

Tight Bounds for Blind Search on the Integers and the Reals

MARTIN DIETZFELBINGER,^{1†} JONATHAN E. ROWE,²
INGO WEGENER^{3‡} and PHILIPP WOELFEL^{4§¶}

¹Fakultät für Informatik und Automatisierung, Technische Universität Ilmenau, 98684 Ilmenau, Germany
(e-mail: martin.dietzfelbinger@tu-ilmenau.de)

²School of Computer Science, University of Birmingham, Birmingham B15 2TT, UK
(email: J.E.Rowe@cs.bham.ac.uk)

³Fakultät für Informatik, Technische Universität Dortmund, 44221 Dortmund, Germany

⁴Department of Computer Science, University of Calgary, Calgary, Alberta T2N 1N4, Canada
(email: woelfel@cpsc.ucalgary.ca)

Received 5 August 2009; first published online 18 December 2009

It was with great sorrow that we learned of the death of Ingo Wegener on November 26, 2008. Ingo brought the rest of us together to work on the problem treated here. He is greatly missed by us all. This paper is dedicated to his memory.

We analyse a simple random process in which a token is moved in the interval $A = \{0, \dots, n\}$. Fix a probability distribution μ over $D = \{1, \dots, n\}$. Initially, the token is placed in a random position in A . In round t , a random *step size* d is chosen according to μ . If the token is in position $x \geq d$, then it is moved to position $x - d$. Otherwise it stays put. Let T_X be the number of rounds until the token reaches position 0. We show tight bounds for the expectation $\mathbf{E}_\mu(T_X)$ of T_X for varying distributions μ . More precisely, we show that $\min_\mu \{\mathbf{E}_\mu(T_X)\} = \Theta((\log n)^2)$. The same bounds are proved for the analogous continuous process, where step sizes and token positions are real values in $[0, n + 1)$, and one measures the time until the token has reached a point in $[0, 1)$. For the proofs, 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.

[†] Work by this author was done in part while visiting ETH Zürich, Switzerland.

[‡] Supported in part by the DFG collaborative research project SFB 531.

[§] Supported in part by NSERC, and while working at the University of Toronto by DFG grant WO 1232/1-1 and by SUN Microsystems.

[¶] A preliminary version of this paper, not including the material in Section 4, appeared in the proceedings of STACS 2008. Joint work on this topic was initiated during Dagstuhl Seminar No. 06111 on *Complexity of Boolean Functions* (2006) and continued during the Oberwolfach workshops *Combinatorics, Probability and Computing* (2006) and *Combinatorics and Probability* (2009).

1. Introduction

1.1. The discrete token process

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

- Initially, place the token in some position chosen uniformly at random from $A - \{0\}$.
- In round t : The token is at position $x \in A$. Choose an element d from D at random, according to μ . If $d \leq x$, move the token to position $x - d$, otherwise leave it where it is.

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

We consider the process $X = (X_0, X_1, \dots)$, where X_t denotes the position of the token after round t , and let $T_X = \min\{t \mid X_t = 0\}$ be the number of rounds needed until position 0 is reached. A basic performance parameter for the process is $\mathbf{E}_\mu(T_X)$, the expectation of T_X for some probability distribution μ . As μ varies, the value $\mathbf{E}_\mu(T_X)$ will vary. The probability distribution μ may be regarded as a strategy. We ask: How should μ be chosen so that $\mathbf{E}_\mu(T_X)$ is as small as possible?

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

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

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

1.2. Motivation and background: blind optimization strategies over the reals

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* [3]. 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 step size $d \in (0, 1]$ 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 [6]. One starts by specifying a minimum perturbation size ε . Then one chooses the probability density function

$$h(t) = \begin{cases} 1/pt & \text{if } \varepsilon \leq t \leq 1, \\ 0 & \text{otherwise,} \end{cases} \quad (1.2)$$

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 [6].) From here on, we focus on the simple case where f is *unimodal*, meaning that it is strictly decreasing in $[0, x^*]$ and strictly increasing in $[x^*, 1]$, where x^* is the unknown minimum point.

Remark. 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 [5], ‘Fibonacci search’ was proposed and analysed, which for a given tolerance ε uses the optimal number of steps in a very strong sense.

The ‘blind search’ strategy from [6] 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^* is $\tau \geq 2\varepsilon$ then every step size 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^* 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^* has been halved is $2p/\ln 2$. We regard the algorithm as successful if the current point has distance smaller than 2ε from x^* . 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 first 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 $x^* = 0$. In this case, the search process is very simple: the actual values of f are irrelevant; going from x to $x + d$ is never an improvement. Actually, the search process is fully described by the simple random process from Section 1.1. How long does it take to reach the optimal point 0, for a distribution 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 leading to the bound $O(\log(1/\varepsilon)^2)$ in the continuous case. The first main result of this paper is that the bound for the discrete case is optimal.

Second, we show that in the continuous case the probability density function (1.2) is (asymptotically) optimal for approximating the minimum of $f : [0, 1] \rightarrow \mathbb{R}$, $f(x) = x$, up to an error of 2ε . We assume ε to be given and let $n = 1/2\varepsilon$ (assuming this to be an integer). Blind search for the minimum until the error is smaller than 2ε is equivalent to the following scaled token process. The start position is uniform on $(0, n + 1)$. In each step a distance d is chosen at random according to some fixed probability distribution μ on $(0, n + 1)$. (It is not necessary that μ has a density.) If $d \leq x$, where x is the current token position, then the token moves to position $x - d$; otherwise it stays put. The process is considered finished when the token has entered the interval $(0, 2\varepsilon)$. As the second main result, we prove that for all μ the expected number of steps one needs until the token reaches a point smaller than 1 is $\Omega((\log n)^2)$. This translates into a lower bound of $\Omega((\log(1/\varepsilon))^2)$ for the 2ε -approximation task by blind search. The result is proved by generalizing the potential function method used for the discrete token process.

