

TIGHT BOUNDS FOR BLIND SEARCH ON THE INTEGERS

MARTIN DIETZFELBINGER¹, JONATHAN E. ROWE², INGO WEGENER³,
AND PHILIPP WOELFEL⁴

¹ Fakultät für Informatik und Automatisierung, Technische Univ. Ilmenau, 98684 Ilmenau, Germany
E-mail address: martin.dietzfelbinger@tu-ilmenau.de

² School of Computer Science, University of Birmingham, Birmingham B15 2TT, United Kingdom
E-mail address: J.E.Rowe@cs.bham.ac.uk

³ FB Informatik, LS2, Universität Dortmund, 44221 Dortmund, Germany
E-mail address: ingo.wegener@uni-dortmund.de

⁴ Department of Computer Science, University of Calgary, Calgary, Alberta T2N 1N4, Canada
E-mail address: woelfel@cpsc.ucalgary.ca

ABSTRACT. We analyze a simple random process in which a token is moved in the interval $A = \{0, \dots, n\}$: Fix a probability distribution μ over $\{1, \dots, n\}$. Initially, the token is placed in a random position in A . In round t , a random value d is chosen according to μ . If the token is in position $a \geq d$, then it is moved to position $a - d$. Otherwise it stays put. Let T be the number of rounds until the token reaches position 0. We show tight bounds for the expectation of T for the optimal distribution μ . More precisely, we show that $\min_{\mu} \{E_{\mu}(T)\} = \Theta((\log n)^2)$. For the proof, a novel potential function argument is introduced. The research is motivated by the problem of approximating the minimum of a continuous function over $[0, 1]$ with a “blind” optimization strategy.

1. Introduction

For a positive integer n , assume a probability distribution μ on $X = \{1, \dots, n\}$ is given. Consider the following random process. A token moves in $A = \{0, \dots, n\}$, as follows:

- Initially, place the token in some position in A .
- In round t : The token is at position $a \in A$. Choose an element d from X at random, according to μ . If $d \leq a$, move the token to position $a - d$ (the step is “accepted”), otherwise leave it where it is (the step is “rejected”).

Work of the first author was done in part while visiting ETH Zürich, Switzerland. The third author was supported in part by the DFG collaborative research project SFB 531. Work of the last author was done in part while at the University of Toronto, supported by DFG grant WO1232/1-1 and by SUN Microsystems. Joint work on this topic was initiated during Dagstuhl Seminar 06111 on Complexity of Boolean Functions (2006).



When the token has reached position 0, no further moves are possible, and we regard the process as finished.

At the beginning the token is placed at a position chosen uniformly at random from $\{1, \dots, n\} = A - \{0\}$. (For simplicity of notation, we prefer this initial distribution over the possibly more natural uniform distribution on $\{0, \dots, n\}$. Of course, there is no real difference between the two starting conditions.) Let T be the number of rounds needed until position 0 is reached. A basic performance parameter for the process is $\mathbf{E}_\mu(T)$. As μ varies, the value $\mathbf{E}_\mu(T)$ will vary. The probability distribution μ may be regarded as a strategy. We ask: How should μ be chosen so that $\mathbf{E}_\mu(T)$ is as small as possible?

It is easy to exhibit distributions μ such that $\mathbf{E}_\mu(T) = O((\log n)^2)$. (All asymptotic notation in this paper refers to $n \rightarrow \infty$.) In particular, we will see that the “harmonic distribution” given by

$$\mu_{\text{har}}(d) = \frac{1}{d \cdot H_n}, \text{ for } 1 \leq d \leq n, \quad (1.1)$$

where $H_n = \sum_{1 \leq d \leq n} \frac{1}{d}$ is the n th harmonic number, satisfies $\mathbf{E}_{\mu_{\text{har}}}(T) = O((\log n)^2)$. As the main result of the paper, we will show that this upper bound is optimal up to constant factors: $\mathbf{E}_\mu(T) = \Omega((\log n)^2)$, for every distribution μ . For the proof of this lower bound, we introduce a novel potential function technique, which may be useful in other contexts.

1.1. Motivation and Background: Blind Optimization Strategies

Consider the problem of minimizing a function $f : [0, 1] \rightarrow \mathbb{R}$, in which the definition of f is unknown: the only information we can gain about f is through trying sample points. This is an instance of a *black box optimization problem* [1]. One algorithmic approach to such problems is to start with an initial random point, and iteratively attempt to improve it by making random perturbations. That is, if the current point is $x \in [0, 1]$, then we choose some distance $d \in [0, 1]$ according to some probability distribution μ on $[0, 1]$, and move to $x + d$ or $x - d$ if this is an improvement. The distribution μ may be regarded as a “search strategy”. Such a search is “blind” in the sense that it does not try to estimate how close to the minimum it is and to adapt the distribution μ accordingly. The problem is how to specify μ . Of course, an optimal distribution μ depends on details of the function f .

The difficulty the search algorithm faces is that for general functions f there is no information about the scale of perturbations which are necessary to get close to the minimum. This leads us to the idea that the distribution might be chosen so that it is *scale invariant*, meaning that steps of all “orders of magnitude” occur with about the same probability. Such a distribution is described in [4]. One starts by specifying a minimum perturbation size ε . Then one chooses the probability density function $h(t) = 1/(pt)$ for $\varepsilon \leq t \leq 1$, and $h(t) = 0$ otherwise, where $p = \ln(1/\varepsilon)$ is the *precision* of the algorithm. (A random number distributed according to this density function may be generated by taking $d = \exp(-pu)$, where u is uniformly random in $[0, 1]$.)

