improving on the current upper bound estimate $\bar{y}_k^{high}$ when sampling from the probability distribution $\zeta_{k+1}(\delta_{k+1})$ is bounded below by some minimum probability γ,

$$P\left(\bar{Y}_{k+1} \leq \bar{y}_k^{high} | \bar{Y}_k = \bar{y}_k\right) \geq \gamma \tag{23}$$

where $0 < \gamma \leq 1$. *We require that the sampling distribution has a minimum probability of improvement. This requirement forces each updated sampling distribution to maintain some probability of sampling within the improving quantile level set associated with the upper bound estimate.*

(iii) The conditional probability that the distribution $\zeta_{k+1}(\delta_{k+1})$ samples within a lower level set given that the previous sampled value was y_k, is non-increasing in $\bar{y}_k$ for all k. This condition can be written as

$$P\left(\bar{Y}_{k+1} \leq y | \bar{Y}_k = \bar{y}_k\right) \geq P\left(\bar{Y}_{k+1} \leq y | \bar{Y}_k = \bar{y}_k'\right) \tag{24}$$

where $\bar{y}_k < \bar{y}_k'$. The sampling distributions cannot perform worse having observed a better point, e.g., $\bar{y}_k < \bar{y}_k'$. This prevents the sampling distribution from getting "stuck" at local minima by arriving at some small value that makes the sampling of further improvement almost impossible.

We now present an analysis of the performance of QAS-E that parallels that of HAS-E. First, in Theorem 4, we show that the iterates of QAS-E stochastically dominate those of a special case of the standard HAS algorithm (called HAS2). Then, we use the special case HAS2 in Theorem 5 to provide an upper bound on the expected number of QAS-E iterations and expected number of function evaluations including replications to achieve an optimal point with a function value within ϵ of the optimal value y_*.

The special case HAS2 uses uniform sampling, i.e., $\zeta^{HAS2} \sim$ Uniform, and the bettering probability is chosen to be constant for all y,

$$b^{HAS2}(y) = \gamma \cdot (1 - \alpha) \cdot q \tag{25}$$

where $0 < \gamma \leq 1$, $0 < \alpha < 1$, and $0 < q < 1$.

Let $\bar{Y}_k^{QASE}$ be the best sampled value by QAS-E on the kth iteration, and let $\bar{Y}_k^{HAS2}$ be the best sampled value on the kth iteration of HAS2. We show that QAS-E stochastically dominates HAS2 in Theorem 4.

Theorem 4 *Given the three conditions in Assumption 2 and setting $R_k = R$ for all k as in (16), then $\bar{Y}_k^{QASE}$ stochastically dominates $\bar{Y}_k^{HAS2}$, that is:*

$$P(\bar{Y}_k^{QASE} \leq y) \geq P(\bar{Y}_k^{HAS2} \leq y) \text{ for } k = 0, 1, \ldots,$$

where $y_* < y \leq y^*$.

Proof The proof is similar to the proof of Theorem 2, but is provided in Appendix B for completeness. □

Theorem 5 provides upper bounds on the expected number of QAS-E iterations and expected number of function evaluations including replications to achieve a point within a target level set $S_{y_*+\epsilon}$. Notice the bounds in Theorem 5 are the same as in Theorem 3, suggesting the importance of an effective sampling distribution.

Theorem 5 *An upper bound on the expected number of QAS-E iterations until the value of $y_* + \epsilon$ or better is sampled, for $\epsilon > 0$, is given by,*

$$E[N_I^{QASE}(y_* + \epsilon)] \leq 1 + \left(\frac{1}{\gamma \cdot (1-\alpha) \cdot q}\right) ln\left(\frac{\nu(S)}{\nu(S_{y_*+\epsilon})}\right). \tag{26}$$

and an upper bound on the expected number of QAS-E function evaluations including replications is

$$E[N_R^{QASE}(y_* + \epsilon)] \leq \left(\frac{\sqrt[n]{q} \cdot 2 \cdot \sigma \cdot z_{\alpha/2}}{(1 - \sqrt[n]{q}) \cdot r_{y_*+\epsilon} \cdot \mathcal{K}_q}\right)^2 E[N_I^{QASE}(y_* + \epsilon)]. \tag{27}$$

Proof The proof is similar to that of Theorem 3. □

The final corollary is analogous to Corollary 1, and states that, when the problem (P) satisfies certain conditions, the expected number of QAS-E iterations is bounded by a linear function in dimension, and the expected number of QAS-E function evaluations including replications is cubic in dimension.

Corollary 2 *When S in (P) is a convex feasible region in n dimensions with a diameter d and $f(x)$ satisfies the Lipschitz condition with Lipschitz constant at most $\mathcal{L}$, then the expected number of iterations for QAS-E to reach a value $y_* + \epsilon$, $\epsilon > 0$, is bounded by,*

$$E[N_I^{QASE}(y_* + \epsilon)] \leq 1 + \left(\frac{n}{\gamma \cdot (1-\alpha) \cdot q}\right) ln\left(\frac{\mathcal{L} \cdot d}{\epsilon}\right) \tag{28}$$

and the expected number of function evaluations (including replications) to achieve a value of $y_* + \epsilon$ or better is bounded by a cubic function of domain dimension,

$$\begin{aligned} &E\left[N_R^{QASE}(y_* + \epsilon)\right] \\ &\leq \left(\left(\frac{q}{1-q} - (n-1)\frac{ln(q)}{(1-q)^2}\right)\left(\frac{2 \cdot \sigma \cdot z_{\alpha/2}}{r_{y_*+\epsilon} \cdot \mathcal{K}_q}\right)\right)^2 \left(1 + \left(\frac{n}{\gamma \cdot (1-\alpha) \cdot q}\right) ln\left(\frac{\mathcal{L} \cdot d}{\epsilon}\right)\right) \\ &\sim O\left(n^3\right). \end{aligned}$$

Proof The proof is similar to that of Corollary 1. □

The analysis of QAS-E parallels that for HAS-E, and highlights the result relating the estimation with a confidence bound and the performance related to the sampling distribution. By making assumptions on the consistency of a sequence of sampling distributions, it is clear that there is flexibility in choosing a parameterized sampling distribution, however, the assumptions must be satisfied. In this paper we emphasize using quantiles as parameters, however, the Boltzmann distribution parameterized by temperature satisfies the assumptions too. Thus, a version of AAS with estimation is captured in the analytical results.