1.3. Formalization of the discrete process as a Markov chain

For the sake of simplicity, from now on we let $\langle a, b \rangle$ denote the discrete interval $\{a, \dots, b\}$ if a and b are integers. Given a probability distribution μ on $\langle 1, n \rangle$, the Markov chain $X = (X_0, X_1, \dots)$ is defined over the state space $A = \langle 0, n \rangle$ as follows. The start state X_0 is uniformly distributed in $A - \{0\} = \langle 1, n \rangle$, and the transition probabilities are defined by

$$\Pr(X_t = x' \mid X_{t-1} = x) = \begin{cases} \mu(x - x') & \text{for } x' < x, \\ 1 - \sum_{1 \leq d \leq x} \mu(d) & \text{for } x' = x, \\ 0 & \text{for } x' > x. \end{cases}$$

Clearly, 0 is an absorbing state. We define the random variable $T_X = \min\{t \mid X_t = 0\}$. Let us write $\mathbf{E}_\mu(T_X)$ for the expectation of T_X for a particular probability distribution μ . We study $\mathbf{E}_\mu(T_X)$ for arbitrary μ . In particular, we wish to identify distributions μ that make $\mathbf{E}_\mu(T_X)$ as small as possible (up to constant factors, where n is growing).

Observation 1.1. *If $\mu(1) = 0$ then $\mathbf{E}_\mu(T_X) = \infty$.*

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

$$\mu(1) > 0. \tag{1.3}$$

It is not hard to derive a ‘closed expression’ for $\mathbf{E}_\mu(T_X)$. For $x \in A$, let $F(x) = \mu(\langle 1, x \rangle) = \sum_{1 \leq d \leq x} \mu(d)$.

Proposition 1.2.

$$\mathbf{E}_\mu(T_X) = \frac{1}{n} \cdot \sum_{1 \leq x_1 < \dots < x_\ell \leq n} \frac{\mu(x_2 - x_1) \cdots \mu(x_\ell - x_{\ell-1})}{F(x_1) \cdots F(x_\ell)}, \tag{1.4}$$

where the sum ranges over all $2^n - 1$ non-empty subsets $\{x_1, \dots, x_\ell\}$ of $\langle 1, n \rangle$.

Proof. Because of (1.3), we have $F(x) > 0$ for all x . Define the expected travel time when starting from position $x \in A$ as $T_x = \mathbf{E}_\mu(T \mid X_0 = x)$. Clearly, $T_0 = 0$. Now assume $X_0 = x \geq 1$. We choose d at random from D , and perform one step. There are two cases: with probability $1 - F(x)$, we have $d > x$, hence $X_1 = x$, and the remaining time is T_x again; otherwise, the remaining travel time is T_{x-d} . Thus,

$$T_x = 1 + (1 - F(x)) \cdot T_x + \sum_{1 \leq d \leq x} \mu(d) \cdot T_{x-d}.$$

Solving for T_x , we get

$$T_x = \frac{1}{F(x)} \cdot \left(1 + \sum_{1 \leq d \leq x} \mu(d) \cdot T_{x-d} \right), \quad \text{for } 1 \leq x \leq n. \tag{1.5}$$

For example, we may now calculate:

$$\begin{aligned} T_1 &= \frac{1}{\mu(1)} = \frac{1}{F(1)}, \\ T_2 &= \frac{1}{F(2)} + \frac{\mu(1)}{F(1)F(2)}, \\ T_3 &= \frac{1}{F(3)} + \frac{\mu(2)}{F(1)F(3)} + \frac{\mu(1)}{F(2)F(3)} + \frac{\mu(1)^2}{F(1)F(2)F(3)}, \end{aligned}$$

and so on. The number of terms in the sum for T_x doubles as x increases by 1. Using (1.5) in an induction on x , it is not hard to prove that

$$T_x = \sum_{1 \leq x_1 < \dots < x_\ell = x} \frac{\mu(x_2 - x_1) \cdots \mu(x_\ell - x_{\ell-1})}{F(x_1) \cdots F(x_\ell)}, \tag{1.6}$$

where the sum ranges over all 2^{x-1} subsets $\{x_1, \dots, x_{\ell-1}\}$ of $\{1, \dots, x-1\}$. By averaging (1.6) over the n possible starting positions, we obtain (1.4). \square

We may conclude from (1.4) that $\mathbf{E}_\mu(T_X)$ is a rational function of $(\mu(1), \dots, \mu(n))$. By compactness, there is some μ that minimizes $\mathbf{E}_\mu(T_X)$. Unfortunately, there does not seem to be an obvious way to use Proposition 1.2 to gain information about the way in which $\mathbf{E}_\mu(T_X)$ depends on μ or what a distribution μ that minimizes $\mathbf{E}_\mu(T_X)$ looks like.

2. Upper bound

In this section, we establish upper bounds on $\mathbf{E}_\mu(T_X)$. We split the state space A and the set D of possible step sizes into ‘orders of magnitude’, arbitrarily choosing 2 as the base.¹ Let $L = \lceil \log n \rceil$, and define $I_i = [2^i, 2^{i+1})$, for $0 \leq i < L$, and $I_L = \langle 2^L, n \rangle$. Define

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

¹ Throughout, log means ‘logarithm to the base 2’.

