

Tight Bounds for the Cover Times of Random Walks with Heterogeneous Step Lengths

Brieuc Guinard 

IRIF, CNRS,
University of Paris, France
guinard@irif.fr

Amos Korman 

IRIF, CNRS,
University of Paris, France
<https://amoskorman.com/>
amos.korman@irif.fr

Abstract

Search patterns of randomly oriented steps of different lengths have been observed on all scales of the biological world, ranging from microscopic to the ecological, including in protein motors, bacteria, T-cells, honeybees, marine predators, and more, see e.g., [21, 22, 31, 33, 34, 35, 36]. Through different models, it has been demonstrated that adopting a variety in the magnitude of the step lengths can greatly improve the search efficiency. However, the precise connection between the search efficiency and the number of step lengths in the repertoire of the searcher has not been identified.

Motivated by biological examples in one-dimensional terrains, a recent paper studied the best cover time on an n -node cycle that can be achieved by a random walk process that uses k step lengths [7]. By tuning the lengths and corresponding probabilities the authors therein showed that the best cover time is roughly $n^{1+\Theta(1/k)}$. While this bound is useful for large values of k , it is hardly informative for small k values, which are of interest in biology [2, 4, 25, 30]. In this paper, we provide a tight bound for the cover time of such a walk, for every integer $k > 1$. Specifically, up to lower order polylogarithmic factors, the cover time is $n^{1+\frac{1}{2k-1}}$. For $k = 2, 3, 4$ and 5 the bound is thus $n^{4/3}$, $n^{6/5}$, $n^{8/7}$, and $n^{10/9}$, respectively. Informally, our result implies that, as long as the number of step lengths k is not too large, incorporating an additional step length to the repertoire of the process enables to improve the cover time by a polynomial factor, but the extent of the improvement gradually decreases with k .

2012 ACM Subject Classification Theory of computation → Random walks and Markov chains; Applied computing → Computational biology

Keywords and phrases Computational Biology, Randomness in Computing, Search Algorithms, Random Walks, Lévy Flights, Intermittent Search, CCRW

Digital Object Identifier 10.4230/LIPIcs.STACS.2020.28

Related Version A full version of this paper is available at <https://hal.archives-ouvertes.fr/hal-02303873v1>.

Funding This work has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreement No 648032).

1 Introduction

This paper follows the “Natural Algorithms” line of research, aiming to contribute to biological studies from an algorithmic perspective [6, 10, 17, 26]. In particular, we follow a similar approach to Chazelle [10, 11], considering a process that has been extensively studied by physicists and biologists, and offering a more uniform algorithmic analysis based on techniques from probability theory. Our subject of interest is random walks with heterogeneous step



lengths, a family of processes that during the last two decades has become a central model for biological movement, see e.g., [12, 21, 22, 25, 27, 29, 31, 32, 34, 35, 36]. Our approach is to quantify by how much can the search efficiency improve when the searcher is allowed to use more steps. Specifically, our goal is to analyze, for every integer k , the best cover time achievable by a random walk that utilizes k step-lengths, and identify the parameters that achieve the optimal cover time. Hence, in some sense, we view the number of steps as a “hardware” constraint on the searcher, and ask what is the best “software” to utilize them, that is, the best way to set the lengths, and the probabilities of taking the corresponding steps. We focus on the one-dimensional terrain (an n -node cycle) as it is both biologically relevant, and, among other Euclidean spaces, it is the most sensitive to step-length variations (e.g., the simple random walk on the two-dimensional plain already enjoys a quasi-linear cover time). A preliminary investigation of this question was recently done by the authors of the current paper together with collaborating researchers [7], yielding asymptotic bounds with respect to k . Unfortunately, these bounds are not very informative for small values of k , which are of particular interest in biology [2, 4, 25, 30]. For example, for processes that can use a small number of step-lengths, say $k = 2$ or $k = 3$, the bound in [7] merely says that the cover time is polynomial in n , which does not even imply that such a process can outperform the simple random walk – whose cover time is known to be $\Theta(n^2)$. In this paper we improve both the lower bound and the upper bound in [7], identifying the tight cover time for every integer k .

1.1 Background and Motivation

The exploration-exploitation dilemma is fundamental to almost all search or foraging processes in biology [19]. An efficient search strategy needs to strike a proper balance between the need to explore new areas and the need to exploit the more promising ones found. At an intuitive level, this is often perceived as a tradeoff between two scales: the global scale of exploration and the local scale of exploitation. This paper studies the benefits of incorporating a hierarchy of multiple scales, where lower scales serve to exploit the exploration made by higher scales. We demonstrate this concept by focusing on random walk search patterns with heterogeneous step lengths, viewing the usage of steps of a given length as searching on a particular scale.

In the last two decades, random walks with heterogeneous step lengths have been used by biologists and physicists to model biological processes across scales, from microscopic to macroscopic, including in DNA binding proteins [5, 14], immune cells [18], crawling amoeba [33], locomotion mode in mussels [15, 22], snails [31], marine predators [21, 34], albatrosses [35, 36], and even in humans [9, 32, 29]. Most of these biological examples concern search contexts, e.g., searching for pathogens or food. Indeed, from a search efficiency perspective, it has been argued that the heterogeneity of step lengths in such processes allows to reduce oversample, effectively improving the balance between global exploration and local exploitation [4, 36]. However, the precise connection between the search efficiency and the number of step lengths in the repertoire of the searcher has not been identified.

Due to possible cognitive conflicts between motion and perception, in some of the aforementioned search contexts it was argued that biological entities are essentially unable to detect targets while moving fast, and hence targets are effectively found only between jumps, see e.g., [4, 25] and the references therein. Those models are often called *intermittent*. When the search is intermittent, we say that a site is *visited* whenever the searcher completes a jump landing on this site. It is also typically assumed that the searcher has some radius of visibility r , and a target can only be detected if it is in the r -vicinity of a site currently

visited by the searcher. Discretizing the space, one may view a Euclidian space as a grid of the appropriate dimension, in which each edge is of length r . In this discretization, sites are nodes, and the searcher can detect a target at a node, only if it makes a random jump that lands on it.