5 Discussion and conclusion

We provide a framework for modeling adaptive random search for problems that require estimation of an objective function. Hesitant Adaptive Search with Estimation has a provable finite-time bound on the expected number of function evaluations until a specified ϵ above the minimum value is reached. Under certain conditions, the expected number of function evaluations including replications is bounded by a cubic function of the domain's dimension.

Furthermore, we introduce an additional adaptive search algorithm, Quantile Adaptive Search with Estimation, that extends HAS-E to a search that adapts the sampling distribution on the domain S based on the quantiles of an objective function. A difference between HAS-E and QAS-E is that HAS-E samples according to a normalized distribution restricted to nested level sets defined by the estimated function values, whereas QAS-E samples on the entire feasible region but parameterizes the sampling distribution to focus on nested level sets defined by quantiles. QAS-E has similar finite-time results as HAS-E controlling the number of replications and iterations.

Using the level of the quantile as a parameter, QAS-E's parameterized sampling distribution is analogous to use of the Boltzmann distribution with a temperature parameter in Annealing Adaptive Search. The analysis of QAS-E can be used to add estimation to AAS for stochastic problems.

An insight that the analysis of HAS-E and QAS-E provides is that the value of consistent improvement in the sampling distribution is more important than the number of replications needed to achieve a close estimate of the objective function at points evaluated during the process. The bounds on expected function evaluations lead us to consider algorithms that use a few replications at the expense of sampling from a larger than needed level set. However, it is still important that the algorithm converges to the true global minimum. Future research will consider the Single Observation Stochastic Algorithm (SOSA) [11, 12, 15] to combine estimation using a shrinking ball with the sampling distribution.

Another avenue for future research would be to use the analysis of QAS-E to develop a means for adaptively setting quantile parameters in a sampling distribution, as was done using AAS to derive an analytical cooling schedule for simulated annealing. Our analysis may help to improve algorithms like Probabilistic Branch and Bound [29], Cross Entropy [22], or reinforcement learning [9] which either explicitly or implicitly attempt to sample from within quantiles of an objective function.

Acknowledgements This research has been supported in part by the National Science Foundation, Grant CMMI-1935403.

Data Availibility Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Appendix A Proofs of Theorems for HAS-E Analysis

A.1 Proof of Theorem 1

Proof of Theorem 1 For any value y_k such that $y_* + \epsilon < y_k \leq y^*$, we start by defining an n-ball $\mathcal{B}_{y_k}$ as the largest n-ball centered at x_* such that $\mathcal{B}_{y_k} \subseteq S_{y_k}$ and let r_{y_k} be its radius. We note that $0 < \nu(\mathcal{B}_{y_k}) \leq \nu(S_{y_k})$. For any value $\hat{y}_k^{high}$, we define $\mathcal{B}_{\hat{y}_k^{high}}$ as the smallest n-ball centered at x_* such that $S_{\hat{y}_k^{high}} \subseteq \mathcal{B}_{\hat{y}_k^{high}}$ and let $r_{\hat{y}_k^{high}}$ be the radius of $\mathcal{B}_{\hat{y}_k^{high}}$.

We examine two cases. First, if $\hat{y}_k^{high} - y_k \leq \kappa_q$ then $\frac{\nu(S_{y_k})}{\nu(S_{\hat{y}_k^{high}})} > q$ by definition in (11), and the theorem is proved.

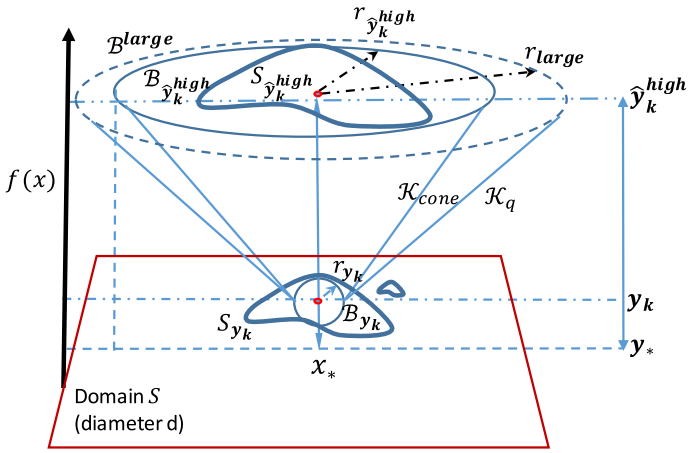


Fig. 5 An illustration of the largest n-ball inscribed in S_{y_k} ($\mathcal{B}_{y_k}$), the smallest n-ball inscribing $S_{\hat{y}_k^{high}}$ ($\mathcal{B}_{\hat{y}_k^{high}}$), and a larger ball defined by the slope $\mathcal{K}_q$ ($\mathcal{B}^{large}$)

Second, consider $\hat{y}_k^{high} - y_k > \kappa_q$. We define $\mathcal{K}_{cone} = \frac{\hat{y}_k^{high} - y_k}{r_{\hat{y}_k^{high}} - r_{y_k}}$, which can be interpreted as the slope that connects the two balls, see Fig. 5. We also write

$$r_{\hat{y}_k^{high}} = r_{y_k} + (\hat{y}_k^{high} - y_k)/\mathcal{K}_{cone}.$$

Since $\hat{y}_k^{high} - y_k > \kappa_q$, the numerator of $\mathcal{K}_{cone}$ is greater than the numerator of $\mathcal{K}_q$ as in (11), and, since $d > r_{\hat{y}_k^{high}} - r_{y_k}$ by definition of the diameter, we have $\mathcal{K}_{cone} > \mathcal{K}_q$. Note that $\mathcal{K}_q$ is independent of the value y_k.