For general functions f , no analysis of this search strategy is known, but in experiments on standard benchmark functions it (or higher dimensional variants) exhibits a good performance. (For details see [4].) From here on, we focus on the simple case where f is *unimodal*, meaning that it is strictly decreasing in $[0, x_0]$ and strictly increasing in $[x_0, 1]$, where x_0 is the unknown minimum point.

Remark 1.1. If one is given the information that f is unimodal, one will use other, deterministic search strategies, which approximate the optimum up to ε within $O(\log(1/\varepsilon))$

steps. As early as 1953, in [3], “Fibonacci search” was proposed and analyzed, which for a given tolerance ε uses the optimal number of steps in a very strong sense.

The “blind search” strategy from [4] can be applied to more general functions f , but the following analysis is valid only for unimodal functions. If the distance of the current point x from the optimum x_0 is $\tau \geq 2\varepsilon$ then every distance d with $\frac{\tau}{2} \leq d \leq \tau$ will lead to a new point with distance at most $\tau/2$. Thus, the probability of at least halving the distance to x_0 in one step is at least $\frac{1}{2} \int_{\tau/2}^{\tau} \frac{dt}{pt} = \frac{\ln 2}{2p}$, which is independent of the current state x . Obviously, then, the expected number of steps before the distance to x_0 has been halved is $2p/\ln 2$. We regard the algorithm to be successful if the current point has distance smaller than 2ε from x_0 . To reach this goal, the initial distance has to be halved at most $\log(1/\varepsilon)$ times, leading to a bound of $O(\log(1/\varepsilon)^2)$ for the expected number of steps.

The question then arises whether this is the best that can be achieved. Is there perhaps a choice for μ that works even better on unimodal functions? To investigate this question, we consider a discrete version of the situation. The domain of f is $A = \{0, \dots, n\}$, and f is strictly increasing, so that f takes its minimum at $x_0 = 0$. In this case, the search process is very simple: the actual values of f are irrelevant; going from a to $a+d$ is never an improvement. Actually, the search process is fully described by the simple random process from Section 1. How long does it take to reach the optimal point 0, for a μ chosen as cleverly as possible? For $\mu = \mu_{\text{har}}$, we will show an upper bound of $O((\log n)^2)$, with an argument very similar to that one leading to the bound $O(\log(1/\varepsilon)^2)$ in the continuous case. The main result of this paper is that the bound for the discrete case is optimal.

1.2. Formalization as a Markov chain

For the sake of simplicity, we let from now on $[a, b]$ denote the discrete interval $\{a, \dots, b\}$ if a and b are integers. Given a probability distribution μ on $[1, n]$, the Markov chain $R = (R_0, R_1, \dots)$ is defined over the state space $A = [0, n]$ by the transition probabilities

$$p_{a,a'} = \begin{cases} \mu(a - a') & \text{for } a' < a; \\ 1 - \sum_{1 \leq d \leq a} \mu(d) & \text{for } a' = a; \\ 0 & \text{for } a' > a. \end{cases}$$

Clearly, 0 is an absorbing state. We define the random variable $T = \min\{t \mid R_t = 0\}$. Let us write $\mathbf{E}_\mu(T)$ for the expectation of T if R_0 is uniformly distributed in $A - \{0\} = [1, n]$. We study $\mathbf{E}_\mu(T)$ in dependence on μ . In particular, we wish to identify distributions μ that make $\mathbf{E}_\mu(T)$ as small as possible (up to constant factors, where n is growing).

Observation 1.2. If $\mu(1) = 0$ then $\mathbf{E}_\mu(T) = \infty$.

This is because with probability $\frac{1}{n}$ position 1 is chosen as the starting point, and from state 1, the process will never reach 0 if $\mu(1) = 0$. As a consequence, for the whole paper we assume that all distributions μ that are considered satisfy

$$\mu(1) > 0. \tag{1.2}$$

Next we note that it is not hard to derive a “closed expression” for $\mathbf{E}_\mu(T)$. Fix μ . For $a \in A$, let $F(a) = \mu([1, a]) = \sum_{1 \leq d \leq a} \mu(d)$. We note recursion formulas for the expected travel time $T_a = \mathbf{E}_\mu(T \mid R_0 = a)$ when starting from position $a \in A$. It is not hard to

obtain (details are omitted due to space constraints)

$$\mathbf{E}_\mu(T) = \frac{1}{n} \cdot \sum_{1 \leq a_1 < \dots < a_\ell \leq n} \frac{\mu(a_2 - a_1) \cdots \mu(a_\ell - a_{\ell-1})}{F(a_1) \cdots F(a_\ell)}, \tag{1.3}$$

where the sum ranges over all $2^n - 1$ nonempty subintervals $[a_1, a_\ell]$ of $[1, n]$. By definition of $F(a)$, we see that $\mathbf{E}_\mu(T)$ is a rational function of $(\mu(1), \dots, \mu(n))$. By compactness, there is some μ that minimizes $\mathbf{E}_\mu(T)$. Unfortunately, there does not seem to be an obvious way to use (1.3) to gain information about the way $\mathbf{E}_\mu(T)$ depends on μ or what a distribution μ that minimizes $\mathbf{E}_\mu(T)$ looks like.