In general, two families of processes with heterogeneous step lengths have been extensively studied in Euclidean spaces: *Lévy Flights* (named after the mathematician Paul Lévy), and *Composite Correlated Random Walks (CCRW)*, see e.g., [2, 4, 25]. Both have been claimed to be optimal under certain conditions and both have certain empirical support. In the Lévy Flight process, step lengths have a probability distribution that is heavy-tailed: at each step a direction is chosen uniformly at random, and the probability to perform a step of length d is proportional to $d^{-\mu}$, for some fixed parameter $1 < \mu < 3$.

Searchers employing a CCRW can potentially alternate between multiple modes of search¹, but apart for few exceptions [30], such patterns have mostly been studied when assuming that the number of search modes is 2. Specifically, a diffusive phase in which targets can be detected and a ballistic phase in which the searcher moves in a random direction in a straight line whose length is exponentially distributed with some mean L . This CCRW with 2 modes can be approximated as a discrete random walk with two step lengths, hereafter called *2-scales search*: first, choose a direction uniformly at random. Then, with some probability p take a step of unit length, and otherwise, with probability $1 - p$, take a step of some predetermined length L .

Lévy Flights and 2-Scales searches have been studied extensively using differential equation techniques and computer simulations. These studies aimed to both compare the performances of these processes as well as to identify the parameters that maximize the rate of target detection or minimize the hitting time under various target distributions [4, 12, 25, 27, 36].

Most of the literature on the subject has concentrated on either one or two dimensional Euclidian spaces. In particular, the one-dimensional case has attracted attention due to several reasons. First, it finds relevance in several biological contexts, including in the reaction pathway of DNA binding proteins [5, 14]. One-dimension can also serve as an approximation to general narrow and long topologies, which can be found for example in blood veins or other organs. Second, from a computational perspective, the one-dimension is the only dimension where the simple random walk has a large cover time, namely, quadratic, whereas in all higher dimensions the cover time is nearly linear. This implies that in terms of the cover time, heterogeneous random walks can potentially play a much more significant role in one-dimension than in higher dimensions.

1.2 Definitions

We model the one-dimension space as an n -node cycle, termed C_n . For an integer k , we define the random walks process with k step lengths as follows.

► **Definition 1** (*k-scales search*). *A random walk process X is called a k -scales search on C_n if there exists a probability distribution $\mathbf{p} = (p_i)_{i=0}^{k-1}$, where $\sum_i p_i = 1$, and integers L_0, L_1, \dots, L_{k-1} such that, on each step, X makes a jump $\{0, -L_i, +L_i\}$ with probability*

¹ CCRW have also been classified as either *cue-sensitive*, i.e., they can change their mode of operation upon detecting a target [3], or *internally-driven*, i.e., their movement pattern depends only on the mechanism internal to the searcher [23]. However, when targets are extremely rare and there is no a-priori knowledge about their distribution, one must cover a large portion of the terrain before finding a target, and hence the aforementioned distinction becomes irrelevant.

respectively $p_i/2, p_i/4, p_i/4$. Overall, with probability $1/2$, the process X stays in place². The numbers (p_i) and (L_i) are called the parameters of the search process X . The speed is assumed to be a unitary constant, that is, it takes L time to do a step of length L .

Our goal is to show upper and lower bounds on the *cover time* of a k -scales search, that is, the expected time to visit every node of the ambient graph C_n , where it is assumed that a jump from some point x to y visits only the endpoint y , and not any of the intermediate nodes. We denote by $\mathbb{E}(t_{cov}(n, k))$ the smallest cover time achievable by a k -scales search over the n -node cycle. The parameters n and k are omitted when clear from the context.

We also define the following k -scales search which is often referred to in the mathematical literature as a *Weierstrassian random walk* [20]. In the biology literature, it has been used as a model for the movement strategy of snails [31] and mussels [30].

► **Definition 2** (Weierstrassian random walk). *Let $b \geq 2$ and k be integers such that $b^{k-1} < n \leq b^k$. The Weierstrassian random walk with parameter b is the k -scales search defined by: $L_i = b^i$ and $p_i = c_b b^{-i}$, for every $0 \leq i \leq k-1$, with the normalizing constant $c_b = \frac{b-1}{b-b^{1-k}}$.*

Note that c_b is an increasing function of $b > 1$, and so $c_b \geq c_2 \geq 1/2$ for $b \geq 2$. Hence, $p_0 = c_b \geq 1/2$. Also $p_0 = c_b \leq 1$, hence $c_b = \Theta(1)$ is indeed a constant.

1.3 Previous Bounds on the Cover Time of k -scales search

The work of Lomholt et al. [25] considered intermittent search on the one-dimensional cycle of length n , and compared the performances of the best 2-scales search to the best Lévy Flight. With the best parameters, they showed that the best 2-scales search can find a target in roughly $n^{4/3}$ expected time, but introducing Lévy distributed relocations with exponent μ close to 2 can reduce the search time to quasi linear.

Taking a more unified computational approach, a recent paper [7] analyzed the impact of having k heterogeneous step lengths on the cover time (or hitting time³) of the n -node cycle C_n . Specifically, the following bounds were established in [7].

► **Theorem** (Upper bound on the cover time of Weierstrassian random walk from [7]). *Let b, n be integers such that $2 \leq b < n$ and set $k = \log n / \log b$. The cover time of the Weierstrassian random walk with parameter b on the n -cycle is at most $\text{poly}(k) \cdot \text{poly}(b) \cdot n \log n$.*

Taking $b = \lceil n^{1/k} \rceil$ yields the following corollary.

► **Corollary** (Upper bound from [7]). *For any $k \leq \frac{\log n}{\log \log n}$, there exists a k -scales search with cover time $n^{1+O(\frac{1}{k})} \log n$.*

Note that for small values of k , this bound is not very informative. For example, for $k = 2, 3$ the bound merely says that the cover time is polynomial in n , which is known already for $k = 1$, i.e., the simple random walk, whose cover time is $\Theta(n^2)$.

² This laziness assumption is used for technical reasons, as is common in many other contexts of random walks. Note that this assumption does not affect the time performance of the process, as we consider it takes time 0 to stay in place.

³ Note that in connected graphs, the notion of *cover time*, namely the expected time until all sites (of a finite domain) are visited when starting the search from the worst case site, is highly related to the *hitting time*, namely, the expected time to visit a node x starting from node y , taken on the worst case pair x and y ; the cover time is always at least the hitting time, and in connected graphs it is at most a logarithmic multiplicative factor more than the hitting time, see [24][Matthews method, Theorem 11.2].