We define $\mathcal{B}^{large}$ as an n-ball centered at x_* with radius r_{large}, where $r_{large} = r_{y_k} + (\hat{y}_k^{high} - y_k)/\mathcal{K}_q$. Here we see that $r_{\hat{y}_k^{high}} \le r_{large}$ since $\mathcal{K}_q \le \mathcal{K}_{cone}$. Therefore $S_{\hat{y}_k^{high}} \subset \mathcal{B}_{\hat{y}_k^{high}} \subset \mathcal{B}^{large}$, as illustrated in Fig. 5.

A lower bound on the ratios of volumes is constructed in terms of the dimension n, using multi-dimensional geometry theorems [10],

$$\frac{\nu(S_{y_k})}{\nu(S_{\hat{y}_k^{high}})} \ge \frac{\nu(\mathcal{B}_{y_k})}{\nu(\mathcal{B}_{\hat{y}_k^{high}})} \ge \frac{\nu(\mathcal{B}_{y_k})}{\nu(\mathcal{B}^{large})} = \left(\frac{r_{y_k}}{r_{y_k} + \frac{\hat{y}_k^{high} - y_k}{\mathcal{K}_q}} \right)^n.$$

Since $\hat{y}_k^{high} - y_k \le \frac{2 \cdot \sigma \cdot z_{\alpha/2}}{\sqrt{R}}$, as given in the theorem statement, we have the following lower bound,

$$\frac{\nu(S_{y_k})}{\nu(S_{\hat{y}_k^{high}})} \ge \left(\frac{r_{y_k}}{r_{y_k} + \frac{2 \cdot \sigma \cdot z_{\alpha/2}}{\mathcal{K}_q \sqrt{R}}} \right)^n.$$

We want to determine R, such that

$$\left(\frac{r_{y_k}}{r_{y_k}+\frac{2\cdot\sigma\cdot z_{\alpha/2}}{\mathcal{K}_q\sqrt{R}}}\right)^n \geq q.$$

Through several algebraic manipulations to isolate R, we determine that (13) holds if $R \geq \left(\frac{\sqrt[n]{q}\cdot 2\cdot\sigma\cdot z_{\alpha/2}}{(1-\sqrt[n]{q})\cdot r_{y_k}\cdot\mathcal{K}_q}\right)^2$.

Finally, since $y_k \geq y_* + \epsilon$, then $r_{y_k} \geq r_{y_*+\epsilon}$, and therefore (13) holds if

$$R \geq \left(\frac{\sqrt[n]{q}\cdot 2\cdot\sigma\cdot z_{\alpha/2}}{(1-\sqrt[n]{q})\cdot r_{y_*+\epsilon}\cdot\mathcal{K}_q}\right)^2$$

which proves Theorem 1. ■

A.2 Lemmas

The bound on replications, which is quadratic in dimension given in (14), uses a bound stated in Lemma 1. Theorems 2 and 4 make use of Lemma 30 from [23], which is repeated here for convenience as Lemma 2.

Lemma 1 *For a given constant a such that $0 < a < 1$, and a variable $n \geq 1$, then the function $f(n) = \frac{a^{1/n}}{1-a^{1/n}}$ is bounded by a linear function of n, that is,*

$$f(n) = \frac{a^{1/n}}{1-a^{1/n}} \leq \frac{a}{1-a} - (n-1)\frac{ln(a)}{(1-a)^2}. \tag{29}$$

A proof of Lemma 1 is omitted. Other bounds are possible that are still linear in n.
The following lemma is used in the proof of Theorem 3.

Lemma 2 *(cf. [23]) Let $\bar{Y}_k^A, k = 0, 1, 2, \ldots$ and $\bar{Y}_k^B, k = 0, 1, 2, \ldots$ be two sequences of objective function values generated by algorithms A and B respectively for solving a minimization problem, such that $\bar{Y}_{k+1}^A \leq \bar{Y}_k^A$ and $\bar{Y}_{k+1}^B \leq \bar{Y}_k^B$ for $k = 0, 1, \ldots$. For $y_* < y, z \leq y^*$ and $k = 0, 1, \ldots$, if*

1. $P(\bar{Y}_{k+1}^A \leq y | \bar{Y}_k^A = z) \geq P(\bar{Y}_{k+1}^B \leq y | \bar{Y}_k^B = z)$
2. $P(\bar{Y}_{k+1}^A \leq y | \bar{Y}_k^A = z)$ *is non-increasing in z, and*
3. $P(\bar{Y}_0^A \leq y) \geq P(\bar{Y}_0^B \leq y)$

then $P(Y_k^A \leq y) \geq P(Y_k^B \leq y)$ for $k = 0, 1, \ldots$ and $y_* < y \leq y^*$.

A proof of Lemma 2 can be found in [23].

A.3 Proof of Theorem 2

Proof of Theorem 2: Using the notation in HAS-E on the kth iteration, and based on Lemma 2, as in [23], if the following conditions hold for $y_* < y$, $\bar{y}_k \leq y^*$ and $k = 0, 1, \ldots$,

(I) $P(\bar{Y}_{k+1}^{HASE} \leq y | \bar{Y}_k^{HASE} = \bar{y}_k) \geq P(\bar{Y}_{k+1}^{HAS1} \leq y | \bar{Y}_k^{HAS1} = \bar{y}_k)$
(II) $P(\bar{Y}_{k+1}^{HASE} \leq y | \bar{Y}_k^{HASE} = \bar{y}_k)$ is non-increasing in $\bar{y}_k$, and
(III) $P(\bar{Y}_0^{HASE} \leq y) \geq P(\bar{Y}_0^{HAS1} \leq y)$

then $P(\bar{Y}_k^{HASE} \leq y) \geq P(\bar{Y}_k^{HAS1} \leq y)$ for $k = 0, 1, \ldots$ and for $y_* < y \leq y^*$.