2. Upper bound

In this section, we establish upper bounds on $\mathbf{E}_\mu(T)$. We split the state space A and the set X of possible distances into “orders of magnitude”, arbitrarily choosing 2 as the base.¹ Let $L = \lfloor \log n \rfloor$, and define $I_i = [2^i, 2^{i+1})$, for $0 \leq i < L$, and $I_L = [2^L, n]$. Define

$$p_i = \sum_{d \in I_i} \mu(d), \text{ for } 0 \leq i \leq L.$$

Clearly, then, $p_0 + p_1 + \dots + p_L = 1$. To simplify notation, we do not exclude terms that mean p_i for $i < 0$ or $i > L$. Such terms are always meant to have value 0. Consider the process $R = (R_0, R_1, \dots)$. Assume $t \geq 1$ and $i \geq 1$. If $R_{t-1} \geq 2^i$ then all numbers $d \in I_{i-1}$ will be accepted as steps and lead to a progress of at least 2^{i-1} . Hence

$$\Pr(R_t \leq R_{t-1} - 2^{i-1} \mid R_{t-1} \geq 2^i) \geq p_{i-1}.$$

Further, if $R_{t-1} \in I_i$, we need to choose step sizes from I_{i-1} at most twice to get below 2^i . Since the expected waiting time for the random distances to hit I_{i-1} twice is $2/p_{i-1}$, the expected time process R remains in I_i is not larger than $2/p_{i-1}$.

Adding up over $1 \leq i \leq L$, the expected time process R spends in the interval $[2, a]$, where $a \in I_j$ is the starting position, is not larger than

$$\frac{2}{p_{j-1}} + \frac{2}{p_{j-2}} + \dots + \frac{2}{p_1} + \frac{2}{p_0}.$$

After the process has left $I_1 = [2, 3]$, it has reached position 0 or position 1, and the expected time before we hit 0 is not larger than $1/p_0 = 1/\mu(1)$. Thus, the expected number T_a of steps to get from $a \in I_j$ to 0 satisfies $T_a \leq \frac{2}{p_{j-1}} + \frac{2}{p_{j-2}} + \dots + \frac{2}{p_1} + \frac{3}{p_0}$. This implies the bound

$$\mathbf{E}_\mu(T) \leq \frac{2}{p_{L-1}} + \frac{2}{p_{L-2}} + \dots + \frac{2}{p_1} + \frac{3}{p_0},$$

for arbitrary μ . If we arrange that

$$p_0 = \dots = p_{L-1} = \frac{1}{L}, \tag{2.1}$$

we will have $T_a \leq (2j + 1)L \leq (2(\log a) + 1)(\log n) = O((\log a)(\log n)) = O((\log n)^2)$. Clearly, then, $\mathbf{E}_\mu(T) = O((\log n)^2)$ as well. The simplest distribution μ with (2.1) is the one that distributes the weight evenly on the powers of 2 below 2^L :

$$\mu_{\text{pow}2}(d) = \begin{cases} 1/L, & \text{if } d = 2^i, 0 \leq i < L, \\ 0, & \text{otherwise.} \end{cases}$$

¹log means “logarithm to the base 2” throughout.

Thus, $\mathbf{E}_{\mu_{\text{pow2}}}(T) = O((\log n)^2)$. The “harmonic distribution” defined by (1.1) satisfies $p_i \approx (\ln(2^{i+1}) - \ln(2^i))/H_n \approx \ln 2 / \ln(n) = 1/\log_2 n$, and we also get $T_a = O((\log a)(\log n))$ and $\mathbf{E}_{\mu_{\text{har}}}(T) = O((\log n)^2)$. More generally, all distributions μ with $p_0, \dots, p_{L-1} \geq \alpha/L$, where $\alpha > 0$ is constant, satisfy $\mathbf{E}_{\mu}(T) = O((\log n)^2)$.

3. Lower bound

We show, as the main result of this paper, that the upper bound of Section 2 is optimal up to a constant factor.

Theorem 3.1. $\mathbf{E}_{\mu}(T) = \Omega((\log n)^2)$ for all distributions μ .

This theorem is proved in the remainder of this section. The distribution μ is fixed from here on; we suppress μ in the notation. Recall that we may assume that $\mu(1) > 0$. We continue to use the intervals $I_0, I_1, I_2, \dots, I_L$ that partition $[1, n]$, as well as the probabilities $p_i, 0 \leq i \leq L$.

3.1. Intuition

The basic idea for the lower bound is the following. For the majority of the starting positions, the process has to traverse all intervals $I_{L-2}, I_{L-3}, \dots, I_1, I_0$. Consider an interval I_i . If the process reaches interval I_{i+1} , then afterwards steps of size 2^{i+2} and larger are rejected, and so do not help at all for crossing I_i . Steps of size from $I_{i+1}, I_i, I_{i-1}, I_{i-2}$ may be of significant help. Smaller step sizes will not help much. So, very roughly, the expected time to traverse interval I_i completely when starting in I_{i+1} will be bounded from below by

$$\frac{1}{p_{i+1} + p_i + p_{i-1} + p_{i-2}},$$

since $1/(p_{i+1} + p_i + p_{i-1} + p_{i-2})$ is the waiting time for the first step with a “significant” size to appear. If it were the case that there is a constant $\beta > 0$ with the property that for each $0 \leq i < L - 1$ the probability that interval I_{i+1} is visited is at least β then it would not be hard to show that the expected travel time is bounded below by

$$\sum_{1 \leq j < L/2} \frac{\beta}{p_{2j+1} + p_{2j} + p_{2j-1} + p_{2j-2}}. \quad (3.1)$$

(We picked out only the even $i = 2j$ to avoid double counting.) Now the sum of the denominators in the sum in (3.1) is at most 2, and the sum is minimal when all denominators are equal, so the sum is bounded below by $\beta \cdot (L/2) \cdot (L/2)/2 = \beta \cdot L^2/8$, hence the expected travel time would be $\Omega(L^2) = \Omega((\log n)^2)$.