► **Theorem** (Lower bound from [7]). *For every $\varepsilon > 0$, there exist sufficiently small constants $c, c' > 0$ such that for $k \leq c' \frac{\log n}{\log \log n}$, any k -scales search cannot achieve a cover time better than $c \cdot n^{1 + \frac{1/2 - \varepsilon}{k+1}}$.*

The aforementioned lower bound of [7] is more precise than the upper bound, but still not tight, as we show in the next subsection. For example, for $k = 2$, the lower bound in [7] gives $n^{7/6}$ instead of $n^{4/3}$, which is the tight bound.

1.4 Our Results

This paper provides tight bounds for the cover times of k -scales searches, for any integer $k > 1$. Specifically, we prove that the optimal cover (or hitting) time achievable by a k -scales search is $n^{1 + \frac{1}{2k-1}}$, up to lower order polylogarithmic factors. Our bound implies that for small k , the improvement in the cover time incurred by employing one more step length is polynomial, but the extent of the improvement gradually decreases with k .

In order to establish the tight bound, we first had to understand what should be a good candidate for the tight bound to aim to. This was not a trivial task, as the precise bound takes an unusual form. After identifying the candidate for the bound, we had to improve both the upper and the lower bounds from [7], which required us to overcome some key technical difficulties. For the lower bound, [7] established that the cover time is bounded from below by a function (specifically the square-root) of the ratio L_{i+1}/L_i , for every i . As it turns out, what was required to tighten the analysis is a better understanding about the relationships between the cover time and the extreme step-lengths, namely, L_0 , L_1 and L_{k-1} . Specifically, in proving the precise lower bound we have two components, one for the “local” part (exploitation) and the other for the “global” part (exploration). We showed that in order to be efficient on the local part, the small step-lengths need to be small, whereas in order to be efficient on the global part (traversing large distances fast), the largest step-length needs to be large. This allowed us to widen the ratios between consecutive step-lengths, consequently increasing the lower bound.

In order to obtain the precise upper bound, we improved the analysis in [7] of the Weierstrassian random walk process. This, in particular, required overcoming non-trivial issues concerning dependencies between variables that were overlooked in [7]. By doing this, we also refined the estimates on the order of magnitude of other dependencies. In addition, we had to incorporate short-time probability bounds for each step-length used by the process, and perform a tighter analysis of the part of the walk that corresponds to the largest step length L_{k-1} .

We next describe our contribution in more details.

1.4.1 The Lower Bound

We begin with the statement of the lower bound. The formal proof is given in Section 2.

► **Theorem 3.** *Let k and n be positive integers. The cover time of any k -scales search X on C_n is:*

$$\mathbb{E}(t_{cov}(n, k)) = n^{1 + \frac{1}{2k-1}} \cdot \Omega(1/k).$$

1.4.2 The Upper Bound

The following theorem implies that up to lower order terms, the cover time of the Weierstrassian random walk matches the lower bound of the cover time of any k -scales search, as given by Theorem 3, for $2 \leq k \leq \log n$, i.e., for all potential scales.

28:6 Tight Bounds for the Cover Times of Random Walks

► **Theorem 4.** *Let k be an integer such that $2 \leq k \leq \log_2 n$. The Weierstrassian random walk with parameter $b = \lfloor n^{\frac{2}{2k-1}} \rfloor$ is a k -scales search that achieves a cover time of:*

$$n^{1+\frac{1}{2k-1}} \cdot O(k^2 \log^2 n).$$

Observe that combining Theorems 3 and 4 we obtain the best cover time $Cov_{k,n}$ achievable by a k -scales search on C_n , which is $\tilde{\Theta}\left(n^{1+\frac{1}{2k-1}}\right)$ for any $2 \leq k \leq \log n$. For particular values of k , we thus have:

k	1	2	3	4	5	...	$\log n$
$\mathbb{E}(t_{cov}(n, k))$	$\Theta(n^2)$	$\tilde{\Theta}(n^{\frac{4}{3}})$	$\tilde{\Theta}(n^{\frac{6}{5}})$	$\tilde{\Theta}(n^{\frac{8}{7}})$	$\tilde{\Theta}(n^{\frac{10}{9}})$...	$O(n \log^3 n)$

Theorem 4 follows immediately from the following more general theorem, by taking $b = n^{\frac{2}{2k-1}}$.

► **Theorem 5.** *Let b, k, n be integers such that $b^{k-1} < n \leq b^k$. The cover time of the Weierstrassian random walk on C_n with parameter b is*

$$O\left(n \max\left\{\frac{b^k}{n}, \frac{n}{b^{k-1}}\right\} \cdot k^2 \cdot \log b \cdot \log n\right) = \tilde{O}\left(\max\left\{b^k, \frac{n^2}{b^{k-1}}\right\}\right).$$

The formal proof of Theorem 5 is deferred to the full version. In Section 3 we provide a sketch of the proof.

As mentioned, Theorem 5 using the particular value $b = n^{\frac{2}{2k-1}}$ gives a tight upper bound for k -scales search. However, since the Weierstrassian random walk is of independent interest as it is used in biology, it might be useful to understand its cover time also for other values of b . Note that Lemmas 6 and 7 below, when applied to the Weierstrassian random walk on C_n , show that the cover time is at least $\Omega\left(\max\{n\sqrt{b}, \frac{n^2}{b^{k-1}}\}\right)$. This is quite close to the bound $\tilde{O}\left(\max\left\{b^k, \frac{n^2}{b^{k-1}}\right\}\right)$ of the theorem. Indeed if $n \geq b^{k-\frac{1}{2}}$, both bounds match, up to logarithmic terms. If $n \leq b^{k-\frac{1}{2}}$, the ratio of the bounds is $\frac{b^{k-\frac{1}{2}}}{n}$.

2 The Lower Bound Proof

The goal of this section is to establish the lower bound in Theorem 3. For this purpose, consider a k -scales search X on the cycle C_n and denote $(L_i)_{i=0}^{k-1}$ its step lengths with $L_i < L_{i+1}$ for all $i \in [k-2]$. For convenience of writing we also set $L_k = n$, but it should be clear that it is actually not a step length of the walk. Let p_i denote the probability of taking the step length L_i .