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 $X = (X_0, X_1, \dots)$. Assume $t \geq 1$ and $i \geq 1$. If $X_{t-1} \geq 2^i$ then all numbers $d \in I_{i-1}$ will be accepted as steps and lead to a ‘progress’ $X_{t-1} - X_t$ of at least 2^{i-1} . Hence

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

Further, if $X_{t-1} \in I_i$, 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 step sizes to hit I_{i-1} twice is $2/p_{i-1}$, the expected time process X remains in I_i is not larger than $2/p_{i-1}$.

Adding up over $1 \leq i \leq L$, the expected time process X spends in the interval $\langle 2, x \rangle$, where $x \in I_j$ 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 $I_1 = \langle 2, 3 \rangle$, it has reached position 0 or position 1, and the expected time before it hits 0 is not larger than $1/p_0 = 1/\mu(1)$. Thus, the expected number T_x of steps to get from $x \in I_j$ to 0 satisfies $T_x \leq \frac{2}{p_{j-1}} + \frac{2}{p_{j-2}} + \dots + \frac{2}{p_1} + \frac{3}{p_0}$. This implies the bound

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

for arbitrary μ . If μ has the property that

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

for some constant $\alpha > 0$, we will have $T_x \leq (2j + 1)(L/\alpha) = O((\log x)(\log n)) = O((\log n)^2)$. Clearly, then, $\mathbf{E}_\mu(T_X) = O((\log n)^2)$ 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 :

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

Thus, $\mathbf{E}_{\mu_{\text{pow}2}}(T_X) = O((\log n)^2)$. 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 $\mathbf{E}_{\mu_{\text{har}}}(T_X) = O((\log n)^2)$.

3. A lower bound for the discrete process

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

Theorem 3.1. $\mathbf{E}_\mu(T_X) = \Omega((\log n)^2)$ 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 proof 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

$$\frac{1}{p_{i+1} + p_i + p_{i-1} + p_{i-2}}, \tag{3.1}$$

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 from below by

$$\sum_{1 \leq j < L/2} \frac{\beta}{p_{2j+1} + p_{2j} + p_{2j-1} + p_{2j-2}}. \tag{3.2}$$

(We picked out only the even $i = 2j$ to avoid double-counting.) Now the sum of the denominators in the sum in (3.2) 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 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 X as another ‘interval process’, which, for 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.

² The authors thank Uri Feige for pointing this out.

3.2. The interval process

We change our point of view on the process X . 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 Y_t , with the following interpretation: if $Y_t > 0$ then X_t is uniformly distributed in $\langle 1, Y_t \rangle$; if $Y_t = 0$ then $X_t = 0$.

What properties should the random process $Y = (Y_0, Y_1, \dots)$ on $\langle 0, n \rangle$ have to be a proper model of the Markov chain X from Section 1.3? Clearly, $Y_0 = n$: the starting position is uniformly distributed in $\langle 1, n \rangle$. Given $y = Y_{t-1} \in \langle 0, n \rangle$, we choose a step length d from D , according to distribution μ . Then there are two cases.

Case 1: $d > y$. If $y \geq 1$, this step cannot be used for any position in $\langle 1, y \rangle$, thus we ‘reject’ it and let $Y_t = y$. If $y = 0$, no further move is possible at all, and we also ‘reject’ and let $Y_t = y = 0$.

Case 2: $d \leq y$. Then $y \geq 1$, and the token is at some position in $\langle 1, y \rangle$. 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.

Case 2(i): The position of the token is larger than d . This happens with probability $(y - d)/y$. In this case we ‘accept’ the step, and now know that the token is in $\langle 1, y - d \rangle$, uniformly distributed; thus, we let $Y_t = y - d$.

Case 2(ii): The position of the token equals d . This happens with probability $1/y$. In this case we ‘finish’ the process, and let $Y_t = 0$.

Case 2(iii): The position of the token is smaller than d . This happens with probability $\frac{d-1}{y}$. In this case we ‘reject’ the step, and now know that the token is in $\langle 1, d - 1 \rangle$, uniformly distributed; thus, we let $Y_t = d - 1$.

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

To summarize, the Markov chain $Y = (Y_0, Y_1, \dots)$ is defined as follows. The state space is $A = \langle 0, n \rangle$ and the start state is $Y_0 = n$. Given a state $y = Y_{t-1}$, $t > 0$, the state Y_t is determined by the following random experiment.

- (1) Pick $d \in D$ according to distribution μ .
- (2) If $d > y$, then $Y_t = y$.
- (3) If $d \leq y$, then $Y_t \in \{y - d, 0, d - 1\}$, distributed as follows:

$$Y_t = \begin{cases} y - d & \text{with probability } \frac{y-d}{y}, \\ 0 & \text{with probability } \frac{1}{y}, \\ d - 1 & \text{with probability } \frac{d-1}{y}. \end{cases}$$

It is not hard to write down transition probabilities of the Markov chain Y :

$$\Pr(Y_t = y' \mid Y_{t-1} = y) = \begin{cases} F(y)/y & \text{if } y > y' = 0, \\ (\mu(y' + 1) + \mu(y - y')) \cdot y'/y & \text{if } y > y' \geq 1, \\ 1 - F(y) & \text{if } y = y'. \end{cases} \quad (3.3)$$

We are interested in the random variable $T_Y = \min\{t \mid Y_t = 0\}$, which counts the number of steps until Y reaches its absorbing state 0.