The first step is to prove (I). When $y \geq \bar{y}_k$, (I) is true trivially (since the conditional probability equals one on both sides). Now, when $y < \bar{y}_k$, we bound the left-hand side of the expression in (I), as,

$$P(\bar{Y}_{k+1}^{HASE} \leq y | \bar{Y}_k^{HASE} = \bar{y}_k) \geq \gamma \cdot P(\bar{Y}_{k+1}^{HASE} \leq y | \bar{Y}_{k+1}^{HASE} \leq \bar{y}_k^{high}, \bar{Y}_k^{HASE} = \bar{y}_k) \quad (30)$$

where we condition on the event that HASE "betters", that is, that HASE samples from the normalized restriction of ζ on $S_{\bar{y}_k^{high}}$, and consequently $\bar{Y}_{k+1}^{HASE} \leq \bar{y}_k^{high}$, which occurs with probability at least γ by the bound on the bettering probability in Assumption 1 (*ii*).

We next consider the event $\{\bar{y}_k \leq \bar{y}_k^{high}\}$, which occurs with probability at least $1 - \alpha$ by (8). We rewrite (30) as,

$$\begin{aligned} &P(\bar{Y}_{k+1}^{HASE} \leq y | \bar{Y}_k^{HASE} = \bar{y}_k) \\ &\quad \geq \gamma \cdot (1 - \alpha) \cdot P(\bar{Y}_{k+1}^{HASE} \leq y | \{\bar{y}_k \leq \bar{y}_k^{high}\}, \bar{Y}_{k+1}^{HASE} \leq \bar{y}_k^{high}) \end{aligned} \quad (31)$$

dropping the condition $\bar{Y}_k^{HASE} = \bar{y}_k$ because it is captured in the other conditions.

To further develop the bound in terms of HAS1, we note that $P(\bar{Y}_{k+1}^{HASE} \leq y) \geq P(\bar{Y}_0^{HASE} \leq y) \geq P(\bar{Y}_0^{HAS1} \leq y)$ by Assumption 1 (*i*), and since $P(\bar{Y}_{k+1}^{HASE} \leq y) = P(\bar{Y}_{k+1}^{HASE} \leq y \bigcap \bar{Y}_{k+1}^{HASE} \leq \bar{y}_k^{high})$ and $P(\bar{Y}_0^{HAS1} \leq y) = P(\bar{Y}_0^{HAS1} \leq y \bigcap \bar{Y}_0^{HAS1} \leq \bar{y}_k^{high})$ we can write

$$\begin{aligned} &P(\bar{Y}_{k+1}^{HASE} \leq y \bigcap \bar{Y}_{k+1}^{HASE} \leq \bar{y}_k^{high}) \left(\frac{P(\bar{Y}_{k+1}^{HASE} \leq \bar{y}_k^{high})}{P(\bar{Y}_{k+1}^{HASE} \leq \bar{y}_k^{high})} \right) \\ &\quad \geq P(\bar{Y}_0^{HAS1} \leq y \bigcap \bar{Y}_0^{HAS1} \leq \bar{y}_k^{high}) \left(\frac{P(\bar{Y}_0^{HAS1} \leq \bar{y}_k^{high})}{P(\bar{Y}_0^{HAS1} \leq \bar{y}_k^{high})} \right) \end{aligned}$$

implying

$$\frac{P(\bar{Y}_{k+1}^{HASE} \leq y \bigcap \bar{Y}_{k+1}^{HASE} \leq \bar{y}_k^{high})}{P(\bar{Y}_{k+1}^{HASE} \leq \bar{y}_k^{high})} \geq \frac{P(\bar{Y}_0^{HAS1} \leq y \bigcap \bar{Y}_0^{HAS1} \leq \bar{y}_k^{high})}{P(\bar{Y}_0^{HAS1} \leq \bar{y}_k^{high})}$$

and

$$P(\bar{Y}_{k+1}^{HASE} \leq y | \{\bar{y}_k \leq \bar{y}_k^{high}\}, \bar{Y}_{k+1}^{HASE} \leq \bar{y}_k^{high}) \geq P(\bar{Y}_0^{HAS1} \leq y | \{\bar{y}_k \leq \bar{y}_k^{high}\}, \bar{Y}_0^{HAS1} \leq \bar{y}_k^{high}).$$

We now rewrite (31) as

$$\begin{aligned} &P(\bar{Y}_{k+1}^{HASE} \leq y | \bar{Y}_k^{HASE} = \bar{y}_k) \\ &\quad \geq \gamma \cdot (1 - \alpha) \cdot P(\bar{Y}_0^{HAS1} \leq y | \{\bar{y}_k \leq \bar{y}_k^{high}\}, \bar{Y}_0^{HAS1} \leq \bar{y}_k^{high}). \end{aligned} \quad (32)$$

Therefore, we can create a lower bound for (32):

$$\begin{aligned} P(\bar{Y}_{k+1}^{HASE} \leq y | \bar{Y}_k^{HASE} = \bar{y}_k) &\geq \gamma \cdot (1 - \alpha) \cdot P(\bar{Y}_0^{HAS1} \leq y | \bar{Y}_0^{HAS1} \leq \bar{y}_k^{high}, \{\bar{y}_k \leq \bar{y}_k^{high}\}) \\ &= \gamma \cdot (1 - \alpha) \cdot \frac{P(\bar{Y}_0^{HAS1} \leq y)}{P(\bar{Y}_0^{HAS1} \leq \bar{y}_k^{high})} \\ &= \gamma \cdot (1 - \alpha) \cdot \frac{P(\bar{Y}_0^{HAS1} \leq y)}{P(\bar{Y}_0^{HAS1} \leq \bar{y}_k)} \cdot \frac{P(\bar{Y}_0^{HAS1} \leq \bar{y}_k)}{P(\bar{Y}_0^{HAS1} \leq \bar{y}_k^{high})} \end{aligned}$$

$$\geq \gamma \cdot (1-\alpha) \cdot \frac{\nu(S_y)}{\nu(S_{\bar{y}_k})} \cdot \frac{\nu(S_{\bar{y}_k})}{\nu(S_{\bar{y}_k^{high}})}$$

$$\geq \gamma \cdot (1-\alpha) \cdot \frac{\nu(S_y)}{\nu(S_{\bar{y}_k})} \cdot q. \tag{33}$$

The last inequality makes use of the lower bound developed in Theorem 1, $\frac{\nu(S_{\bar{y}_k})}{\nu(S_{\bar{y}_k^{high}})} \geq q$.