The theorem will follow from the combination of two lemmas. The first one, Lemma 6, stems from the analysis of the number of nodes that can be visited during L_{i+1} time steps. It forces $L_0 L_1$ as well as the ratios L_{i+1}/L_i for all $1 \leq i \leq k-1$ to be small enough in order to have a small cover time. The second one, Lemma 7, comes from bounding the cover time by the time it takes to go to a distance of at least $n/3$. It forces L_{k-1} to be big enough to have a small cover time.

► **Lemma 6.** *The cover time of X is at least*

- $\mathbb{E}(t_{cov}) = \Omega(n\sqrt{L_0 L_1})$.
- $\mathbb{E}(t_{cov}) = \Omega\left(\frac{n}{k} \sqrt{\frac{L_{i+1}}{L_i}}\right)$ for any $1 \leq i \leq k-1$.

The second part of Lemma 6 was already given in [7]. We sketch here the ideas behind the proof of the first part, namely, that the cover time is at least of order $n\sqrt{L_0L_1}$. Essentially, we count the expected number of nodes that can be visited in a time duration of L_1 , which we call a *phase*. A jump of length $L_i \geq L_1$ will not contribute to visiting a new node during this time duration. Thus, we may suppose that there are only jumps of length L_0 . Since $L_1 \leq n$, the process does not do a turn of the cycle and, therefore, it can be viewed as a walk on \mathbb{Z} . Furthermore, since every jump has length L_0 , we can couple this walk by a corresponding simple random walk, that does steps of length 1, during a time duration of L_1/L_0 . The expected number of nodes visited during a phase is thus of order $\sqrt{L_1/L_0}$. It follows that we need at least $n/(\sqrt{L_1/L_0})$ such phases before covering the cycle. Since a phase lasts for L_1 time, the cover time is at least of order $n\sqrt{L_0L_1}$. The full proof of Lemma 6, including the part that was proven in [7], appears in the full version.

► **Lemma 7.** *The cover time of X is at least $\Omega(n^2 \frac{\mu}{\sigma^2})$, where $\mu = \frac{1}{2} \sum_{i \leq k-1} p_i L_i$ and $\sigma^2 = \frac{1}{2} \sum_{i \leq k-1} p_i L_i^2$ are the mean and variance of the jump lengths, respectively. In particular, the cover time is:*

$$\mathbb{E}(t_{cov}) = \Omega\left(\frac{n^2}{L_{k-1}}\right).$$

Proof. Let m_{cov} denote the random number of *steps* before all nodes of C_n are covered, and let t_{cov} be the random cover time of the process. By Wald's identity, we have:

$$\mathbb{E}(t_{cov}) = \mathbb{E}(m_{cov}) \cdot \mu, \tag{1}$$

where $\mu = \frac{1}{2} \sum_{i=0}^{k-1} p_i L_i$ is the expected length, and hence the expected time, of a jump (the factor $\frac{1}{2}$ comes from the laziness). By Markov's inequality, we have:

$$\Pr(m_{cov} < 2\mathbb{E}(m_{cov})) \geq 1/2.$$

Let N_m be the (random) number of nodes visited by step m . We have:

$$\mathbb{E}(N_{2\mathbb{E}(m_{cov})}) \geq \mathbb{E}(N_{2\mathbb{E}(m_{cov})} \mid m_{cov} < 2\mathbb{E}(m_{cov})) \cdot \Pr(m_{cov} < 2\mathbb{E}(m_{cov})) \geq n \cdot \frac{1}{2}.$$

Define D_m as the maximal distance of the process from step 0 up to step m , i.e., $D_m = \max_{s \leq m} |X(s)|$. Since $N_m \leq 2D_m + 1$, we have:

$$2\mathbb{E}(D_{2\mathbb{E}(m_{cov})}) + 1 \geq \mathbb{E}(N_{2\mathbb{E}(m_{cov})}) \geq n/2.$$

As shown in [13] for general one-dimensional random walks, we have $\mathbb{E}(D_m) = O(\sigma\sqrt{m})$, where σ is the standard deviation of the length distribution, i.e., $\sigma^2 = \frac{1}{2} \sum_i p_i L_i^2$. Thus, we have:

$$\sqrt{\mathbb{E}(m_{cov})} \sigma = \Omega(n),$$

and so:

$$\mathbb{E}(m_{cov}) = \Omega\left(\frac{n^2}{\sigma^2}\right),$$

and by Eq. (1), we get:

$$\mathbb{E}(t_{cov}) = \Omega\left(n^2 \frac{\mu}{\sigma^2}\right) = \Omega\left(n^2 \frac{\sum_{i=0}^{k-1} L_i p_i}{\sum_{i=0}^{k-1} L_i^2 p_i}\right),$$

28:8 Tight Bounds for the Cover Times of Random Walks

which proves the first part of the lemma.

In order to prove the second part, note that since L_{k-1} is the biggest step length, we have $\sum_{i=0}^{k-1} p_i L_i (1 - \frac{L_i}{L_{k-1}}) \geq 0$, and so $\frac{\sum_{i=0}^{k-1} L_i p_i}{\sum_{i=0}^{k-1} L_i^2 p_i} \geq \frac{1}{L_{k-1}}$. Therefore,

$$\mathbb{E}(t_{cov}) = \Omega\left(\frac{n^2}{L_{k-1}}\right),$$

which completes the proof of Lemma 7. \blacktriangleleft

Next, it remains to show how Theorem 3 follows by combining Lemma 6 and Lemma 7. First, consider the lower bound of $\Omega(n^2/L_{k-1})$ in Lemma 7. If $L_{k-1} \leq n^{1-\frac{1}{2k-1}}$ then the bound in Theorem 3 immediately follows. Let us therefore assume that $L_{k-1} > n^{1-\frac{1}{2k-1}}$.