The idea to derive the interval process in the way described above is not new. For example, Aspnes, Diamadi and Shah [1, Section 4.2.2], analyse a so-called *aggregate chain* that is derived from a Markov chain in the same way as we derived our interval process from the token process. They prove that both processes have the same expected running time. In our case, essentially the same proof yields the following lemma.

Lemma 3.2. $E(T_X) = E(T_Y)$.

For completeness, we provide a full proof of this lemma in Section 4.2, where we deal with the continuous case.

3.3. Potential function: definition and application

We introduce a potential function Φ on the state space $A = \langle 0, n \rangle$ to bound the progress of process Y . Our main lemma states that for any $y \in A$, for a random transition from $Y_{t-1} = y$ to Y_t , the expected loss in potential is at most constant (i.e., $E(\Phi(Y_{t-1}) - \Phi(Y_t) \mid Y_{t-1} = y) = O(1)$). We will see that this implies that $E(T_Y) = \Omega(\Phi(Y_0))$. Since the potential function will satisfy $\Phi(Y_0) = \Omega((\log n)^2)$, 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 $\sum_{k \geq 0} c^k = \frac{1}{1-c}$. Extending the idea from (3.1), 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 $y \in I_k$ has potential about

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

It turns out that analysing process Y 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 following.³ Assumption (1.3) guarantees that in the formulas to follow all denominators are non-zero.

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

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

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

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

Lemma 3.3.

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

Proof. (a) is clear since Y_t , $t \geq 0$, is non-increasing and the terms φ_a are positive. (b) is obvious since $\Phi_t = 0$ if and only if $\Phi(Y_t)$ is the empty sum, which is the case if and only if $Y_t = 0$. 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 $c = 1/\sqrt{2}$. Then

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

Hence

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

with ψ_i from (3.4). Thus,

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

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

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

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 3.3. \square

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

Lemma 3.4 (Main Lemma). *There is a constant C such that, for $0 \leq y \leq n$, we have $\mathbf{E}(\Phi_{t-1} - \Phi_t \mid Y_{t-1} = y) \leq C$.*

The proof of Lemma 3.4 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 [4]. (It is a one-sided variant of Wald’s identity.)

Lemma 3.5. *Let R_1, R_2, \dots denote random variables with bounded range, let $g > 0$, and let $S = \min\{t \mid R_1 + \dots + R_t \geq g\}$. If $\mathbf{E}(S) < \infty$ and $\mathbf{E}(R_t \mid S \geq t) \leq C$ for all $t \in \mathbb{N}$, then $\mathbf{E}(S) \geq g/C$.*

Proof of Theorem 3.1. Since $Y_t = 0$ if and only if $\Phi_t = 0$ (Lemma 3.3(b)), the step count $T_\Phi = \min\{t \mid \Phi_t = 0\}$ for the potential reaching 0 satisfies $T_\Phi = T_Y$. Thus, to prove Theorem 3.1, it is sufficient to show that $\mathbf{E}(T_\Phi) = \Omega((\log n)^2)$. For this, we let $R_t = \Phi_{t-1} - \Phi_t$, the progress made in step t in terms of the potential. By Lemma 3.4, $\mathbf{E}(R_t \mid Y_{t-1} = y) \leq C$, for all $y \geq 1$, and hence

$$\mathbf{E}(R_t \mid T_Y \geq t) = \mathbf{E}(R_t \mid Y_{t-1} > 0) \leq C.$$

Observe that $R_1 + \dots + R_t = \Phi_0 - \Phi_t$ and hence $T_\Phi = \min\{t \mid R_1 + \dots + R_t \geq \Phi_0\}$. Applying Lemma 3.5, and combining with Lemma 3.3, we get that $\mathbf{E}(T_\Phi) \geq \Phi_0/C = \Omega((\log n)^2)$, which proves Theorem 3.1. \square

The only missing part to fill in is the proof of Lemma 3.4.

3.4. Proof of the Main Lemma (Lemma 3.4)

Fix $y \in \langle 1, n \rangle$, and assume $Y_{t-1} = y$. Our aim is to show that the ‘expected potential loss’ is constant, i.e., that

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

Clearly, $\mathbf{E}(\Phi_{t-1} - \Phi_t \mid Y_{t-1} = y) = \sum_{0 \leq y' \leq y} \Delta(y, y')$, where

$$\Delta(y, y') = (\Phi(y) - \Phi(y')) \cdot \Pr(Y_t = y' \mid Y_{t-1} = y). \tag{3.8}$$

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

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

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

Bounding $\Delta(y, \mathbf{0})$. According to (3.3), a step from $Y_{t-1} = y$ to $Y_t = 0$ has probability $F(y)/y$. Since $\Phi(0) = 0$, the potential difference is $\Phi(y)$. Thus, we obtain

$$\begin{aligned} \Delta(y, \mathbf{0}) &= \frac{1}{y} \cdot \left(\sum_{d \leq y} \mu(d) \right) \cdot \left(\sum_{a \leq y} \varphi_a \right) \\ &= \frac{1}{y} \cdot \sum_{a \leq y} \frac{\sum_{d \leq y} \mu(d)}{\sum_{b \leq a} \mu(b) \sqrt{ab} + \sum_{a < b \leq n} \mu(b) a^{3/2} / \sqrt{b}} \\ &\leq \frac{1}{y} \cdot \sum_{a \leq y} \delta(a), \quad \text{where } \delta(a) = \frac{\sum_{b \leq y} \mu(b)}{\sum_{b \leq a} \mu(b) \sqrt{ab} + \sum_{a < b \leq y} \mu(b) a^{3/2} / \sqrt{b}}. \end{aligned} \tag{3.10}$$