We similarly expand the expression for HAS1 in the right-hand side of (I), noting that HAS1 either improves or stays where it is, yielding,

$$P(\bar{Y}_{k+1}^{HAS1} \leq y | \bar{Y}_k^{HAS1} = \bar{y}_k) = b^{HAS1}(\bar{y}_k) P\left(\bar{Y}_{k+1}^{HAS1} \leq y | \bar{Y}_k^{HAS1} = \bar{y}_k\right)$$

where X_{k+1}^{HAS1} is sampled according to the normalized restriction of the uniform distribution on the improving level set. Combining this with the bettering probability of HAS1, $b(y) = \gamma \cdot (1-\alpha) \cdot q$, and when HAS1 "betters", we have,

$$P(\bar{Y}_{k+1}^{HAS1} \leq y | \bar{Y}_k^{HAS1} = \bar{y}_k) = \gamma \cdot (1-\alpha) \cdot q \cdot \frac{\nu(S_y)}{\nu(S_{\bar{y}_k})}. \tag{34}$$

Combining (33) and (34) proves condition (I).

We go on to prove (II), that $P(\bar{Y}_{k+1}^{HASE} \leq y | \bar{Y}_k^{HASE} = \bar{y}_k)$ is non-increasing in $\bar{y}_k$. Suppose that $\bar{y}_k$ and $\bar{y}_k'$ are such that $\bar{y}_k < \bar{y}_k'$. To show (II) we want to show that:

$$P(\bar{Y}_{k+1}^{HASE} \leq y | \bar{Y}_k^{HASE} = \bar{y}_k) \geq P(\bar{Y}_{k+1}^{HASE} \leq y | \bar{Y}_k^{HASE} = \bar{y}_k').$$

The approach is to condition on the value of $\bar{y}_k^{high}$, and since HAS-E samples on $S_{\bar{y}_k^{high}}$ in Step 2 of the algorithm, we know that $P(\bar{Y}_{k+1}^{HASE} \leq y | \bar{Y}_k^{HASE} = \bar{y}_k, \bar{y}_k^{high} = u)$ is non-increasing, therefore, we have,

$$P(\bar{Y}_{k+1}^{HASE} \leq y | \bar{Y}_k^{HASE} = \bar{y}_k)$$
$$= \int_{-\infty}^{\infty} P(\bar{Y}_{k+1}^{HASE} \leq y | \bar{Y}_k^{HASE} = \bar{y}_k, \bar{y}_k^{high} = z) \cdot dP(\bar{y}_k^{high} \leq z | \bar{Y}_k^{HASE} = \bar{y}_k)$$

and because $\int_{-\infty}^{z} dP\left(\bar{Y}_{k+1}^{HASE} \leq y | \bar{Y}_k^{HASE} = \bar{y}_k, \bar{y}_k^{high} = u\right) = P\left(\bar{Y}_{k+1}^{HASE} \leq y | \bar{Y}_k^{HASE} = \bar{y}_k, \bar{y}_k^{high} = z\right) - P\left(\bar{Y}_{k+1}^{HASE} \leq y | \bar{Y}_k^{HASE} = \bar{y}_k, \bar{y}_k^{high} = -\infty\right)$, and since $P\left(\bar{Y}_{k+1}^{HASE} \leq y | \bar{Y}_k^{HASE} = \bar{y}_k, \bar{y}_k^{high} = -\infty\right) = 1$ (trivially), we substitute $P\left(\bar{Y}_{k+1}^{HASE} \leq y | \bar{Y}_k^{HASE} = \bar{y}_k, \bar{y}_k^{high} = z\right)$ as follows,

$$= \int_{-\infty}^{\infty} \left(1 + \int_{-\infty}^{z} dP\left(\bar{Y}_{k+1}^{HASE} \leq y | \bar{Y}_k^{HASE} = \bar{y}_k, \bar{y}_k^{high} = u\right)\right) \cdot dP\left(\bar{y}_k^{high} \leq z | \bar{Y}_k^{HASE} = \bar{y}_k\right)$$

and reversing the order of integration, we get

$$= 1 + \int_{-\infty}^{\infty} \int_{u}^{\infty} dP\left(\bar{y}_k^{high} \leq z | \bar{Y}_k^{HASE} = \bar{y}_k\right) \cdot dP\left(\bar{Y}_{k+1}^{HASE} \leq y | \bar{Y}_k^{HASE} = \bar{y}_k, \bar{y}_k^{high} = u\right)$$
$$= 1 + \int_{-\infty}^{\infty} \left(1 - P\left(\bar{y}_k^{high} \leq u | \bar{Y}_k^{HASE} = \bar{y}_k\right)\right) \cdot dP(\bar{Y}_{k+1}^{HASE} \leq y | \bar{Y}_k^{HASE} = \bar{y}_k, \bar{y}_k^{high} = u).$$

Now to show a lower bound in terms of $\bar{y}_k'$. However, since $dP\left(\bar{Y}_{k+1}^{HASE} \leq y | \bar{Y}_k^{HASE} = \bar{y}_k, \bar{y}_k^{high} = u\right) \leq 0$, and since $P\left(\bar{Y}_{k+1}^{HASE} \leq y | \bar{Y}_k^{HASE} = \bar{y}_k, \bar{y}_k^{high} = u\right)$ is

non-increasing in $\bar{y}_k^{high}$, and since, $P\left(\bar{y}_k^{high} \leq u | \bar{Y}_k^{HASE} = \bar{y}_k\right) \geq P\left(\bar{y}_k^{high} \leq u | \bar{Y}_k^{HASE} = \bar{y}_k'\right)$, the probability that $\bar{y}_k^{high}$ is lower than u is always greater for $\bar{y}_k < \bar{y}_k'$, then