Define $\alpha_0 = L_0 L_1$ and $\alpha_i = \frac{L_{i+1}}{L_i}$ for $i \in \{1, 2, \dots, k-2\}$. As

$$\prod_{i=0}^{k-2} \alpha_i = L_0 L_{k-1},$$

there must exist an index $0 \leq i \leq k-2$ such that $\alpha_i \geq (L_0 L_{k-1})^{\frac{1}{k-1}}$. Thus, by Lemma 6, the cover time is at least

$$\Omega\left(\frac{n}{k} (L_0 L_{k-1})^{\frac{1}{2(k-1)}}\right).$$

Since $L_{k-1} > n^{1-\frac{1}{2k-1}} = n^{\frac{2k-2}{2k-1}}$ and $L_0 \geq 1$, we conclude that the cover time is at least

$$\mathbb{E}(t_{cov}) = \Omega\left(\frac{n}{k} \cdot n^{\frac{1}{2k-1}}\right),$$

as desired. This completes the proof of Theorem 3. \blacktriangleleft

3 Upper Bound Proof (Sketch)

Let us give the key ideas of the proof of Theorem 5. Some of the initial steps in the proof follow the technique in [7] (by doing so, we also corrected some mistakes in [7]). These parts are clearly mentioned below. Our main technical contribution that allowed us to obtain the precise upper bound, is the use of short-time probability bounds (see Eq. (6)), and a tighter analysis of the part of the walk that corresponds to the largest step length L_{k-1} .

In more details, let us consider the Weierstrassian walk on C_n , termed X . The following lemma establishes a link between the cover time of X and the point-wise probabilities of X . For completeness, we provide a formal proof of it in the full version, although it is not hard to obtain it using the technique in [7].

► **Lemma 8.** *If $p > 0$ and $m_0 > 0$ are such that, for any $x \in \{0, \dots, n-1\}$,*

$$\frac{\sum_{m=m_0}^{2m_0} \Pr(X(m) = x)}{\sum_{m=0}^{m_0} \Pr(X(m) = 0)} \geq p, \quad (2)$$

then the cover time of the Weierstrassian random walk X on the cycle C_n is $O(m_0 p^{-1} k \log n)$.

Using Lemma 8, the bound of Theorem 5 can be established by proving bounds on the probability to visit node $x \in [0, n)$ at step m .

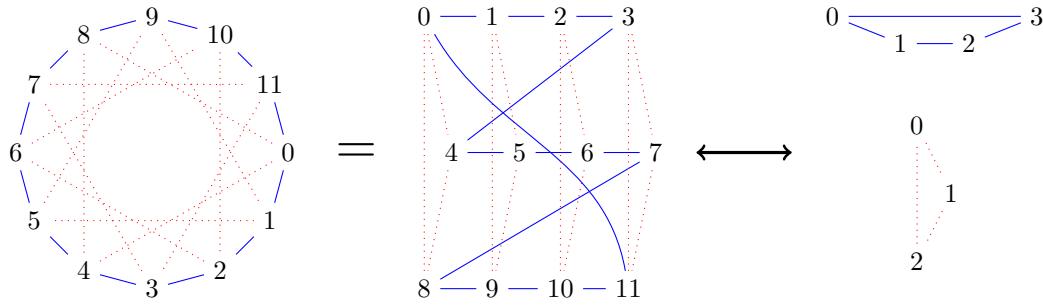


Figure 1 The first two graphs represent, in different node disposition, the Weierstrassian walk on C_{12} with parameter $b = 4$. There are $k = 2$ jump lengths, $L_0 = 1$ (blue edges) and $L_1 = b = 4$ (red, dotted edges). To the right, we show the decomposition of C_{12} as $C_4 \times C_3$. For instance the node $x = 7 \in C_{12}$ will be represented by $x_0 = 3 \in C_4$ and $x_1 = 1 \in C_3$.

In order to simplify the presentation, assume first that $n = b^k$. Proceeding first as in [7], we view the k -lengths Weierstrassian random walks as k (dependent) random walks, by grouping together the jumps of the same length (see Figure 1). Define $S_i(m)$ as the algebraic count of the jumps of lengths b^i . E.g., if, by step m , there are exactly four positive jumps of length b^i , and one negative, then $S_i(m) = 3$. We have:

$$X(m) = \sum_{i=0}^{k-1} S_i(m)b^i.$$

Define also the following decomposition of C_n .

► **Definition 9** (Base b decomposition). For any $x \in C_n$, we may decompose x in base b as

$$x = \sum_{i=0}^{k-1} x_i b^i,$$

with $0 \leq x_i < b$. We call x_i the i -th coordinate of x (in base b).

It follows from Euclidean division, and the fact that $n = b^k$, that the base b decomposition is well-defined and unique for every $x \in C_n$. This decomposition is illustrated in Figure 1 (where we have taken $n = \hat{n}b^{k-1}$ to anticipate the more general case to follow).

Note that $X(m) = x$ in C_n if and only if

$$\sum_i (S_i(m) - x_i)b^i = 0 \pmod n. \tag{3}$$

By taking Eq. (3) modulo b^i , for $i \leq k - 1$, it is easy to show that Eq. (3) is equivalent to

$$S_i(m) = y_i \pmod b,$$

for $y_i := x_i - b^{-i} \sum_{j < i} (S_j(m) - x_j)b^j \pmod b$.

Thus, $X(m) = x$ is equivalent to $R_i(m) = y_i$ for all i , where $R_i = S_i \pmod b$ is a random walk on C_b that moves with probability $\frac{b^i}{2}$. This process is illustrated in Figure 1, where $X(m) = 7$ is equivalent to $R_0(m) = 3$ and $R_1(m) = 2$.

Unfortunately, the R_i 's and the y_i 's are not independent, due to the fact that only one of the R_i can change between steps m and $m + 1$, however, let us overlook this issue in this

28:10 Tight Bounds for the Cover Times of Random Walks

informal outline. We then have:

$$\Pr(X(m) = x) \approx \prod_{i=0}^{k-1} \Pr(R_i(m) = y_i). \quad (4)$$

Recall that R_i is a random walk over C_b that moves with probability p_i . The following is a well-known property of the random walk over a cycle (see, e.g., Example 5.7 and Proposition 6.18 in [1]):

▷ **Claim 10.** For a simple random walk R on C_b that moves with probability $\frac{1}{2}$, and any $y \in C_b$,

$$\Pr(R(m) = y) = \begin{cases} O(1/\sqrt{m}) & \text{if } m < b^2 \\ b^{-1}(1 \pm \varepsilon_m) & \text{if } m \geq b^2, \end{cases} \quad (5)$$

with $\varepsilon_m = O(e^{-cm})$ where $c > 0$.