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 $1/\sqrt{ab} \leq \sqrt{a}/a \leq \sqrt{y}/a$. For $a < b$ and $\mu(b) \neq 0$, the quotient is $\sqrt{b}/a^{3/2} \leq \sqrt{y}/a$. Thus, factoring out \sqrt{y}/a from the sum in the denominator of $\delta(a)$, and using $\mu(1) > 0$, we obtain $\delta(a) \leq \sqrt{y}/a$. This implies (recall that $H_y = \sum_{1 \leq a \leq y} \frac{1}{a}$):

$$\Delta(y, \mathbf{0}) \leq \frac{1}{y} \cdot \sum_{a \leq y} \sqrt{y}/a \leq \frac{H_y}{\sqrt{y}} \leq \frac{\ln(y) + 1}{\sqrt{y}} < 2. \tag{3.11}$$

Bounding $\sum_{1 \leq y' < y} \Delta(y, y')$. Assume $1 \leq y' < y$. According to (3.3),

$$\Pr(Y_t = y' \mid Y_{t-1} = y) = \frac{y'}{y} \cdot (\mu(y' + 1) + \mu(y - y')).$$

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

$$\begin{aligned} \sum_{1 \leq y' < y} \Delta(y, y') &= \sum_{1 \leq y' < y} \sum_{y' < a \leq y} \varphi_a \cdot \frac{y'}{y} \cdot (\mu(y' + 1) + \mu(y - y')) \\ &= \sum_{1 < a \leq y} \sum_{1 \leq y' < a} \varphi_a \cdot \frac{y'}{y} \cdot (\mu(y' + 1) + \mu(y - y')) \\ &= \frac{1}{y} \cdot \sum_{1 < a \leq y} (\lambda_a + \gamma_a), \end{aligned} \tag{3.12}$$

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

$$\begin{aligned} \lambda_a &= \varphi_a \cdot \sum_{2 \leq y' \leq a} \mu(y')(y' - 1) \\ &\leq \frac{\sum_{1 \leq y' \leq a} \mu(y')(y' - 1)}{\sum_{1 \leq b \leq a} \mu(b) \cdot \sqrt{ab} + \sum_{a < b \leq n} \mu(b) a^{3/2} / \sqrt{b}} \leq \frac{\sum_{1 \leq b \leq a} \mu(b)(b - 1)}{\sum_{1 \leq b \leq a} \mu(b) \sqrt{ab}}. \end{aligned} \tag{3.13}$$

(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 (3.13) is smaller than the sum in the denominator, and we get $\lambda_a < 1$.

Next, we bound γ_a for $a \leq y$:

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

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

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

(If $a \leq y/2$, 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 $y - a < b \leq a$, this quotient is

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

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

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

Hence, $\gamma_a \leq \sqrt{y/(y - a + 1)} + \sqrt{y/a}$. Plugging this bound on γ_a and the bound $\lambda_a < 1$ into (3.12), and using that

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

we obtain

$$\begin{aligned} \sum_{1 \leq y' < y} \Delta(y, y') &< \frac{1}{y} \cdot \sum_{1 < a \leq y} \left(1 + \sqrt{\frac{y}{a}} + \sqrt{\frac{y}{y - a + 1}} \right) \\ &< 1 + \frac{1}{\sqrt{y}} \left(\sum_{1 < a \leq y} \sqrt{\frac{1}{a}} + \sum_{1 \leq a < y} \sqrt{\frac{1}{a}} \right) < 1 + \frac{2}{\sqrt{y}} \sum_{1 \leq a \leq y} \frac{1}{\sqrt{a}} < 1 + \frac{2}{\sqrt{y}} \cdot 2\sqrt{y} = 5. \end{aligned} \tag{3.14}$$

Summing up the bounds from (3.9), (3.11) and (3.14), we obtain

$$\mathbf{E}(\Phi_{t-1} - \Phi_t \mid Y_{t-1} = y) \leq \Delta(y, 0) + \sum_{1 \leq y' < y} \Delta(y, y') + \Delta(y, y) < 2 + 5 + 0 = 7.$$

Thus Lemma 3.4 is proved. □

4. The continuous process

We now analyse a continuous version of the process, where token positions and step sizes are real numbers. All definitions and proofs to follow are generalizations of the definitions and proofs for the discrete process.

We use the standard notation $([a, b], [a, b), \text{etc.})$ to denote intervals over the reals.

The continuous random process $X = (X_0, X_1, \dots)$ is defined as follows. Let μ be some probability distribution over $(0, n + 1)$. At the beginning, the token is placed in a point X_0 chosen uniformly at random from $[1, n + 1)$. Assume that after $t - 1$ steps the token is in position $X_{t-1} \in [1, n + 1)$. We pick a step size d according to probability distribution μ , and proceed as follows.

- If $X_{t-1} < d$, then the token stays put and $X_t = X_{t-1}$.
- If $X_{t-1} \in [d, d + 1)$, then the process is finished, denoted by $X_t = 0$.
- If $X_{t-1} \geq d + 1$, then the token moves to position $X_t = X_{t-1} - d$.

The process ends when state 0 is reached. Our goal is to prove a lower bound on the expectation of the random variable $T_X = \min\{t \mid X_t = 0\}$.