It turns out that it is not straightforward to turn this informal argument into a rigorous proof. First, there are (somewhat strange) distributions μ for which it is not the case that each interval is visited with constant probability. (For example, let $\mu(d) = B^{d-1} \cdot (B-1)/(B^n-1)$, for a large base B like $B = n^3$. Then the “correct” jump directly to 0 has an overwhelming probability to be chosen first.²) Even for reasonable distributions μ , it may happen that some intervals or even blocks of intervals are jumped over with high probability. This means that the analysis of the cost of traversing I_i has to take into account that this traversal might happen in one big jump starting from an interval I_j with j much

²The authors thank Uri Feige for pointing this out.

larger than i . Second, in a formal argument, the contribution of the steps of size smaller than 2^{i-2} must be taken into account.

In the remainder of this section, we give a rigorous proof of the lower bound. For this, some machinery has to be developed. The crucial components are a reformulation of process R as another process, which as long as possible defers decisions about what the (randomly chosen) starting position is, and a potential function to measure how much progress the process has made in direction to its goal, namely reaching position 0.

3.2. Reformulation of the process

We change our point of view on the process R (with initial distribution uniform in $[1, n]$). The idea is that we do not have to fix the starting position right at the beginning, but rather make partial decisions on what the starting position is as the process advances. The information we hold on for step t is a random variable S_t , with the following interpretation: if $S_t > 0$ then R_t is uniformly distributed in $[1, S_t]$; if $S_t = 0$ then $R_t = 0$.

What properties should the random process $S = (S_0, S_1, \dots)$ on $[0, n]$ have to be a proper model of the Markov chain R from Section 1.2? We first give an intuitive description of process S , and later formally define the corresponding Markov chain. Clearly, $S_0 = n$: the starting position is uniformly distributed in $[1, n]$. Given $s = S_{t-1} \in [0, n]$, we choose a step length d from X , according to distribution μ . Then there are two cases.

Case 1: $d > s$. — If $s \geq 1$, this step cannot be used for any position in $[1, s]$, thus we reject it and let $S_t = s$. If $s = 0$, no further move is possible at all, and we also reject.

Case 2: $d \leq s$. — Then $s \geq 1$, and the token is at some position in $[1, s]$. What happens now depends on the position of the token relative to d , for which we only have a probability distribution. We distinguish three subcases:

- (i) The position of the token is larger than d . — This happens with probability $(s-d)/s$. In this case we “accept” the step, and now know that the token is in $[1, s-d]$, uniformly distributed; thus, we let $S_t = s-d$.
- (ii) The position of the token equals d . — This happens with probability $1/s$. In this case we “finish” the process, and let $S_t = 0$.
- (iii) The position of the token is smaller than d . — This happens with probability $\frac{d-1}{s}$. In this case we “reject” the step, and now know that the token is in $[1, d-1]$, uniformly distributed; thus, we let $S_t = d-1$.

Clearly, once state 0 is reached, all further steps are rejected via Case 1.

We formalize this idea by defining a new Markov chain $S = (S_0, S_1, \dots)$, as follows. The state space is $A = [0, n]$. For a state s' , we collect the total probability that we get from s to s' . If $s' > s$, this probability is 0; if $s' = s$, this probability is $\sum_{s < d \leq n} \mu(d) = 1 - F(s)$; if $s' = 0$, this probability is $\sum_{1 \leq d \leq s} \mu(d)/s = F(s)/s$; if $1 \leq s' < s$, this probability is $(\mu(s'+1) + \mu(s-s')) \cdot s'/s$, since d could be $s'+1$ or $s-s'$. Thus, we have the following transition probabilities:

$$p_{s,s'} = \begin{cases} F(s)/s & \text{if } s > s' = 0; \\ (\mu(s'+1) + \mu(s-s')) \cdot s'/s & \text{if } s > s' \geq 1; \\ 1 - F(s) & \text{if } s = s'. \end{cases}$$

Again, several initial distributions are possible for process S . The version with initial distribution with $\Pr(S_0 = n) = 1$ is meant to describe process R . Define the stopping time

$$T_S = \min\{t \mid S_t = 0\}.$$

We note that it is sufficient to analyze process S (with the standard initial distribution).

Lemma 3.2. $\mathbf{E}(T) = \mathbf{E}(T_S)$.

Proof. For $0 \leq s \leq n$, consider the version $R^{(s)}$ of process R induced by choosing the uniform distribution on $[1, s]$ (for $s \geq 1$) resp. $\{0\}$ (for $s = 0$) as the initial distribution. We let

$$A^{(s)} = \mathbf{E}(\min\{t \mid R_t^{(s)} = 0\}).$$

Clearly, $A^{(n)} = \mathbf{E}(T)$ and $A^{(0)} = 0$. We derive a recurrence for $(A^{(0)}, \dots, A^{(n)})$. Let $s \geq 1$, and assume the starting point R_0 is chosen uniformly at random from $[1, s]$. We carry out the first step of $R^{(s)}$, which starts with choosing d . The following situations may arise.