Considering that R_i moves with probability $\frac{p_i}{2} = \Theta(b^{-i})$, we can expect that, at step m , $R_i(m)$ has the same distribution as the lazy random walk with mp_i steps that moves with probability $\frac{1}{2}$. This is proved formally in the full version. Hence, by substituting m with mp_i in Claim 10, we obtain:

$$\Pr(R_i(m) = y_i) = \begin{cases} O(1/\sqrt{mp_i}) & \text{if } m < b^{i+2} \\ b^{-1}(1 \pm \varepsilon_{mp_i}) & \text{if } m \geq b^{i+2}. \end{cases} \quad (6)$$

Theorem 5 then follows from Eq. (4), Eq. (6) and Lemma 8. Essentially, to cover C_n , we need that each $R_i(m)$ is mixed, i.e., has some significant probability to visit any node y_i in C_b , which happens, as shown by Eq. (6), for $m > b^{k-1+2} = b^{k+1}$. Let us apply Lemma 8 with

$$m_0 := b^{k+1}.$$

We first establish a lower bound on $\sum_{m=m_0}^{2m_0} \Pr(X(m) = x)$. By Eq. (4) and Eq. (6), we have, for $m > m_0$,

$$\Pr(X(m) = x) \approx \prod_{0 \leq i \leq k-1} b^{-1}(1 - \varepsilon_{mp_i}) = \Theta(b^{-k}),$$

where the last equality is justified in the full version. Thus,

$$\sum_{m=m_0}^{2m_0} \Pr(X(m) = x) = \Omega(m_0 b^{-k}) = \Omega(b).$$

We need also to upper bound $\sum_{m=0}^{m_0} \Pr(X(m) = 0)$, which is the expected number of returns to the origin up to step m_0 . To do this, we shall use the short-time bounds of Eq. (6).

Let us decompose the aforementioned sum as follows.

$$\sum_{m=0}^{m_0} \Pr(X(m) = 0) = 1 + \frac{1}{2} + \sum_{j=0}^{k-1} \sum_{m=1+b^j}^{b^{j+1}} \Pr(X(m) = 0) + \sum_{m=1+b^k}^{m_0} \Pr(X(m) = 0). \quad (7)$$

Fix j , such that $1 \leq j \leq k-1$ and let $m \in (b^j, b^{j+1}]$. By Eq. (4), in order to upper bound $\Pr(X(m) = 0)$ it is enough to bound $\Pr(R_i(m) = y_i)$ for every $i \leq k-1$. For $i > j$, we

bound $\Pr(R_i(m) = y_i)$ by 1. For $i \leq j - 2$, we use Eq. (6) to upper bound $\Pr(R_i(m) = y_i)$ by $b^{-1}(1 + \varepsilon_{mp_j})$. For $i = j - 1$ and $i = j$, we bound $\Pr(R_i(m) = y_i)$ by $O(1/\sqrt{mp_{j-1}})$ and $O(1/\sqrt{mp_j})$, respectively. We thus obtain, by Eq. (4),

$$\begin{aligned} \Pr(X(m) = x) &= O\left(\frac{1}{\sqrt{mp_{j-1}}} \cdot \frac{1}{\sqrt{mp_j}} \cdot \prod_{0 \leq i \leq j-2} b^{-1}(1 + \varepsilon_{mp_j})\right) \\ &= O\left(b^{-(j-1)} \cdot \frac{\sqrt{bb^{j-1}}}{m}\right) = O\left(\frac{\sqrt{b}}{m}\right), \end{aligned}$$

where we justify in the full version that $\prod_{0 \leq i \leq j-2} (1 + \varepsilon_{mp_j}) = O(1)$. Hence, we get:

$$\sum_{m=1+b^j}^{b^{j+1}} \Pr(X(m) = 0) = O(\sqrt{b} \log b), \quad (8)$$

by using that $\sum_{m=1+b^j}^{b^{j+1}} m^{-1} = \Theta\left(\int_{m=b^j}^{b^{j+1}} u^{-1} du\right) = \Theta(\log b)$. For the case $j = 0$, we bound $\Pr(R_i(m) = y_i)$ by 1 for $i > 1$ and $\Pr(R_0(m) = y_0)$ by $O(m^{-\frac{1}{2}})$, so that, by Eq. (4), $\Pr(X(m) = 0) = O(\frac{1}{\sqrt{m}})$. Hence, we get:

$$\sum_{m=2}^b \Pr(X(m) = 0) = O(\sqrt{b}). \quad (9)$$

Similarly, for $m \in (b^k, b^{k+1}]$, $\Pr(R_i(m) = y_i)$ is bounded by $b^{-1}(1 + \varepsilon_{mp_i})$ for $i \leq k - 2$, and by $\frac{1}{\sqrt{mp_{k-1}}}$ for $i = k - 1$. Thus, for $m \in (b^k, b^{k+1}]$,

$$\Pr(X(m) = 0) = O\left(\frac{1}{\sqrt{m}\sqrt{b^{k-1}}}\right)$$

and, since $\sum_{m=1+b^k}^{b^{k+1}} \frac{1}{\sqrt{m}} = O\left(\int_{b^k}^{b^{k+1}} \frac{1}{\sqrt{u}} du\right) = O(\sqrt{b^{k+1}})$, we get:

$$\sum_{m=1+b^k}^{b^{k+1}} \Pr(X(m) = x) = O\left(\frac{\sqrt{b^{k+1}}}{\sqrt{b^{k-1}}}\right) = O(b). \quad (10)$$

In total, by Eq. (7), combining Eqs. (8), (9) and (10), we find that the expected number of returns to the origin up to step b^{k+1} is

$$\sum_{m=0}^{m_0} \Pr(X(m) = 0) = O\left(k\sqrt{b} \log b + b\right) = O(kb \log b).$$

So that all together we have:

$$\frac{\sum_{m=m_0}^{2m_0} \Pr(X(m) = x)}{\sum_{m=0}^{m_0} \Pr(X(m) = 0)} = \Omega\left(\frac{b}{kb \log b}\right) = \Omega\left(\frac{1}{k \log b}\right).$$

Thus, by Lemma 8, the cover time of X is at most:

$$O(m_0 \cdot k \log b \cdot k \log n) = O(b^{k+1} k^2 \log b \log n) = O(nbk^2 \log b \log n), \quad (11)$$

as claimed by Theorem 5, for the case where $n = b^k$.