Note that if $\mu((0, 2)) = 0$, then $\mathbf{E}(T_X) = \infty$, because otherwise with positive probability the initial token position is in $[1, 2)$, and the token cannot be moved. Therefore, we assume in the following that $\mu((0, 2)) > 0$.

Note that the continuous token process is a generalization of the discrete token process; therefore we use the same symbol X for both. The following is the second main result of this paper, which settles the question for optimal blind search strategies over real intervals.

Theorem 4.1. *Let X be the continuous process on $[0, n + 1)$ as described. Then $\mathbf{E}_\mu(T_X) = \Omega((\log n)^2)$ for all distributions μ .*

In the remainder of this section we prove Theorem 4.1. For this, we first generalize the definition of the discrete interval process to a continuous version, and then apply a potential function argument as before.

4.1. The continuous interval process

As in the discrete case we describe an interval process $Y = (Y_0, Y_1, \dots)$ that is ‘equivalent’ to the token process. The state space is $B = [0, n]$. State $y \in B - \{0\}$ corresponds to a situation in which the token is uniformly distributed over the interval $[1, y + 1)$, and state 0 corresponds to a situation in which the token is at 0, *i.e.*, the token process is finished.

The start state Y_0 is n . Now consider a state $y \in B$ and let $Y_{t-1} = y$. The next state, $Y_t = y'$, is determined by a random experiment. The idea is the following.

Suppose that the token is uniformly distributed in $J_y := [1, y + 1)$ and step size d is picked at random. We split the interval J_y into three parts:

$$I_1 = [1, d), \quad I_2 = [d, d + 1) \quad \text{and} \quad I_3 = [d + 1, y + 1).$$

(If $a \geq b$, then $[a, b)$ denotes the empty interval.) Suppose the token is somewhere in I_1 . Then applying step size d cannot move the token, so it remains in I_1 , uniformly distributed. If the token is in I_2 , then the step of length d moves it to 0. Finally, if the token is in I_3 ,

applying step size d moves it to a position in $[1, y - d + 1)$, again maintaining uniformity. The idea is to first randomly pick one of the intervals I_1, I_2, I_3 , thus deciding in which of these intervals x is located. We then ‘apply’ step size d , and choose y' as either d , 0 , or $y - d$, which corresponds to the new distribution of the token, given that it was in I_1, I_2 , or I_3 , respectively.

If the token is uniformly distributed in J_y , it is located in $I_\gamma, \gamma \in \{1, 2, 3\}$, with probability $p_{y,d}^{(\gamma)}$, where

$$\begin{aligned}
 p_{y,d}^{(1)} &= \frac{|I_1 \cap J_y|}{|J_y|} = \max\left\{0, \frac{d-1}{y}\right\}, \\
 p_{y,d}^{(2)} &= \frac{|I_2 \cap J_y|}{|J_y|} = \frac{\min\{y+1, d+1\} - \max\{1, d\}}{y}, \quad \text{and} \\
 p_{y,d}^{(3)} &= \frac{|I_3 \cap J_y|}{|J_y|} = \max\left\{0, \frac{y-d}{y}\right\}.
 \end{aligned}$$

To summarize, we obtain $Y_t = y'$ by the following random experiment.

- (1) Pick d at random according to distribution μ .
- (2) Pick $\gamma \in \{1, 2, 3\}$ at random, such that $\Pr(\gamma = j) = p_{y,d}^{(j)}$ for $j = 1, 2, 3$.
- (3) Let

$$Y_t = y' = \begin{cases} d-1 & \text{if } \gamma = 1, \\ 0 & \text{if } \gamma = 2, \\ y-d & \text{if } \gamma = 3. \end{cases}$$

The interval process ends when it enters state 0 . Thus, we are interested in the random variable $T_Y = \min\{t \mid Y_t = 0\}$.

4.2. Equivalence of the token process and the interval process

We first show that our construction of Y_t matches the intuitive connection between X_t and Y_t . The proof of the following lemma is a straightforward extension to the continuous case of the proof of Lemma 4 in [2].

Lemma 4.2. *Let X_t and Y_t be the random variables as defined above (both using the same distribution μ). Let Z_t be chosen uniformly at random from $[1, Y_t + 1)$ if $Y_t > 0$, and otherwise let $Z_t = 0$. Then X_t and Z_t are identically distributed.*

Proof. By induction on t . For $t = 0$ the claim is trivially true, as X_0 is chosen uniformly at random from $[1, n + 1)$ and $Y_0 = n$.

Now let $t > 0$. Consider two random experiments leading to characterizations of the random variables X_t and Z_t , in terms of triples of random variables (X_{t-1}, d, Y_{t-1}) and (X', d, Y_{t-1}) , respectively. We argue that (X_{t-1}, d, Y_{t-1}) and (X', d, Y_{t-1}) are identically distributed. Moreover, the characterization of X_t in terms of (X_{t-1}, d, Y_{t-1}) is the same as the characterization of Z_t in terms of (X', d, Y_{t-1}) . This yields the desired result.

Let d be the step size used in the t th step of the token process. Since (X_{t-1}, d, Y_{t-1}) and (X', d, Y_{t-1}) use the same random variable d , we may assume that d is fixed. According to

the rules of the token game we have

$$\begin{aligned}
 d > X_{t-1} &\Rightarrow X_t = X_{t-1}, \\
 d \in (X_{t-1} - 1, X_{t-1}] &\Rightarrow X_t = 0, \\
 d \leq X_{t-1} - 1 &\Rightarrow X_t = X_{t-1} - d,
 \end{aligned} \tag{4.1}$$

where by the induction hypothesis X_{t-1} is uniformly distributed in $[1, Y_{t-1} + 1)$, given Y_{t-1} .