- (i) $d > s$. — This happens with probability $1 - F(s) < 1$. This distance will be rejected for all starting points in $[1, s]$, so the expected remaining travel time is $A^{(s)}$ again.
- (ii) $1 \leq d \leq s$. For each d , the probability for this to happen is $\mu(d)$. For the starting point R_0 there are three possibilities:
 - $R_0 \in [1, d - 1]$ (only possible if $d > 1$). — This happens with probability $\frac{d-1}{s}$. The remaining expected travel time is $A^{(d-1)}$.
 - $R_0 = d$. — This happens with probability $\frac{1}{s}$. The remaining travel time is 0.
 - $R_0 \in [d + 1, s]$ (only possible if $d < s$). — This happens with probability $\frac{s-d}{s}$. The remaining expected travel time in this case is $A^{(s-d)}$.

We obtain:

$$A^{(s)} = 1 + (1 - F(s))A^{(s)} + \sum_{1 \leq d \leq s} \mu(d) \left(\frac{d-1}{s} \cdot A^{(d-1)} + \frac{s-d}{s} \cdot A^{(s-d)} \right).$$

We rename $d - 1$ into s' in the first sum and $s - d$ into s' in the second sum and rearrange to obtain

$$A^{(s)} = \frac{1}{F(s)} \cdot \left(1 + \sum_{1 \leq s' < s} (\mu(s' + 1) + \mu(s - s')) \cdot (s'/s) \cdot A^{(s')} \right). \quad (3.2)$$

Next, we consider process S . For $0 \leq s \leq n$, let $S^{(s)}$ be the process obtained from S by choosing s as the starting point. Clearly, $S^{(0)}$ always sits in 0, and $S^{(n)}$ is just S . Let

$$B^{(s)} = \mathbf{E}(\min\{t \mid S_t^{(s)} = 0\}),$$

the expected number of steps process S needs to reach 0 when starting in s . Then $B^{(0)} = 0$ and $B^{(n)} = \mathbf{E}(T_S)$. We derive a recurrence for $(B^{(0)}, \dots, B^{(n)})$. Let $s \geq 1$. Carry out the first step of $S^{(s)}$, which leads to state s' . The following situations may arise.

- (i) $s = s' \geq 1$. — This occurs with probability $1 - F(s)$, and the expected remaining travel time is $B^{(s)}$ again.
- (ii) $s' = 0$. — In this case the expected remaining travel time is $B^{(0)} = 0$.
- (iii) $s > s' \geq 1$. — This occurs with probability $(\mu(s' + 1) + \mu(s - s')) \cdot s'/s$. The expected remaining travel time is $B^{(s')}$.

Summing up, we obtain

$$B^{(s)} = 1 + (1 - F(s))B^{(s)} + \sum_{1 \leq s' < s} (\mu(s' + 1) + \mu(s - s')) \cdot (s'/s) \cdot B^{(s')}.$$

Solving for $B^{(s)}$ yields:

$$B^{(s)} = \frac{1}{F(s)} \cdot \left(1 + \sum_{1 \leq s' < s} (\mu(s' + 1) + \mu(s - s')) \cdot (s'/s) \cdot B^{(s')} \right). \tag{3.3}$$

Since $A^{(0)} = 0 = B^{(0)}$ and the recurrences (3.2) and (3.3) are identical, we have $\mathbf{E}(T) = A^{(n)} = B^{(n)} = \mathbf{E}(T_S)$, as claimed. ■

3.3. Potential function: Definition and application

We introduce a potential function Φ on the state space $A = [0, n]$ to bound the progress of process S . Our main lemma states that for any $s \in A$, for a random transition from $S_i = s$ to S_{i+1} the expected loss in potential is at most constant (i.e., $\mathbf{E}(\Phi(S_{i+1}) - \Phi(S_i) \mid S_i = s) = O(1)$). This implies that $\mathbf{E}(T_S) = \Omega(\Phi(S_0))$. Since the potential function will satisfy $\Phi(S_0) = \Omega(\log^2 n)$, the lower bound follows.

We start by trying to give intuition for the definition. A rough approximation to the potential function we use would be the following: For interval I_i there is a term

$$\psi_i = \frac{1}{\sum_{0 \leq j \leq L} p_j \cdot c^{|j-i|}}, \tag{3.4}$$

for some constant c with $\frac{1}{2} < c < 1$, e. g., $c = 1/\sqrt{2}$. For later use we note that

$$\sum_{1 \leq i < L} \psi_i^{-1} = \sum_{1 \leq i < L} \sum_{0 \leq j \leq L} p_j \cdot c^{|j-i|} = \sum_{0 \leq j \leq L} p_j \sum_{1 \leq i < L} c^{|j-i|} = O(1), \tag{3.5}$$

since $\sum_{0 \leq j \leq L} p_j = 1$ and $\sum_{k \geq 0} c^k = \frac{1}{1-c}$. The term ψ_i tries to give a rough lower bound for the expected number of steps needed to cross I_i in the following sense: The summands $p_j \cdot c^{|j-i|}$ reflect the fact that step sizes that are close to I_i will be very helpful for crossing I_i , and step sizes far away from I_i might help a little in crossing I_i , but they do so only to a small extent ($j \ll i$) or with small probability ($j \gg i$). The idea is then to arrange that a state $s \in I_k$ has potential about

$$\Psi_k = \sum_{i \leq k} \psi_i. \tag{3.6}$$

It turns out that analyzing process S on the basis of a potential function that refers to the intervals I_i is possible but leads to messy calculations and numerous cases. The calculations become cleaner if one avoids the use of the intervals in the definition and in applying the potential function. The following definition derives from (3.4) and (3.6) by splitting up the summands ψ_i into contributions from all positions $a \in I_i$ and smoothing out the factors $c^{|j-i|} = 2^{|j-i|/2}$, for $a \in I_i$ and $d \in I_j$, into $2^{-|\log a - \log d|/2}$, which is $\sqrt{a/d}$ for $a \leq d$ and $\sqrt{d/a}$ for $d \leq a$. This leads to the following³. Assumption (1.2) guarantees that in the formulas to follow all denominators are nonzero.