$$1 + \int_{-\infty}^{\infty} \left(1 - P\left(\bar{y}_k^{high} \leq u | \bar{Y}_k^{HASE} = \bar{y}_k\right)\right) \cdot dP\left(\bar{Y}_{k+1}^{HASE} \leq y | \bar{Y}_k^{HASE} = \bar{y}_k, \bar{y}_k^{high} = u\right)$$
$$\geq 1 + \int_{-\infty}^{\infty} \left(1 - P\left(\bar{y}_k^{high} \leq u | \bar{Y}_k^{HASE} = \bar{y}_k'\right)\right) \cdot dP\left(\bar{Y}_{k+1}^{HASE} \leq y | \bar{Y}_k^{HASE} = \bar{y}_k, \bar{y}_k^{high} = u\right)$$

which is equivalent to

$$= 1 + \int_{-\infty}^{\infty} \int_{u}^{\infty} \left(dP\left(\bar{y}_k^{high} \leq z | \bar{Y}_k^{HASE} = \bar{y}_k'\right)\right) \cdot dP(\bar{Y}_{k+1}^{HASE} \leq y | \bar{Y}_k^{HASE} = \bar{y}_k, \bar{y}_k^{high} = u)$$

and reversing the order of integration:

$$= 1 + \int_{-\infty}^{\infty} \int_{-\infty}^{z} dP\left(\bar{Y}_{k+1}^{HASE} \leq y | \bar{Y}_k^{HASE} = \bar{y}_k, \bar{y}_k^{high} = u\right) \cdot \left(dP\left(\bar{y}_k^{high} \leq z | \bar{Y}_k^{HASE} = \bar{y}_k'\right)\right)$$
$$= \int_{-\infty}^{\infty} P(\bar{Y}_{k+1}^{HASE} \leq y | \bar{Y}_k^{HASE} = \bar{y}_k, \bar{y}_k^{high} = z) \cdot dP(\bar{y}_k^{high} \leq z | \bar{Y}_k^{HASE} = \bar{y}_k')$$

therefore, since $P(\bar{Y}_{k+1}^{HASE} \leq y | \bar{Y}_k^{HASE} = \bar{y}_k, \bar{y}_k^{high} = z) = P(\bar{Y}_{k+1}^{HASE} \leq y | \bar{Y}_k^{HASE} = \bar{y}_k', \bar{y}_k^{high} = z)$, we write:

$$= \int_{-\infty}^{\infty} P(\bar{Y}_{k+1}^{HASE} \leq y | \bar{Y}_k^{HASE} = \bar{y}_k', \bar{y}_k^{high} = z) \cdot dP(\bar{y}_k^{high} \leq z | \bar{Y}_k^{HASE} = \bar{y}_k')$$
$$= P(\bar{Y}_{k+1}^{HASE} \leq y | \bar{Y}_k^{HASE} = \bar{y}_k')$$

which proves (II).

Lastly, condition (III) from Lemma 2 is true by Assumption 1 (*i*) that $P(\bar{Y}_0^{HASE} \leq y) \geq P(\bar{Y}_0^{HAS1} \leq y)$. This proves the theorem through reference to Lemma 2. ■

A.4 Proof of Theorem 3

Proof of Theorem 3: By stochastic dominance in Theorem 2, the expected number of iterations to achieve a value within $S_{y_*+\epsilon}$ for HAS-E is less than or equal to the number for HAS1. Since the bettering probability for HAS1 is $b(y) = \gamma \cdot (1-\alpha) \cdot q$ for all $y_* < y \leq y^*$, using (4), we have

$$E[N_I^{HASE}(y_* + \epsilon)] \leq 1 + \int_{y_*+\epsilon}^{\infty} \frac{d\rho(t)}{\gamma \cdot (1-\alpha) \cdot q \cdot p(t)}$$

and since HAS1 uses uniform sampling, i.e., $p(y) = \frac{\nu(S_y)}{\nu(S)}$, we have

$$= 1 + \frac{1}{\gamma \cdot (1-\alpha) \cdot q} \cdot ln\left(\frac{\nu(S)}{\nu(S_{y_*+\epsilon})}\right).$$

Using a constant number of replications R for each iteration, yields

$$E[N_R^{HASE}(y_* + \epsilon)] = R \cdot E[N_I^{HASE}(y_* + \epsilon)].$$

Setting $R_k = R$ as in (16) yields the result in (18). ■

Appendix B Proofs of Theorems for QAS-E Analysis

The proofs of the theorems for QAS-E are similar to the proofs for HAS-E. The proof of Theorem 4 for QAS-E is provided for completeness.

Proof of Theorem 4 Similar to the proof of Theorem 2, if the three conditions listed in Lemma 2 hold, then $P(\bar{Y}_k^{QASE} \leq y) \geq P(\bar{Y}_k^{HAS2} \leq y)$ for $k = 0, 1, \ldots$ and for $y_* < y \leq y^*$.

We start by proving the first condition in Lemma 2, that is, we show that

$$P(\bar{Y}_{k+1}^{QASE} \leq y | \bar{Y}_k^{QASE} = \bar{y}_k) \geq P(\bar{Y}_{k+1}^{HAS2} \leq y | \bar{Y}_k^{HAS2} = \bar{y}_k)$$

for $y_* < y, \bar{y}_k \leq y^*$ and $k = 0, 1, \ldots$. When $y \geq \bar{y}_k$, $P(\bar{Y}_{k+1}^{QASE} \leq y | \bar{Y}_k^{QASE} = \bar{y}_k) = P(\bar{Y}_{k+1}^{HAS2} \leq y | \bar{Y}_k^{QASE} = \bar{y}_k) = 1$, and the first condition holds.

Now, when $y < \bar{y}_k$, we bound the left-hand side of the expression in (I) by conditioning on the event that X_{k+1}^{QASE} "betters", that is, the event that $\bar{Y}_{k+1}^{QASE} \leq \bar{y}_k^{high}$, yielding

$$P(\bar{Y}_{k+1}^{QASE} \leq y | \bar{Y}_k^{QASE} = \bar{y}_k) \geq \gamma \cdot P(\bar{Y}_{k+1}^{QASE} \leq y | \bar{Y}_{k+1}^{QASE} \leq \bar{y}_k^{high}, \bar{Y}_k^{QASE} = \bar{y}_k) \quad (35)$$

by Assumption 2 (*ii*).