28:12 Tight Bounds for the Cover Times of Random Walks

Consider now a more general case, in which n is a multiple of b^{k-1} . Here, we can write $n = \hat{n}b^{k-1}$, where $\hat{n} \in (0, b]$ is an integer. What changes in this case is that the last coordinate, R_{k-1} , is now a random walk over $C_{\hat{n}}$ instead of over C_b , as depicted in Figure 1. R_{k-1} is thus mixed after a number of steps:

$$\hat{n}^2 p_{k-1}^{-1} = \Theta(b^{k-1} \hat{n}^2) = \Theta(n^2 / b^{k-1}).$$

On the other hand, after $\Theta(b^{k-2+2}) = \Theta(b^k)$ steps, the other coordinates are mixed. Thus, the number of steps needed before every coordinate R_i is mixed is:

$$m_0 = \Theta(\max\{b^k, n^2 / b^{k-1}\}), \quad (12)$$

which is again the order of magnitude of the cover time of X , up to polylogarithmic factors. Note that when $n = b^k$, Eq. (12) recovers the cover time of order $\tilde{\Theta}(b^{k+1})$. Furthermore, the ratio of the cover time for $n = b^k$ and $n = \hat{n}b^{k-1}$ is of order $\frac{b^{k+1}}{\max\{b^k, b^{k-1} \hat{n}^2\}} = \min\{b, \frac{b^2}{\hat{n}^2}\}$. When b is large (which corresponds to k being small), this can be significant. Hence, naively bounding \hat{n} from above by b would not suffice to yield an optimal bound.

The general case, when n is not necessarily a multiple of b^{k-1} , needs to be treated with more care. What changes in this case is that we can no longer decompose X as k dependent random walks on $C_b \times \cdots \times C_b \times C_{\frac{n}{b^{k-1}}}$, since $\frac{n}{b^{k-1}}$ is not an integer. Instead, we define Z as the process that does the same jumps as X , but on the infinite line \mathbb{Z} , and we also define

$$\hat{n} := \lfloor n / b^{k-1} \rfloor.$$

Then, we use almost the same decomposition, where Z is viewed as k dependent random walks over $C_b \times \cdots \times C_b \times \mathbb{Z}$. The process corresponding to the last coordinate, R_{k-1} , is now a random walk on \mathbb{Z} , and we are interested especially on the probability of the event $R_{k-1}(m) = x_{k-1}$ for $x_{k-1} \in [0, \hat{n}]$. As the coordinate R_{k-1} is not restricted to $[0, \hat{n}]$, we need to pay attention that the walk does not go too far.

4 Discussion

The upper bound in Theorem 4 implies that almost linear time performances, as those obtained by Lévy Flights, can be achieved with a number of step lengths that ranges from logarithmic to linear. This further suggests that cover time performances similar to those of Lévy Flights can be seen by a large number of different processes. In practice, if one aims to fit empirical statistics of an observed process to a theoretical model of a particular heterogeneous step length distribution, the large degree of freedom can make this task extremely difficult, if not impossible. On the other hand, the fact that so many processes yield similar cover times may justify viewing all of them as essentially equivalent. This interpretation may also be relevant to the current debate regarding whether animals' movement is better represented by Lévy Flights or by CCRW distributions with 2 or 3 scales [28, 15, 22, 30]. Moreover, the fact that many heterogeneous step processes yield similar performances to Lévy Flights may imply that limiting the empirical fit to either Lévy Flights or CCRW searches with 2 or 3 scales may be too restrictive. Our work may suggest that instead, the focus could shift to identifying the number of scales involved in the search.

When combined with appropriate empirical measurements, our lower bound can potentially be used to indirectly show that a given intermittent process uses strictly more than a certain number of step lengths. For example, if the process is empirically shown as a heterogeneous random walk whose cover time is almost linear, then Theorem 3 implies that it must use

roughly logarithmic number of step lengths. From a methodological perspective, such a result would be of particular appeal as demonstrating lower bounds in biology through mathematical arguments is extremely rare [8, 16].

Finally, we note that most of the theoretical research on heterogeneous search processes which is based on differential equation techniques and computer simulations. In contrast, and similarly to [7], our methodology relies on algorithmic analysis techniques and discrete probability arguments, which are more commonly used in theoretical computer science. We believe that the computational approach presented here can contribute to a more fundamental understanding of these search processes.

References

- 1 D. Aldous and J. A. Fill. Reversible markov chains and random walks on graphs, 2002.
- 2 M. Auger-Méthé, A. Derocher, M. Plank, E. Codling, and M. Lewis. Differentiating the lévy walk from a composite correlated random walk. *Methods in Ecology and Evolution*, 6(10):1179–1189, 2015.
- 3 Simon Benhamou and Julien Collet. Ultimate failure of the lévy foraging hypothesis: Two-scale searching strategies outperform scale-free ones even when prey are scarce and cryptic. *Journal of theoretical biology*, 387, October 2015.
- 4 Olivier Bénichou, C Loverdo, M Moreau, and R Voituriez. Intermittent search strategies. *Reviews of Modern Physics*, 83(1), 2011.
- 5 Otto G. Berg, Robert B. Winter, and Peter H. Von Hippel. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. models and theory. *Biochemistry*, 20(24), 1981.
- 6 L. Boczkowski, O. Feinerman, A. Korman, and E. Natale. Limits for rumor spreading in stochastic populations. In *9th Innovations in Theoretical Computer Science Conference*, pages 49:1–49:21, 2018.
- 7 L. Boczkowski, B. Guinard, A. Korman, Z. Lotker, and M. Renault. Random walks with multiple step lengths. *LATIN 2018: Theoretical Informatics*, pages 174–186, January 2018.
- 8 L. Boczkowski, E. Natale, O. Feinerman, and A. Korman. Limits on reliable information flows through stochastic populations. *PLOS Computational Biology*, 14(6):1–15, June 2018.
- 9 D. Boyer, G. Ramos-Fernandez, O. Miramontes, J. Mateos, G. Cocho, H. Larralde, H. Ramos, and F. Rojas. Scale-free foraging by primates emerges from their interaction with a complex environment. *Proceedings of the Royal Society of London B: Biological Sciences*, 273(1595), 2006.
- 10 Bernard Chazelle. Natural algorithms. In *Proceedings of the twentieth annual ACM-SIAM symposium on Discrete algorithms*, pages 422–431, 2009.
- 11 Bernard Chazelle. The convergence of bird flocking. *J. ACM*, 61(4):21:1–21:35, 2014.
- 12 M. Chupeau, O. Bénichou, and R. Voituriez. Cover times of random searches. *Nature Physics*, 11(10):844, 2015.
- 13 A. Comtet and S. N. Majumdar. Precise asymptotics for a random walker’s maximum. *Journal of Statistical Mechanics: Theory and Experiment*, 6:06013, June 2005.
- 14 M. Coppey, O. Bénichou, R. Voituriez, and M. Moreau. Kinetics of Target Site Localization of a Protein on DNA: A Stochastic Approach. *Biophysical Journal*, 87(3), 2004.
- 15 Monique de Jager, Franz J Weissing, Peter MJ Herman, Bart A Nolet, and Johan van de Koppel. Lévy walks evolve through interaction between movement and environmental complexity. *Science*, 332(6037):1551–1553, 2011.
- 16 Ofer Feinerman and Amos Korman. Theoretical distributed computing meets biology: A review. In *International Conference on Distributed Computing and Internet Technology*, pages 1–18. Springer, 2013.
- 17 Ofer Feinerman and Amos Korman. The ANTS problem. *Distributed Computing*, 30(3):149–168, 2017.