Definition 3.3. For $1 \leq a \leq n$ let

$$\sigma_a = \sum_{1 \leq d \leq n} \mu(d) \cdot 2^{-|\log a - \log d|/2} = \sum_{1 \leq d \leq a} \mu(d) \sqrt{\frac{d}{a}} + \sum_{a < d \leq n} \mu(d) \sqrt{\frac{a}{d}}$$

³Whenever in the following we use letters a, b, d , the range $[1, n]$ is implicitly understood.

and $\varphi_a = 1/(a\sigma_a)$. For $0 \leq s \leq n$ define $\Phi(s) = \sum_{1 \leq a \leq s} \varphi_a$. The random variable Φ_t , $t = 0, 1, 2, \dots$, is defined as $\Phi_t = \Phi(S_t)$.

We note some easy observations and one fundamental fact about Φ_t , $t \geq 0$.

Lemma 3.4.

- (a) Φ_t , $t \geq 0$, is nonincreasing for t increasing.
- (b) $\Phi_t = 0 \Leftrightarrow S_t = 0$.
- (c) $\Phi_0 = \Omega((\log n)^2)$ (Φ_0 is a number that depends on n and μ).

Proof. (a) is clear since S_t , $t \geq 0$, is nonincreasing and the terms φ_a are positive. — (b) is obvious since $\Phi_t = 0$ if and only if $\Phi(S_t)$ is the empty sum, which is the case if and only if $S_t = 0$. — We prove (c). In this proof we use the intervals I_i and the probabilities p_i , $0 \leq i \leq L$, from Section 2. We use the notation $i(a) = \lfloor \log a \rfloor = \max\{i \mid 2^i \leq a\}$. We start with finding an upper bound for σ_a by grouping the summands in σ_a according to the intervals. Let $c = 1/\sqrt{2}$.

$$\begin{aligned} \sigma_a &= \sum_{1 \leq d \leq n} \mu(d) \cdot 2^{-|\log a - \log d|/2} \\ &\leq \sum_{j \leq i(a)} \sum_{d \in I_j} \mu(d) \cdot 2^{(j+1-i(a))/2} + \sum_{j > i(a)} \sum_{d \in I_j} \mu(d) \cdot 2^{(i(a)+1-j)/2} \\ &= \sum_{j \leq i(a)} p_j \cdot 2^{(j+1-i(a))/2} + \sum_{j > i(a)} p_j \cdot 2^{(i(a)+1-j)/2} = 2c \cdot \left(\sum_{0 \leq j \leq L} p_j \cdot c^{|j-i(a)|} \right). \end{aligned}$$

Hence

$$\sum_{a \in I_i} \varphi_a = \sum_{a \in I_i} \frac{1}{a\sigma_a} \geq \frac{2^i}{2c \cdot 2^{i+1} \cdot \left(\sum_{0 \leq j \leq L} p_j \cdot c^{|j-i|} \right)} = \frac{\psi_i}{4c},$$

with ψ_i from (3.4). Thus,

$$\Phi_0 \geq \sum_{0 \leq i < L} \frac{\psi_i}{4c}. \quad (3.7)$$

Let $u_i = 4c/\psi_i$ be the reciprocal of the summand for i in (3.7), $0 \leq i < L$. From (3.5) we read off that $\sum_{0 \leq i < L} u_i \leq k$, for some constant k . Now $\sum_{0 \leq i < L} \frac{1}{u_i}$ with $\sum_{0 \leq i < L} u_i \leq k$ is minimal if all u_i are equal to k/L . Together with (3.7) this entails $\Phi_0 \geq L \cdot (L/k) = L^2/k = \Omega((\log n)^2)$, which proves part (c) of Lemma 3.4. ■

The crucial step in the lower bound proof is to show that the progress made by process S in one step, measured in terms of the potential, is bounded:

Lemma 3.5 (Main Lemma). *There is a constant C such that for $0 \leq s \leq n$, we have $\mathbf{E}(\Phi_{t-1} - \Phi_t \mid S_{t-1} = s) \leq C$.*

The proof of Lemma 3.5 is the core of the analysis. It will be given in Section 3.4. To prove Theorem 3.1, we need the following lemma, which is stated and proved (as Lemma 12) in [2]. (It is a one-sided variant of Wald's identity.)

Lemma 3.6. *Let X_1, X_2, \dots denote random variables with bounded range, let $g > 0$ and let $T = \min\{t \mid X_1 + \dots + X_t \geq g\}$. If $\mathbf{E}(T) < \infty$ and $\mathbf{E}(X_t \mid T \geq t) \leq C$ for all $t \in \mathbf{N}$, then $\mathbf{E}(T) \geq g/C$.*

Proof of 3.1: Since $S_t = 0$ if and only if $\Phi_t = 0$ (Lemma 3.4(b)), the stopping time $T_\Phi = \min\{t \mid \Phi_t = 0\}$ of the potential reaching 0 satisfies $T_\Phi = T_S$. Thus, to prove Theorem 3.1, it is sufficient to show that $\mathbf{E}(T_\Phi) = \Omega((\log n)^2)$. For this, we let $X_t = \Phi_{t-1} - \Phi_t$, the progress made in step t in terms of the potential. By Lemma 3.5, $\mathbf{E}(X_t \mid S_{t-1} = s) \leq C$, for all $s \geq 1$, and hence

$$\mathbf{E}(X_t \mid T \geq t) = \mathbf{E}(X_t \mid \Phi(S_{t-1}) > 0) \leq C .$$

Observe that $X_1 + \dots + X_t = \Phi_0 - \Phi_t$ and hence $T_\Phi = \min\{t \mid X_1 + \dots + X_t \geq \Phi_0\}$. Applying Lemma 3.6, and combining with Lemma 3.4, we get that $\mathbf{E}(T_\Phi) \geq \Phi_0/C = \Omega((\log n)^2)$, which proves Theorem 3.1. ■

The only missing part to fill in is the proof of Lemma 3.5.