We next consider the event $\{\bar{y}_k \leq \bar{y}_k^{high}\}$, which occurs with probability at least $1 - \alpha$, by (8). We rewrite (35) as,

$$\begin{aligned} &P(\bar{Y}_{k+1}^{QASE} \leq y | \bar{Y}_k^{QASE} = \bar{y}_k) \\ &\quad \geq \gamma \cdot (1 - \alpha) \cdot P(\bar{Y}_{k+1}^{QASE} \leq y | \bar{Y}_{k+1}^{QASE} \leq \bar{y}_k^{high}, \{\bar{y}_k \leq \bar{y}_k^{high}\}, \bar{Y}_k^{QASE} = \bar{y}_k). \end{aligned} \quad (36)$$

From Assumption 2 (*i*), we have

$$P(\bar{Y}_{k+1}^{QASE} \leq y | \bar{Y}_k^{QASE} = \bar{y}_k) \geq \gamma \cdot (1 - \alpha) \cdot P(\bar{Y}_0^{HAS2} \leq y | \bar{Y}_0^{HAS2} \leq \bar{y}_k^{high}).$$

Making use of the lower bound developed in Theorem 1, $\frac{\nu(S_{\bar{y}_k})}{\nu(S_{\bar{y}_k^{high}})} \geq q$, we have,

$$\begin{aligned} P(\bar{Y}_{k+1}^{QASE} \leq y | \bar{Y}_k^{QASE} = \bar{y}_k) &\geq \gamma \cdot (1 - \alpha) \cdot P(\bar{Y}_0^{HAS2} \leq y | \bar{Y}_0^{HAS2} \leq \bar{y}_k^{high}) \\ &= \gamma \cdot (1 - \alpha) \cdot \frac{P(\bar{Y}_0^{HAS2} \leq y)}{P(\bar{Y}_0^{HAS2} \leq \bar{y}_k^{high})} \\ &= \gamma \cdot (1 - \alpha) \cdot \frac{P(\bar{Y}_0^{HAS2} \leq y)}{P(\bar{Y}_0^{HAS2} \leq \bar{y}_k)} \cdot \frac{P(\bar{Y}_0^{HAS2} \leq \bar{y}_k)}{P(\bar{Y}_0^{HAS2} \leq \bar{y}_k^{high})} \\ &\geq \gamma \cdot (1 - \alpha) \cdot \frac{\nu(S_y)}{\nu(S_{\bar{y}_k})} \cdot \frac{\nu(S_{\bar{y}_k})}{\nu(S_{\bar{y}_k^{high}})} \\ &\geq \gamma \cdot (1 - \alpha) \cdot \frac{\nu(S_y)}{\nu(S_{\bar{y}_k})} \cdot q. \end{aligned} \quad (37)$$

We similarly expand the expression for HAS2 in the right-hand side of (I), noting that HAS2 either improves or stays where it is, yielding,

$$P(\bar{Y}_{k+1}^{HAS2} \leq y | \bar{Y}_k^{HAS2} = \bar{y}_k) = b^{HAS2}(\bar{y}_k) P\left(\bar{Y}_{k+1}^{HAS2} \leq y | \bar{Y}_k^{HAS2} = \bar{y}_k\right)$$

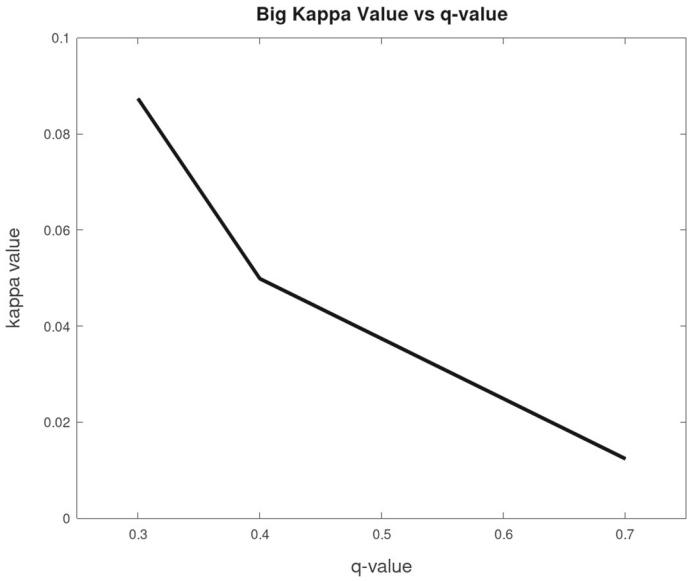


Fig. 6 Values for $\mathcal{K}_q$ for the one-dimensional sample problem for $q \in [0.3, 0.7]$

and since the bettering probability of HAS2 equals $\gamma \cdot (1-\alpha) \cdot q$, and when HAS2 "betters", it samples uniformly on the improving level set, we have,

$$P(\bar{Y}_{k+1}^{HAS2} \leq y | \bar{Y}_k^{HAS2} = \bar{y}_k) = \gamma \cdot (1-\alpha) \cdot q \cdot \frac{\nu(S_y)}{\nu(S_{\bar{y}_k})}. \tag{38}$$

Combining (37) and (38) proves condition (I).
The second condition in Lemma 2 is satisfied directly by Assumption 2 (*iii*). The third condition in Lemma 2 is satisfied by Assumption 2 (*i*). This proves the theorem by Lemma 2. ■

Appendix C Details for Sample Problem

The one-dimensional sample problem $f(x)$, illustrated in Fig. 3, is defined for $x \in [-4, 4]$. The calculations use the following parameter settings: $\sigma = 1$, $\alpha = 0.05$, $\gamma = 0.5$, and $\epsilon = 0.3$. The values for $\mathcal{K}_q$ are calculated numerically for a range of values of q, and are shown in Fig. 6.