Now we characterize Z_t in terms of (X', d, Y_{t-1}) , where X' is defined as follows. Consider the t th step of the interval process leading from state Y_{t-1} to Y_t and using step size d . Then Z_t is uniformly distributed over $[1, Y_t)$, given Y_t . The definition of X' depends on the value of γ that was picked in this step as follows.

Case 1: $\gamma = 1$. We let $X' = Z_t$. Hence, X' is uniformly distributed over $[1, Y_t + 1) = [1, d)$, as this case implies $Y_t = d - 1$.

Case 2: $\gamma = 2$. We pick X' uniformly at random from $[d, d + 1)$. Note that this case implies $Z_t = 0$.

Case 3: $\gamma = 3$. We let $X' = Z_t + d$. Hence, X' is distributed uniformly over $[d + 1, d + Y_t + 1) = [d + 1, Y_{t-1} + 1)$, as this case implies $Y_t = Y_{t-1} - d$.

Given the probabilities for each of the three options, X' is uniformly distributed over $[1, Y_{t-1} + 1)$. As this is the case for any choice of d , (X', d, Y_{t-1}) and (X_t, d, Y_{t-1}) are identically distributed. From the three cases above, we read off

$$\begin{aligned}
 X' < d &\Rightarrow Z_t = X', \\
 d \leq X' < d + 1 &\Rightarrow Z_t = 0, \\
 X' \geq d + 1 &\Rightarrow Z_t = X' - d.
 \end{aligned}$$

This is the same as (4.1) with X' instead of X_{t-1} and Z_t instead of X_t . Hence, Z_t and X_t are identically distributed. □

Corollary 4.3. *For all probability distributions μ , we have $\mathbf{E}_\mu[T_X] = \mathbf{E}_\mu[T_Y]$.*

Note that Lemma 3.2 (the same statement for the discrete case) is merely a special case of this corollary.

4.3. Definition of the potential function

We analyse the continuous interval process using almost the same potential function argument as for the discrete process. For integers $d, a \in \{1, \dots, n\}$, define

$$\begin{aligned}
 \hat{\mu}(d) &= \mu((d - 1, d + 1) \cap (0, n]), \\
 \hat{\sigma}_a &= \sum_{1 \leq d \leq n} \hat{\mu}(d) \cdot 2^{-|\log a - \log d|/2}, \\
 \hat{\varphi}_a &= \frac{1}{a \hat{\sigma}_a},
 \end{aligned}$$

$$\hat{\Phi}(y) = \sum_{1 \leq a \leq \lceil y \rceil} \hat{\phi}_a, \quad \text{and}$$

$$\hat{\Phi}_t = \hat{\Phi}(Y_t).$$

Note that if one replaces $\hat{\mu}$ with μ in the definitions of $\hat{\sigma}_a$, $\hat{\phi}_a$, and $\hat{\Phi}(y)$, then one obtains σ_a , ϕ_a and $\Phi(y)$, respectively. Moreover, $\hat{\mu} = \mu$ for distributions μ concentrated on $\{1, \dots, n\}$. Also, note that due to our assumption $\mu((0, 2)) > 0$, we have $\hat{\mu}(1) > 0$ and thus $\hat{\sigma}_a > 0$ for all $a \in \{1, \dots, n\}$.

Hence, Lemma 3.3 also holds for $\hat{\Phi}_t$ instead of Φ_t .

4.4. Bounding the expected loss of potential

We analyse the expected loss of potential, given that the interval process is in state $Y_{t-1} = y$ at some point of time $t - 1$. More precisely, we show that Lemma 3.4 holds for $\hat{\Phi}_t$ instead of Φ_t . Then Theorem 4.1 follows, with the same arguments as in the discrete case.

Lemma 4.4. *There is a constant C such that $\mathbf{E}(\hat{\Phi}_{t-1} - \hat{\Phi}_t \mid Y_{t-1} = y) \leq C$, for $0 \leq y \leq n$.*

Proof. It is immediate from the definitions that for any integer $z \in \{0, \dots, n\}$, all points $y' \in (z - 1, z] \cap B$ have the same potential $\hat{\Phi}(z)$. For an integer z , let

$$\Delta(y, z) = (\hat{\Phi}(y) - \hat{\Phi}(z)) \cdot \Pr(Y_t \in (z - 1, z] \mid Y_{t-1} = y).$$

Then clearly

$$\mathbf{E}(\hat{\Phi}_{t-1} - \hat{\Phi}_t \mid Y_{t-1} = y) = \sum_{0 \leq z \leq n} \Delta(y, z) = \sum_{0 \leq z \leq \lceil y \rceil} \Delta(y, z).$$

We consider $\Delta(y, \lceil y \rceil)$, $\Delta(y, 0)$, and $\sum_{1 \leq z < \lceil y \rceil} \Delta(y, z)$ separately.

Bounding $\Delta(y, \lceil y \rceil)$. Since $\Phi(y) = \Phi(\lceil y \rceil)$, clearly $\Delta(y, \lceil y \rceil) = 0$.