3.4. Proof of the Main Lemma (Lemma 3.5)

Fix $s \in [1, n]$, and assume $S_{t-1} = s$. Our aim is to show that the “expected potential loss” is constant, i. e., that

$$\mathbf{E}(\Phi_t - \Phi_{t-1} \mid S_{t-1} = s) = O(1).$$

Clearly, $\mathbf{E}(\Phi_t - \Phi_{t-1} \mid S_{t-1} = s) = \sum_{0 \leq x \leq s} \Delta(s, x)$, where

$$\Delta(s, x) = (\Phi(s) - \Phi(x)) \cdot \Pr(S_t = x \mid S_{t-1} = s). \tag{3.8}$$

We show that $\sum_{0 \leq x \leq s} \Delta(s, x)$ is bounded by a constant, by considering $\Delta(s, s)$, $\Delta(s, 0)$, and $\sum_{1 \leq x < s} \Delta(s, x)$ separately.

For $x = s$, the potential difference $\Phi(s) - \Phi(x)$ is 0, and thus

$$\Delta(s, s) = 0. \tag{3.9}$$

Bounding $\Delta(s, 0)$: According to the definition of the process S , a step from $S_{t-1} = s$ to $S_t = 0$ has probability $F(s)/s$. Since $\Phi(0) = 0$, the potential difference is $\Phi(s)$. Thus, we obtain

$$\begin{aligned} \Delta(s, 0) &= \frac{1}{s} \cdot \left(\sum_{d \leq s} \mu(d) \right) \cdot \left(\sum_{a \leq s} \varphi_a \right) = \frac{1}{s} \cdot \sum_{a \leq s} \frac{\sum_{d \leq s} \mu(d)}{\sum_{b \leq a} \mu(b)\sqrt{ab} + \sum_{a < b \leq n} \mu(b)a^{3/2}/\sqrt{b}} \\ &\leq \frac{1}{s} \cdot \sum_{a \leq s} \delta(a), \quad \text{where } \delta(a) = \frac{\sum_{b \leq s} \mu(b)}{\sum_{b \leq a} \mu(b)\sqrt{ab} + \sum_{a < b \leq s} \mu(b)a^{3/2}/\sqrt{b}} . \end{aligned}$$

We bound $\delta(a)$. For $b \leq a$ and $\mu(b) \neq 0$, the quotient of the summands in the numerator and denominator of $\delta(a)$ that correspond to b is $1/\sqrt{ab} \leq \sqrt{a}/a \leq \sqrt{s}/a$. For $a < b$ and $\mu(b) \neq 0$, the quotient is $\sqrt{b}/a^{3/2} \leq \sqrt{s}/a$. Thus, $\delta(a) \leq \frac{\sqrt{s}}{a}$. This implies (recall that $H_s = \sum_{a \leq d \leq s} \frac{1}{a}$):

$$\Delta(s, 0) \leq \frac{1}{s} \cdot \sum_{a \leq s} \sqrt{s}/a \leq \frac{H_s}{\sqrt{s}} \leq \frac{\ln(s) + 1}{\sqrt{s}} < 2. \tag{3.10}$$

Bounding $\sum_{1 \leq x < s} \Delta(s, x)$: Assume $1 \leq x < s$. According to the definition of the process S ,

$$\Pr(S_{t-1} = x \mid S_t = s) = \frac{x}{s} \cdot (\mu(x+1) + \mu(s-x)).$$

The potential difference is $\Phi(s) - \Phi(x) = \sum_{x < a \leq s} \varphi_a$. Thus we have

$$\sum_{1 \leq x < s} \Delta(s, x) = \sum_{1 \leq x < s} \sum_{x < a \leq s} \varphi_a \cdot \frac{x}{s} \cdot (\mu(x+1) + \mu(s-x)) = \frac{1}{s} \cdot \sum_{1 < a \leq s} (\lambda_a + \gamma_a), \quad (3.11)$$

where $\lambda_a = \varphi_a \cdot \sum_{1 \leq x < a} \mu(x+1)x$ and $\gamma_a = \varphi_a \cdot \sum_{1 \leq x < a} \mu(s-x)x$. We bound λ_a and γ_a separately. Observe first that

$$\begin{aligned} \lambda_a &= \varphi_a \cdot \sum_{2 \leq x \leq a} \mu(x)(x-1) \\ &\leq \frac{\sum_{1 \leq x \leq a} \mu(x)(x-1)}{\sum_{1 \leq b \leq a} \mu(b) \cdot \sqrt{ab} + \sum_{a < b \leq n} \mu(b)a^{3/2}/\sqrt{b}} \leq \frac{\sum_{1 \leq b \leq a} \mu(b)(b-1)}{\sum_{1 \leq b \leq a} \mu(b)\sqrt{ab}}. \end{aligned} \quad (3.12)$$

(We used the definition of φ_a , and omitted some summands in the denominator.) Recall that $\mu(1) > 0$, so the denominator is not zero. For each $b \leq a$ we clearly have $\mu(b)(b-1) \leq \mu(b)\sqrt{ab}$, thus the sum in the numerator in (3.12) is smaller than the sum in the denominator, and we get $\lambda_a < 1$.