- 18 T. Harris, E. Banigan, D. Christian, et al. Generalized lévy walks and the role of chemokines in migration of effector cd8(+) t cells. *Nature*, 486(7404), 2012.
- 19 T. Hills, P. Todd, D. Lazer, A. Redish, I. Couzin, and Cognitive Search Research Group. Exploration versus exploitation in space, mind, and society. *Trends in cognitive sciences*, 19(1):46–54, 2015.
- 20 B. D. Hughes, M. F. Shlesinger, and E. W. Montroll. Random Walks with Self-Similar Clusters. *Proceedings of the National Academy of Science*, 78:3287–3291, June 1981.
- 21 N. E. Humphries, N. Queiroz, J. R. M. Dyer, N. G. Pade, M. K. Musyl, K. M. Schaefer, D. W. Fuller, J. M. Brunnschweiler, T. K. Doyle, J. D. R. Houghton, G. C. Hays, C. S. Jones, L. R. Noble, V. J. Wearmouth, E. J. Southall, and D. W. Sims. Environmental context explains Lévy and Brownian movement patterns of marine predators. *Nature*, 465:1066–1069, June 2010.
- 22 Vincent A. A. Jansen, Alla Mashanova, and Sergei Petrovskii. Comment on “lévy walks evolve through interaction between movement and environmental complexity”. *Science*, 335(6071):918–918, 2012.
- 23 A. Koelzsch, A. Alzate, F. Bartumeus, M. de Jager, E. Weerman, G. Hengeveld, M. Naguib, B. Nolet, and J. van de Koppel. Experimental evidence for inherent lévy search behaviour in foraging animals. *Proceedings of the Royal Society B*, 282, May 2015.
- 24 David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Society, 2008.
- 25 Michael Lomholt, Koren Tal, Ralf Metzler, and Joseph Klafter. Lévy strategies in intermittent search processes are advantageous. *Proceedings of the National Academy of Sciences*, 105(32), 2008.
- 26 Cameron Musco, Hsin-Hao Su, and Nancy A. Lynch. Ant-inspired density estimation via random walks: Extended abstract. In *Proceedings of the 2016 ACM Symposium on Principles of Distributed Computing, PODC 2016, Chicago, IL, USA, July 25-28, 2016*, pages 469–478, 2016.
- 27 G. Oshanin, H. Wio, K. Lindenberg, and S. Burlatsky. Intermittent random walks for an optimal search strategy: one-dimensional case. *Journal of Physics: Condensed Matter*, 19(6), 2007.
- 28 G. H. Pyke. Understanding movements of organisms: it’s time to abandon the lévy foraging hypothesis. *Methods in Ecology and Evolution*, 6(1):1–16, 2015.
- 29 D. A. Raichlen, B. M. Wood, A. D. Gordon, A. Z. P. Mabulla, F. W. Marlowe, and H. Pontzer. Evidence of Lévy walk foraging patterns in human hunter-gatherers. *Proceedings of the National Academy of Science*, 111:728–733, January 2014.
- 30 A. Reynolds. Mussels realize Weierstrassian Lévy walks as composite correlated random walks. *Scientific Reports*, 4:4409, March 2014.
- 31 A. Reynolds, G. Santini, G. Chelazzi, and S. Focardi. The weierstrassian movement patterns of snails. *Royal Society Open Science*, 4(6):160941, 2017.
- 32 I. Rhee, M. Shin, S. Hong, K. Lee, and S. Chong. On the lévy-walk nature of human mobility. In *IEEE INFOCOM 2008*, 2011.
- 33 F. L. Schuster and M. Levandowsky. Chemosensory responses of *acanthamoeba castellanii*: Visual analysis of random movement and responses to chemical signals. *Journal of Eukaryotic Microbiology*, 43(2):150–158, 1996.
- 34 D. W. Sims, E. J. Southall, N. E. Humphries, G. C. Hays, C. J. A. Bradshaw, J. W. Pitchford, A. James, M. Z. Ahmed, A. S. Brierley, M. A. Hindell, D. Morritt, M. K. Musyl, D. Righton, E. L. C. Shepard, V. J. Wearmouth, R. P. Wilson, M. J. Witt, and J. D. Metcalfe. Scaling laws of marine predator search behaviour. *Nature*, 451:1098–1102, February 2008.
- 35 G. Viswanathan, V. Afanasyev, S. Buldyrev, E. Murphyt, P. Princet, and H. Stanley. Lévy flight search patterns of wandering albatrosses. *Nature*, 381(6581), 1996.
- 36 G. Viswanathan, S. Buldyrev, S. Havlin, M. da Luz, E. Raposo, and E. Stanley. Optimizing the success of random searches. *Nature*, 401(6756), 1999.