Bounding $\Delta(y, 0)$. According to the definition of the process, a step from $Y_{t-1} = y$ to $Y_t = 0$, given that a step size d is chosen in that step, has conditional probability $p_{y,d}^{(2)}$ for $d \leq y$ and probability 0 for $d > y$. Since $\hat{\Phi}(0) = 0$, the potential difference is $\hat{\Phi}(y)$. Thus, using $p_{y,d}^{(2)} \leq m := \min\{1/y, 1\}$, we have

$$\Delta(y, 0) \leq m \cdot \mu((0, y]) \cdot \hat{\Phi}(y) \leq m \cdot \left(\sum_{1 \leq d \leq \lceil y \rceil} \hat{\mu}(d) \right) \cdot \left(\sum_{1 \leq a \leq \lceil y \rceil} \hat{\phi}_a \right).$$

Following the calculations starting with (3.10) and ending with (3.11), and replacing y with $\lceil y \rceil$, μ with $\hat{\mu}$, and ϕ with $\hat{\phi}$, we obtain

$$\begin{aligned} \Delta(y, 0) &\leq m \cdot \sum_{a \leq \lceil y \rceil} \sqrt{\lceil y \rceil} / a \leq \min\{1/y, 1\} \cdot \sqrt{\lceil y \rceil} \sum_{a \leq \lceil y \rceil} 1/a \\ &\leq \min\{\sqrt{\lceil y \rceil} / y, \sqrt{\lceil y \rceil}\} \cdot \ln(\lceil y \rceil) = O(1). \end{aligned}$$

(As in the discrete version of these calculations, we use $\hat{\mu}(1) > 0$ and $\hat{\sigma}_a > 0$ so that the denominators of the fractions are non-zero.)

Bounding $\sum_{1 \leq z < \lceil y \rceil} \Delta(y, z)$. Fix $y \in B$ and $z \in \{1, \dots, n\}$ such that $0 < z < y$. Suppose the process is in state $Y_{t-1} = y$ and the next step leads to some state $Y_t \in (z - 1, z]$, $0 < z < y$. According to the definition of the process, this event happens if and only if one of the following two events occurs for some step size d that is used in the step from Y_{t-1} to Y_t .

- $\gamma = 1$ was chosen and thus $Y_t = d - 1$. Hence, $d \in (z, z + 1]$, and consequently $p_{y,d}^{(1)} \leq (d - 1)/y \leq z/y$. It follows that this event occurs with probability at most

$$\mu((z, z + 1]) \cdot \frac{z}{y} \leq \hat{\mu}(z + 1) \cdot \frac{z}{y}.$$

- $\gamma = 3$ was chosen and thus $Y_t = y - d$. Hence, $d \in [y - z, y - (z - 1))$, and consequently $p_{y,d}^{(3)} = (y - d)/y \leq z/y$. It follows that this event occurs with probability at most

$$\mu([y - z, y - z + 1)) \cdot \frac{z}{y} \leq \hat{\mu}(\lceil y \rceil - z) \cdot \frac{z}{y}.$$

If one of the two events above occurs (for fixed z), the loss of potential is $\hat{\Phi}(y) - \hat{\Phi}(z) = \sum_{z < a \leq \lceil y \rceil} \hat{\varphi}_a$. Thus, weighting this loss of potential with the sum of probabilities for the events (1) and (2), and summing over all z , we obtain

$$\begin{aligned} \sum_{1 \leq z < \lceil y \rceil} \Delta(y, z) &\leq \sum_{1 \leq z < \lceil y \rceil} \sum_{z < a \leq \lceil y \rceil} \hat{\varphi}_a \cdot \frac{z}{y} (\hat{\mu}(z + 1) + \hat{\mu}(\lceil y \rceil - z)) \\ &< 2 \sum_{1 \leq z < \lceil y \rceil} \sum_{z < a \leq \lceil y \rceil} \hat{\varphi}_a \cdot \frac{z}{\lceil y \rceil} (\hat{\mu}(z + 1) + \hat{\mu}(\lceil y \rceil - z)). \end{aligned}$$

(For the last inequality we used $y > 1$ and thus $1/y < 2/\lceil y \rceil$.) Except for the factor of 2, this is the same as (3.12), for $\mu = \hat{\mu}$, $\varphi = \hat{\varphi}$, and replacing $\lceil y \rceil$ with y . Thus, with exactly the same calculations as those following (3.12), we obtain $\sum_{1 \leq z < \lceil y \rceil} \Delta(y, z) = O(1)$. \square

References

- [1] Aspnes, J., Diamadi, Z. and Shah, G. (2002) Fault-tolerant routing in peer-to-peer systems. In *Proc. 21st Annual Symposium on Principles of Distributed Computing (PODC 2002)*, pp. 223–232.
- [2] Aspnes, J., Diamadi, Z. and Shah, G. (2003) Fault-tolerant routing in peer-to-peer systems. [arXiv:cs/0302022v2](https://arxiv.org/abs/cs/0302022v2).
- [3] Droste, S., Jansen, T. and Wegener, I. (2006) Upper and lower bounds for randomized search heuristics in black-box optimization. *Theory Comput. Syst.* **39** 525–544.
- [4] Jägersküpfer, J. (2007) Algorithmic analysis of a basic evolutionary algorithm for continuous optimization. *Theoret. Comput. Sci.* **379** 329–347.
- [5] Kiefer, J. (1953) Sequential minimal search for a maximum. *Proc. Amer. Math. Soc.* **4** 502–506.
- [6] Rowe, J. E. and Hidović, D. (2004) An evolution strategy using a continuous version of the Gray-code neighbourhood distribution. In *Genetic and Evolutionary Computation (GECCO 2004)*, Part I, Vol. 3102 of *Lecture Notes in Computer Science*, Springer, pp. 725–736.