Next, we bound γ_a for $a \leq s$:

$$\begin{aligned} \gamma_a &= \varphi_a \cdot \sum_{1 \leq x < a} \mu(s-x)x = \varphi_a \cdot \sum_{s-a < x < s} \mu(x)(s-x) \\ &= \frac{\sum_{s-a < x \leq a} \mu(x)(s-x) + \sum_{\max\{a, s-a\} < x < s} \mu(x)(s-x)}{\sum_{1 \leq b \leq a} \mu(b)\sqrt{ab} + \sum_{a < b \leq n} \mu(b)a^{3/2}/\sqrt{b}}. \end{aligned}$$

The denominator is not zero because $\mu(1) > 0$. Hence, if $\mu(x) = 0$ for all $s-a < x < s$, then $\gamma_a = 0$. Otherwise, by omitting some of the summands in the denominator we obtain

$$\gamma_a \leq \frac{\sum_{s-a < b \leq a} \mu(b)(s-b) + \sum_{\max\{a, s-a\} < b < s} \mu(b)(s-b)}{\sum_{s-a < b \leq a} \mu(b)\sqrt{ab} + \sum_{\max\{a, s-a\} < b < s} \mu(b)a^{3/2}/\sqrt{b}}$$

(If $a \leq s/2$, the first sum in both numerator and denominator is empty.) Now consider the quotient of the summands for each b with $\mu(b) > 0$. For $s-a < b \leq a$, this quotient is

$$\frac{\mu(b)(s-b)}{\mu(b)\sqrt{ab}} \leq \frac{a-1}{\sqrt{a} \cdot (s-a+1)} < \sqrt{\frac{a}{s-a+1}} \leq \sqrt{\frac{s}{s-a+1}}.$$

For $\max\{a, s-a\} < b < s$, the quotient of the corresponding summands is

$$\frac{\mu(b)(s-b)}{\mu(b)a^{3/2}/\sqrt{b}} \leq \frac{\min\{a, s-a\} \cdot \sqrt{b}}{a^{3/2}} \leq \frac{a \cdot \sqrt{s}}{a^{3/2}} = \sqrt{\frac{s}{a}}.$$

Hence, $\gamma_a \leq \sqrt{s/(s-a+1)} + \sqrt{s/a}$. Plugging this bound on γ_a and the bound $\lambda_a < 1$ into (3.11), and using that

$$\sum_{1 \leq a \leq s} \frac{1}{\sqrt{a}} = 1 + \sum_{2 \leq a \leq s} \frac{1}{\sqrt{a}} < 1 + \int_1^s \frac{dx}{\sqrt{x}} = 1 + [2\sqrt{x}]_1^s = 1 + 2\sqrt{s} - 2 < 2\sqrt{s},$$

we obtain

$$\begin{aligned} \sum_{1 \leq x < s} \Delta(s, x) &< \frac{1}{s} \cdot \sum_{1 < a \leq s} \left(1 + \sqrt{\frac{s}{a}} + \sqrt{\frac{s}{s-a+1}} \right) \\ &< 1 + \frac{1}{\sqrt{s}} \left(\sum_{1 < a \leq s} \sqrt{\frac{1}{a}} + \sum_{1 \leq a < s} \sqrt{\frac{1}{a}} \right) < 1 + \frac{2}{\sqrt{s}} \sum_{1 \leq a \leq s} \frac{1}{\sqrt{a}} < 1 + \frac{2}{\sqrt{s}} \cdot 2\sqrt{s} = 5. \end{aligned} \quad (3.13)$$

Summing up the bounds from (3.9), (3.10), and (3.13), we obtain

$$\mathbf{E}(\Phi_t - \Phi_{t-1} \mid S_{t-1} = s) \leq \Delta(s, 0) + \sum_{1 \leq x < s} \Delta(s, x) + \Delta(s, s) < 2 + 5 + 0 = 7.$$

Thus Lemma 3.5 is proved. ■

4. Open problems

1. We conjecture that the method can be adapted to the continuous case to prove a lower bound of $\Omega((\log(1/\varepsilon))^2)$ for approximating the minimum of some unimodal function f by a scale-invariant search strategy (see Section 1.1).

2. It is an open problem whether our method can be used to prove a lower bound of $\Omega((\log n)^2)$ for finding the minimum of an arbitrary unimodal function $f: \{0, \dots, n\} \rightarrow \mathbb{R}$ by a scale invariant search strategy.

Acknowledgement

The authors thank two anonymous referees for their careful reading of the manuscript and for providing several helpful comments.

References

- [1] Droste, S., Jansen, T., and Wegener, I., Upper and lower bounds for randomized search heuristics in black-box optimization, *Theory Comput. Syst.* **39**(4) 525–544 (2006).
- [2] Jägerskupper, J., Algorithmic analysis of a basic evolutionary algorithm for continuous optimization, *Theor. Comput. Sci.* **379**(3) 329–347 (2007).
- [3] Kiefer, J., Sequential minimal search for a maximum, *Proc. Amer. Math. Soc.* **4** 502–506 (1953).
- [4] Rowe, J.E., and Hidović, D., An evolution strategy using a continuous version of the Gray-code neighbourhood distribution, in: K. Deb *et al.*, *Eds.*, Proc. GECCO 2004, Part 1, LNCS Vol. 3102, Springer-Verlag, pp. 725–736.