References

1. Andradóttir, S., Prudius, A.A.: Adaptive random search for continuous simulation optimization. Naval Res. Logist. **57**, 583–604 (2010)
2. Baritompa, William P., Bulger, David W., Wood, Graham R.: Generating functions and the performance of backtracking adaptive search. J. Glob. Optim. **37**(2), 159–175 (2007)
3. Guus, C., Boender, E., Edwin Romeijn, H.: Stochastic methods. In *Handbook of global optimization*, pages 829–869. Springer: Berlin. 1995

4. Bulger, David W., Wood, Graham R.: Hesitant adaptive search for global optimisation. Math. Prog. **81**(1), 89–102 (1998)
5. Michael, CFu.: Handbook of Simulation Optimization, vol. 216. Springer, New York, New York, NY (2015)
6. Ho, Y.C., Cassandras, C.G., Chen, C.H., Dai, L.: Ordinal optimisation and simulation. J. Oper. Res. Soc. **51**, 490–500 (2000)
7. Ho, Y.C., Zhao, Q.C., Jia, Q.S.: Ordinal optimization: soft optimization for hard problems. Springer, Berlin, Germany (2007)
8. Hu, Jiaqiao., Wang, Yongqiang., Zhou, Enlu., Fu, Michael C., Marcus, Steven I.: A survey of some model-based methods for global optimization. In *Optimization, Control, and Applications of Stochastic Systems*, pages 157–179. Birkhäuser Boston, 2012
9. Jinyang, J., Hu, J., Peng, Y.: Quantile-based policy optimization for reinforcement learning, 2022. available on arXiv:2201.11463
10. Kendall, Maurice G.: *A Course in the Geometry of n Dimensions*. Courier Corporation, (2004)
11. Kiatsupaibul, S., Smith, R.L., Zabinsky, Z.B.: Single observation adaptive search for continuous simulation. Oper. Res. **66**, 1713–1727 (2018)
12. Kiatsupaibul, S., Smith, R.L., Zabinsky, Z.B.: Single observation adaptive search for discrete and continuous simulation. Oper. Res. Lett. **48**, 666–673 (2020)
13. Kleywegt, A.J., Shapiro, A., Homem-de Mello, T.: The sample average approximation method for stochastic discrete optimization. SIAM J. Optim. **12**, 479–502 (2002)
14. Linz, David D.: Optimizing population healthcare resource allocation under uncertainty using global optimization methods. *University of Washington Dissertation*, (2018)
15. Linz, D.D., Zabinsky, Z.B., Kiatsupaibul, S, Smith, R.L.: A computational comparison of simulation optimization methods using single observations within a shrinking ball on noisy black-box functions with mixed integer and continuous domains. In Chan, W.K.V., D'Ambrogio, A., Zacharewicz, G., Mustafee, N., Wainer, G., Page, E., (eds.) *Proceedings of the 2017 Winter Simulation Conference*, pages 2045 – 2056, Washington, DC, 2017
16. Locatelli, M., Schoen, F.: *Global optimization: theory, algorithms, and applications*, volume 15. SIAM, 2013
17. Locatelli, Marco, Schoen, Fabio: (Global) optimization: historical notes and recent developments. EURO J. Comput. Optim. **9**, 100012 (2021)
18. Pardalos, Panos M., Edwin Romeijn, H., Tuy, Hoang: Recent developments and trends in global optimization. J. Comput. Appl. Math. **124**(1–2), 209–228 (2000)
19. Raphael, Benny, Smith, Ian F. C.: A direct stochastic algorithm for global search. Appl. Math. Comput. **146**(2–3), 729–758 (2003)
20. Raphael, Benny., Smith, Ian F. C.: Global search through sampling using a PDF. In *Stochastic Algorithms: Foundations And Applications*, volume 2827, pages 71–82. Springer (2003)
21. Edwin Romeijn, H., Smith, Robert L.: Simulated annealing and adaptive search in global optimization. Prob. Eng. Inf. Sci. **8**(4), 571–590 (1994)
22. Rubinstein, Reuven Y., Kroese, Dirk P.: The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization. Monte-Carlo Simulation and Machine Learning. Springer, Cambridge, UK (2004)
23. Shen, Yanfang.: Annealing Adaptive Search With Hit-and-Run Sampling Methods for Global Optimization. *University of Washington Dissertation* (2005)
24. Shen, Yanfang, Kiatsupaibul, Seksan, Zabinsky, Zelda B., Smith, Robert L.: An analytically derived cooling schedule for simulated annealing. J. Glob. Optim. **38**(3), 333–365 (2007)
25. Wood, Graham R., Bulger, David W., Baritompa, William P., Alexander, D.: Backtracking adaptive search: distribution of number of iterations to convergence. J. Optim. Theory Appl. **128**(3), 547–562 (2006)
26. Wood, Graham R., Zabinsky, Zelda B., Kristinsdottir, Birna P.: Hesitant adaptive search: the distribution of the number of iterations to convergence. Math. Progr. **89**(3), 479–486 (2001)
27. Zabinsky, Zelda B.: Stochastic adaptive search for global optimization. Kluwer Academic Publishers originally, Springer Science & Business Media (2003)
28. Zabinsky, Zelda B., Bulger, David, Khompatraporn, Charoenchai: Stopping and restarting strategy for stochastic sequential search in global optimization. J. Glob. Optim. **46**, 273–286 (2010)
29. Zabinsky, Zelda B., Huang, Hao: A partition-based optimization approach for level set approximation: Probabilistic branch and bound. In: Alice, S. (ed.) Women in Industrial and Systems Engineering: Key Advances and Perspectives on Emerging Topics. Springer, Berlin (2020)
30. Zabinsky, Zelda B., Smith, Robert L.: Pure adaptive search in global optimization. Math. Progr. **53**(1–3), 323–338 (1992)
31. Zabinsky, Zelda B., Wood, Graham R., Steel, Mike A., Baritompa, William P.: Pure adaptive search for finite global optimization. Math. Progr. **69**(1–3), 443–448 (1995)